ASHG 2023 Poster Abstracts

As of May 24, 2024

Instructions

1. View the Table of Contents on the next page.
    a. The Table of Contents is sorted chronologically by Poster Board Number.
2. Click on the Abstract Title of the abstract you would like to view, and you will be taken to the page the abstract is on.

Table of Contents

Session Title: Molecular Effects of Genetic Variation Poster Session I

PB1001 3D Structure-based development of prediction model for the functional assessment of genetic variation of RAG1 and RAG2 complex

Authors:

N. Haque[1], T. Kawai[2], B. Ratnasinghe[1], R. Urrutia[1], L. Notarangelo[3], M. Zimmermann[1]; [1]Med. Coll. of Wisconsin, Milwaukee, WI, [2]Natl. Inst. of Allergy and Infectious Disease, NIH, Bethesda, Bethesda, MD, [3]Natl. Inst. of Allergy and Infectious Disease, NIH, Bethesda, Wauwatosa, MD

The current pace of genomic sequencing for patient-derived samples is unprecedented. However, the methods available for the interpretation of the effects of genetic changes on the patient's phenotype is in its early stage of development and is largely dependent of the information derived from individual sequence or their horologes. Sequence-based pathogenicity prediction performs well for some genomic variants but fails to confidently rank variants of uncertain significance (VUS). There are numerous immune deficiency syndromes that are associated with genetic variations unique to individual patients. In this study we have developed a method for genomic interpretation of variation by mathematical combination of features derived from sequence and their 3D structures. The method has been developed on one of the most vulnerable RAG1/RAG2 protein complex, central to the immune system. The variant effect prediction model was developed by leveraging enzymatic activity values of 182 clinically observed missense variants of RAG1 and RAG2. An efficient regression model was built and tested on a set of experimental data, a test set not used in model development, separately for RAG1 (R2 = 0.914 and enzyme activity RMSE = 10.34%) and RAG2 (R2 = 0.973; RMSE = 9.57%) to predict RAG activity changes from a combination of sequence- and structure-based scores. Such a model with ~10% of RMSE was further used for assessment of 711 RAG1/RAG2 variants obtained from various databases. Thus, we have demonstrated the development of advanced methods which has potential generalize the assessment of effect of genetic variation of clinically significant genetic products in rare diseases and cancers.

Session Title: Molecular Effects of Genetic Variation Poster Session II

PB1002 3DVariantVision: Multimodal framework to interpret genetic variant effects based on 3D chromatin

Authors:

J. Yang, H. Wang, J. Wang; Michigan State Univ., East Lansing, MI

Deciphering the effects and mechanisms of genetic variants is essential in human genetics and precision medicine. Traditional GWAS approach suffers from low statistical power and lack of causality. Recently, deep-learning based models were developed to extract sequence information de novo to predict functional variant effects on nearby regions. However, due to the large-scale cell-type specific multi-omic data and complex chromatin folding conformation, it is extremely difficult to prioritize functional genetic variants and delineate their multi-level effects, including TF binding, epigenetics, gene expression, and long-range chromatin contacts. Here we developed 3DVariantVision, a new deep learning-based multi-modal framework to provide the comprehensive vision on genetic variant effects from genotype to phenotype by integrating diverse panels of genome-wide datasets, including ChIP-seq, Hi-C, and eQTL summary statistics. Relying on DNA sequence only, 3DVariantVision demonstrates superior performance to predict candidate functional genetic variants and their disruptive effects on specific TF bindings, histone modifications, enhancer-promoter interactions, and gene expression, leading to boosted capability of eQTL discoveries. By capturing the non-linear complex regulatory grammar across multiple levels of features and across long-range genomic distances, 3DVairantVision accurately identifies the distal target genes of non-coding variants, revealing novel mechanistic insights into trait-associated variants beyond traditional GWAS and TWAS studies.

Session Title: Molecular Effects of Genetic Variation Poster Session III

PB1003 A biobank catalogue of transcriptomes and associated genetic effects based on 2,000 subjects uncovers the causal effects of Middle Eastern genetic variation and uncovers novel disease mechanisms

Authors:

Y. Mokrab[1], H. Naeem[1], I. Diboun[1], M. Ghorbani[1], R. Razali[1], E. Aliyev[1], M. Hashmi[1], R. Mathew[1], V. Mattei[1], N. James[1], L. Mathew[1], L. Silcock[1], K. Wang[1], M. K. R. Kalikiri[1], F. Vempalli[1], M. Tamanni[1], A. Khouly[1], S. Poolat[1], A. Akil[1], S. Lorenz[1], R. Ali[1], Qatar Genome Program Research Consortium, K. Fakhro[1], S. Tomei[1], D. Chaussabel[1], S. Montgomery[2]; [1]Sidra Med., Doha, Qatar, [2]Stanford Univ., Stanford, CA

Mapping genetic variation to transcriptome activity at a population level is a robust approach to linking regulatory mechanisms to traits and diseases, allowing both association and causality to be inferred. Here, we present a large study of predominantly healthy individuals from Qatar Biobank, combining whole genome sequencing (WGS) (30x coverage; n=6,216) and whole blood RNA-Seq (20-100 M reads; n=2,127) data to build a comprehensive catalogue of genetic variants regulating various transcription traits. Investigating the effects of single nucleotide polymorphisms (SNPs), insertions/deletions (Indels), and structural variants (SVs) on gene expression, splicing (isoform-level and intron-level), and allelic expression (variant-level and gene-level) in cis, identifying ~12 million novel associations of eQTLs, isoQTLs, sQTLs, and aseQTLs. We co-localized the QTL signals with a GWAS on 6,218 subjects from the same QBB cohort, featuring 45 clinically relevant traits. With this analysis, we colocalized 40 causal SNPs in 30 genes and validating 8 which were previously linked to the GWAS traits, while establishing novel trait relationship between 22 genes. Our colocalization analysis also led to finetuning the associations for 9 SNPs where the identified QTL gene was other than the most proximal gene to the GWAS SNP, an example is our identified QTL rs11234557 for SUGP1 with an established role in lipid homeostasis. Allele frequency analysis of our fine-mapped QTL SNPs revealed significant differences between the Arab major sub ancestries and the East Asian (EAS) ancestry from the 1KG. IPA enrichment analysis for the corresponding genes revealed a signature for viral infection, cancer, endocrine disorders, immunological and inflammatory disease for the causal variants with varying frequencies with respect to EAS as opposed to respiratory, hematological, cardiovascular, and hepatic disease for South Asians (SAS). This study produced the largest whole blood multi-QTL resource available to date based on non-imputed data and samples, uniformly generated and processed in a single center. By revealing novel loci in a Middle Eastern population, this resource illustrates the need for greater diversity in public reference datasets in order to gain more complete understanding of functional impact of genetic variation in the global human population.

Session Title: Molecular Effects of Genetic Variation Poster Session II

PB1004 A common flanking variant is associated with enhanced meiotic stability of the *FGF14*-SCA27B locus

Authors:

M. Danzi[1], D. Pellerin[2], G. Del Gobbo[3], M. Couse[4], E. Dolzhenko[5], M-J. Dicaire[2], A. Rebelo[1], V. Roth[6], M. Wandzel[6], C. Bonnet[6], C. Ashton[2], P. Lamont[7], N. Laing[8], M. Renaud[6], G. Ravenscroft[8], H. Houlden[9], M. Synofzik[10], M. Eberle[5], K. Boycott[3], T. Pastinen[11], All of Us Long Reads Working Group, B. Brais[2], S. Zuchner[1]; [1]Univ. of Miami, Miami, FL, [2]McGill Univ., Montreal, QC, Canada, [3]CHEO Res. Inst., Ottawa, ON, Canada, [4]BC Children and Women's Hosp., Vancouver, BC, Canada, [5]Pacific BioSci.s, Oceanside, CA, [6]Univ. Hosp. of Nancy, Nancy, France, [7]Royal Perth Hosp., Perth, Australia, [8]Harry Perkins Inst. of Med. Res., Perth, Australia, [9]Univ. Coll. London, London, United Kingdom, [10]Hertie Inst. for Clinical Brain Res., Tuebingen, Germany, [11]Children's Mercy Hosp. and Clinics, Kansas City, MO

The factors driving initiation of pathological expansion of tandem repeats remain largely unknown. Late-onset spinocerebellar ataxia type 27B caused by intronic GAA repeat expansions in *FGF14* has recently been discovered as one of the most common genetic causes of adult-onset ataxia worldwide. Initial observations suggested that intermediate and expanded *FGF14* alleles are highly unstable in the germline. Leveraging long-read PacBio HiFi sequencing on 2,191 individuals and Sanger sequencing on 339 individuals, we here investigated if sequence variants in or near the *FGF14* repeat locus affect the meiotic stability of the locus. We identified a 5'-flanking 17-bp deletion-insertion in 70.34% of alleles (3,463/4,923) which is uniquely associated with alleles containing fewer than 30 GAA repeat units (odds ratio: 44,158, 95% confidence interval: 13,981 - 139,474). Furthermore, this common flanking variant also predicted meiotic stability of the repeat locus (odds ratio: 47.04, 95% confidence interval: 25.24 - 85.15) based on a total of 470 meiosis events measured by PacBio sequencing (403 events) and Sanger sequencing (67 events). To our knowledge, this is the first evidence of a stabilizing variant flanking the *FGF14*-SCA27B locus and the first description of a genomic element that so strongly predicts the stability of a short tandem repeat in humans. This finding may yield further insight into the mechanisms underlying tandem repeat expansions and facilitate the identification of similar sequence variants at other known pathogenic repeat loci.

Session Title: Molecular Effects of Genetic Variation Poster Session III

PB1005 A compendium of HLA allelic associations with plasma protein levels in the UK Biobank

Authors:

C. Adams, C. Benner, R. K. Pendergrass, M. I. McCarthy, A. Mahajan, C. Hammer; Genentech, South San Francisco, CA

Human leukocyte antigen (HLA) alleles are established risk factors for many immune-related diseases. They can also affect protein expression, in particular through the initiation of signaling cascades in response to immune activation upon antigen presentation and recognition. For example, cytokine expression, which is important in immunity and inflammation, is partially determined by MHC peptide recognition by CD4 T-cells. The systematic measurement of plasma protein levels (PPLs) in a subset of participants from the UK Biobank (UKB) has previously been used to identify more than 10,000 non-MHC quantitative trait loci (pQTL). Here, we present an HLA allele-wide association study of PPLs in UKB participants. We estimated associations between 2,940 PPLs and imputed HLA alleles in 35,571 UKB participants of European ancestry. PPLs were measured using the Olink Explore 3072 platform and inverse rank normalized. Statistically independent HLA allele associations were identified with stepwise conditional testing. We also estimated associations for HLA class I alleles grouped according to their interactions with killer cell immunoglobulin-like receptors (KIR) (Bw4 vs. Bw6, C1 vs. C2). Regression analyses were adjusted for age, sex, batch, UKB center, array, time between blood sampling and measurement, and 20 principal components. We also examined interaction between PPLs associated with disease-relevant HLA alleles, making use of genetically predicted protein levels (GPPLs) composed of non-MHC pQTLs. We identified 4,408 statistically independent associations between 1,809 PPLs and 97 HLA alleles. Of these, 4,035 were trans associations with 1,775 proteins outside the extended MHC region. To demonstrate the complex relationship between HLA status, protein expression and disease risk, we investigated the interaction between a GPPL for CCL19 and HLA-DRB1*03:01 or HLA-DRB1*04:01 in type 1 diabetes, finding significant interaction terms ($CCL19_{GPPL}$×$DRB1*03:01_{DOM}$: OR_{IXN}=2.26, p=0.001; $CCL19_{GPPL}$×$DRB1*04:01_{DOM}$: OR_{IXN}=1.95, p=0.03). Both alleles were associated with CCL19 PPLs, as well as the disease phenotype. Furthermore, 109 PPLs were significantly associated with at least one KIR ligand group after Bonferroni correction, including proteins encoded by genes located in the leukocyte receptor complex on chromosome 19, which harbors all KIR and other immune-relevant genes. Our results demonstrate and build on the important contribution of HLA alleles to protein levels and disease risk. This resource will be useful to improve our understanding of the relationship between complex traits, immunogenetics, and disease-relevant protein networks.

Session Title: Molecular Effects of Genetic Variation Poster Session I

PB1006 A DNA repair-reporter deep mutational scan to comprehensively map missense effects in *MUTYH*

Authors:

S. Hemker, A. Bashir, B. Burugula, G. Clark, S. Jayakody, E. Kirk, J. Kitzman; Univ. of Michigan, Ann Arbor, MI

For the vast majority of genetic variants, the impacts upon gene function and organismal phenotype are unknown. This remains the case even at deeply studied genetic disease loci such as *MUTYH*, which underlies a recessive cancer predisposition syndrome for which up to ~1% of individuals worldwide are carriers. Of the ~800 *MUTYH* missense variants reported in ClinVar, 95% remain as variants of uncertain significance (VUS). This reflects the difficulty of classifying rare missense variants, which can range from having no effects to complete functional impairment. Traditional experimental and bioinformatic methods lack the scale and accuracy to resolve these and as-yet unseen *MUTYH* VUSs. Therefore, we established a multiplexed assay of variant effects (MAVE) to profile every *MUTYH* missense variant. Because MAVEs typically rely upon cell survival to measure variant function, non-cell-essential disease genes have mostly been off-limits to this powerful approach. We overcome this limitation using a high-throughput DNA damage reporter. In this assay, the repair event catalyzed by MUTYH - excision at a 8oxoG:A mispair lesion, the major post-replication byproduct of oxidative damage - results in GFP expression. We generated CRISPR-induced *MUTYH* knockout human cell lines, into which we stably integrate libraries of *MUTYH* missense variants. Variants that restore MUTYH function, and therefore GFP expression, can be sorted *en masse* via flow cytometry. Following deep sequencing, a function score is calculated based upon each variant's enrichment/depletion between the starting and the sorted, repair-positive cell populations. In pilot studies, we have profiled all 969 missense variants in a 51-codon region overlapping a key catalytic domain. Nonsense and synonymous variants are well separated in this assay, and we observed loss of function effects at known pathogenic variants such as the p.Tyr151Cys founder allele. Function scores are being generated for the remainder of the protein, and are being calibrated using clinical variants with standing interpretation. These results will inform mechanistic studies of *MUTYH* and related DNA glycosylases and allow for resolution of clinically reported VUS.

Session Title: Molecular Effects of Genetic Variation Poster Session II

PB1007 A high-resolution view of human gene expression and splicing diversity with long-read sequencing

Authors:

S. Yan, H. Eden, M. Li, A. Battle, W. Timp, R. C. McCoy; Johns Hopkins Univ., Baltimore, MD

Differences in gene expression are the primary source of phenotypic diversity within and between species. Variation in both RNA transcript abundance and splicing can have dramatic consequences for the function of proteins and non-coding transcripts, in turn driving phenotypes and disease. Short-read sequencing, which still dominates functional genomic datasets, requires computational reconstruction of RNA isoforms. In contrast, long-read sequencing enables the direct identification of complete RNA isoforms, allowing for accurate resolution of splicing and exon structure across full transcripts. Recently published long-read studies have demonstrated that more than half of all RNA isoforms observed in a single long-read sequenced sample are absent from existing annotations based on short reads. However, these studies—as well as many other RNA sequencing datasets—predominantly include individuals of European ancestry, thus limiting our view of global gene expression diversity and evolutionTo address these limitations, we generated Nanopore long-read RNA sequencing data from five lymphoblastoid cell lines, derived from globally diverse individuals in the 1000 Genomes Project (representing populations from Africa, the Americas, Europe, East Asia, and South Asia). By aligning this sequencing data to the complete T2T-CHM13 reference genome, we observe improved resolution of transcripts within repetitive genomic regions, such as the HLA and immunoglobulin loci, when compared to short-read RNA-sequencing of the same samples. We additionally discover 158,735 novel isoforms unannotated in GENCODE, underscoring the power of long reads for transcript discovery. Finally, we intersect these novel transcripts with rare splice-altering variation in the respective genomes, establishing mechanisms that explain the observed isoform diversity. Together, our results demonstrate how long-read sequencing of diverse individuals improves understanding of human gene expression and splicing diversity and evolution.

Session Title: Molecular Effects of Genetic Variation Poster Session III

PB1008 A high-throughput screen to nominate phenotypically relevant asthma genes from genome-wide association studies.

Authors:

I. Salamone, E. Zhong, C. Ober, X. He, M. Nobrega; Univ. of Chicago, Chicago, IL

Genome-wide association studies (GWAS) often implicate non-coding variants in disease risk, suggesting that common variants modulate disease risk through transcriptional regulation of downstream genes. Although advances in functionally informed fine-mapping methods have led to increasingly confident identification of causal variants, it is still difficult to pinpoint the gene or genes that are regulatory targets of these variants. The difficulty of this task, which is commonly described as the variant-to-function problem, is further compounded in complex diseases such as asthma, where phenotypes arise as a result of perturbations of transcriptional programs in many cell types and causal variants may have phenotypic impact on multiple genes and across multiple tissues or cell types. In fact, asthma is considered a group of diseases with shared symptoms that result from many distinct mechanisms including airway inflammation, mucus overproduction, and epithelial barrier dysfunction. To improve nomination of target genes associated with asthma, we have developed a high-throughput CRISPR/Cas9-based screen. This screen allows us to test the role of all putative target genes at a GWAS locus and select candidates that modulate disease-relevant cellular phenotypes. As proof of principle, we have nominated a pool of putative target genes using functionally informed fine-mapping of adult- and childhood-onset GWAS loci. By performing our screen in bronchial epithelial cells, which play a primary role in the etiology of asthma, we are able to evaluate how candidate genes influence barrier maintenance. Overall, this screen will identify both candidate genes at asthma risk-associated loci that may play a role in complex disease pathogenesis as well as the cell type in which they are acting. These results will not only allow us to address the variant-to-function problem, but also elucidate the genetic architecture of this group of diseases called asthma that affect hundreds of millions of people worldwide.

Session Title: Molecular Effects of Genetic Variation Poster Session I

PB1009 A hyperVNTR is located in intron 8 of the human dopamine transporter gene (SLC6A3).

Authors:

D. Vandenbergh[1], A. Apsley[2], M. Verbiest[3], E. Domico[4], M. Anisimova[3]; [1]Penn State Univ, University Park, PA, [2]The Pennsylvania State Univ., University Park, PA, [3]Zürich Univ. of Applied Sci., Wädenswil, Switzerland, [4]Johns Hopkins Univ., Baltimore, MD

Multiple genetic analyses of SLC6A3 have generated conflicting results for dopamine-related phenotypes, which might result from an inability to test relevant loci in the gene. We identified 5 Tandem Repeats that vary in copy number (VNTR) in SLC6A3 using Tandem Repeat Annotation Library (TRAL), and characterized 3 that have not been studied. Copy number and sequence similarity of each repeat unit of the five VNTRs is reported, along with the correlations of SNP-SNP, SNP-VNTR and VNTR-VNTR alleles across the gene. One VNTR in intron 8 is hyper-variable (hyVNTR) with a range of 3.4-133.4 repeat copies and 46 alleles in the 64 long-read haplotype-phased chromosomes analyzed (93% heterozygosity). The consensus sequence of the repeat unit is 38 bp with 82% G+C content. The repeat is predicted to form G-quadruplexes (G4s, G-tetrads) in silico, which was confirmed by circular dichroism spectroscopy. Multiple putative PRDM9 binding sites (recruiter of recombination promoting proteins) are present in the repeat, and it is in low linkage disequilibrium (LD) with flanking genetic markers (r2=0.016 and 0.001), suggesting it is a hotspot for recombination. Studies of other sites in SLC6A3 could not estimate genetic effects of this hyVNTR due to lack of LD.

Session Title: Molecular Effects of Genetic Variation Poster Session II

PB1010 A likely pathogenic variant in *RYR2* causing Catecholaminergic polymorphic ventricular tachycardia in an Iranian family.

Authors:

N. Naderi, M. Maleki, S. Kalayinia; Rajaie Cardiovascular Med. and Res. Ctr., Tehran, Iran, Islamic Republic of

Background: Catecholaminergic polymorphic ventricular tachycardia (CPVT) is an arrhythmia disorder from stress or physical activity. CPVT is often associated with syncope or sudden cardiac death, especially in childhood, and accounts for about 50-60% of cases. Genetic changes also have a significant role in CPVT etiology. Next-generation sequencing (NGS) technology can be helpful in the identification of heterogenic disease causes such as CPVT. **Method:** After confirming the CPVT in the patient by a cardiologist, blood samples of the patient and her family members were received. DNA extraction and whole-exome sequencing (WES) were performed to identify the causative variants. Finally, PCR and Sanger-sequencing were done for confirmation and segregation of the candidate variant. **Result:** A likely pathogenic heterozygous missense variant, c.A12280G:p.I4094V, was detected in the *RYR2* gene. The *in silico* analysis of rs1573933519 indicated the pathogenicity of this variant. The c.A12280G was confirmed in the patient and segregated in the pedigree available members. **Conclusion:** The *RYR2* mutations are involved in Arrhytmogenic right ventricular dyspasia, ventricular arrhythmias and Catecholaminergic polymorphic ventricular tachycardia. Considering that the CPVT leads to syncope and cardiac arrest in 30% of cases, genetic evaluation of these patients using the WES technique is important approach.**Keywords:** Catecholaminergic polymorphic ventricular tachycardia (CPVT), *RYR2* gene, whole exome sequencing (WES)

Session Title: Molecular Effects of Genetic Variation Poster Session III

PB1011 A multi-omics approach to characterize genetic architecture of plasma lipidome in Hispanics/Latinos

Authors:

W. Zhu[1], H-H. Chen[2], A. Petty[3], H. Highland[4], L. Petty[1], M. Yaser[5], K. Young[6], P. Meikle[7], J. Curran[8], J. McCormick[9], S. Fisher-Hoch[10], K. North[11], E. Gamazon[12], J. Below[13]; [1]Vanderbilt Univ., Nashville, TN, [2]Academia Sinica, Taipei, Taiwan, [3]Vanderbilt Univ. Med. Ctr., Nashville, TN, [4]Univ North Carolina at Chapel Hill, Chapel Hill, NC, [5]Univ. of North Carolina Chapel Hill, Chapel Hill, NC, [6]Univ North Carolina, Chapel Hill, NC, [7]Baker Heart and Diabetes Inst., Melbourne, Australia, [8]Univ. of Texas Rio Grande Valley, Brownsville, TX, [9]UTH SPH Brownsville, Brownsville, TX, [10]Univ. of Texas Sch. of Publ. Hlth., Brownsville, TX, [11]Univ North Carolina, Wake Forest, NC, [12]VUMC Clare Hall, Univ. of Cambridge, Nashville, TN, [13]Vanderbilt Univ Med Ctr., Nashville, TN

Background: Large GWAS studies in Europeans identified hundreds of plasma lipid-related variants, yet these findings may not capture genetic risk factors in non-European populations. Similarly, prediction models trained on European datasets tend to have poor transferability across populations. Multiple studies have revealed that one of the major risk factors of cardiometabolic disorder, dyslipidemia, is highly prevalent in H/Ls (30%). However, H/Ls population remains underrepresented in genetic studies despite of their profound health disparities in metabolic diseases. While traditional lipid measures such as HDL and LDL are easily accessible in many cohorts, they are insufficient to characterize the dynamic and subtle changes of the entire lipidome. Alternatively, lipidomics provides a complete snapshot of the underlying lipid metabolism within individuals, providing a valuable resource to understand the biological mechanisms of dyslipidemia in genetic studies. **Methods:** Subjects from the Cameron County Hispanic Cohort (CCHC) were genotyped on the Illumina MEGA^{EX} array and imputed to the TOPMed phase 8 reference panel following standard protocols. Plasma lipidomic profiles from 2500 individuals were obtained using a chromatography-mass spectrometry pipeline. We estimated SNP-based heritability and performed lipidome-wide GWASs in 2289 individuals across 830 lipid species and 49 lipid classes, using linear mixed models adjusted for sex, age, age2, BMI, PCs and relatedness. Within the maximum independent set (N=1680) constructed by PRIMUS, we filtered variants by GWAS p-values and developed lipidomic prediction models by elastic net regression. **Results:** Our GWASs identified approximately 2500 significant associations between genetic variants and 830 lipid species from the 49 lipid classes. In addition to associations of 35 previously reported genes, we also identified novel signals on chromosomes 1,2, 5, 6, 8, 9 and 22 near genes such as BASP1, PACSIN2 and MCAT that have been previously linked to lipid binding and lipid metabolism. For lipid classes, we estimated heritability between 0.12 to 0.55, and between 0.09 to 0.89 for lipid species. Performances of our prediction models aligned with heritability estimations.

Session Title: Molecular Effects of Genetic Variation Poster Session I

PB1012† A new hypothesis for the molecular mechanisms underlying dominant diseases

Authors:

D. Stainier, C. Dooley, L. Falcucci, K. Hsu, G. Jakutis, B. Juvik, V. Serobyan, J. Welker, L. Xie; Max Planck Inst. for Heart and Lung Res., Bad Nauheim, Germany

The onset and progression of dominant diseases are thought to be due to haploinsufficiency, gain-of-function, or the action of dominant negative/poison proteins. Here, we propose an alternative hypothesis, namely a newly discovered cellular response to deleterious mutations called transcriptional adaptation (TA). During TA, mutant mRNA degradation products, or their derivatives, translocate to the nucleus where they modulate the expression of so-called adapting genes. We initially identified this process in zebrafish while investigating the discrepancy between knockout and knockdown phenotypes, and subsequently in mouse and human cells in culture, as well as in *Caenorhabditis elegans*. Depending on the adapting genes whose expression is modulated, TA can lead to genetic compensation or a worsening of the phenotype. Importantly, TA is a dominant phenomenon in the sense that it can be observed in heterozygous cells and animals.
Recent studies have challenged the current concepts of haploinsufficiency or dominant negative/poison proteins as the underlying mechanism for certain dominant diseases including frontotemporal lobar degeneration (FTLD), Brugada syndrome (BrS), Marfan syndrome (MFS), Dravet syndrome, and hypertrophic cardiomyopathy (HCM). We hypothesize that for these, and other, dominant diseases, when the underlying mutations lead to mutant mRNA degradation, the phenotypes are due, at least in part, to the dysregulation of gene expression due to TA.
This presentation will briefly summarize the key findings about TA, both published and unpublished, and go over recent data from our own experiments, and others, that aim to test this new hypothesis in the context of FTLD, BrS, MFS, and HCM.

Session Title: Molecular Effects of Genetic Variation Poster Session II

PB1013 A novel framework for functional annotation of ASD and ID linked variants of uncertain significance.

Authors:

A. Bhattacharya[1], P. Parlanti[1], L. Cavallo[1], E. Farrow[2], T. Spivey[2], A. Renieri[3], F. Mari[3], M. Manzini[1]; [1]Child Hlth.Inst. of New Jersey and Dept. of NeuroSci. and Cell Biology, Rutgers – RWJMS, New Brunswick, NJ, [2]Dept. of Pharmacology and Physiology, Sch. of Med. and Hlth.Sci., George Washington Univ., Washington, DC, [3]Med. Genetics, Univ. of Siena, Siena, Italy

Autism Spectrum Disorder (ASD) and **Intellectual Disability** (ID) are clinically heterogeneous neurodevelopmental disorders (NDD) with a complex genetic architecture. Over a hundred risk genes with highly penetrant, rare loss of function mutations have been discovered and clinical exome sequencing is increasingly adding to the list of ASD and ID linked variants in these risk genes. However, most of the variants discovered are variants of uncertain significance (**VUS**), which represent a major bottleneck in the field, preventing the use of genetic knowledge to inform clinical decisions. We sought to develop a framework to characterize VUS in *Coiled-coil and C2 domain containing 1A* (*CC2D1A*; MIM *610055), a gene leading to moderate to severe autosomal recessive ID with comorbid ASD in 40% of cases.We report four ASD-ID cases (three males, one female) with compound heterozygous missense VUS (p.Pro319Leu, p.Ser327Leu, p.Gly441Val, p.Val449Met, p.Thr580Ile, p.Arg886His and p.Glu910Lys) that were cloned and overexpressed either individually or in patient combination in HEK293 cells. Overexpression of VUS neither affected protein stability nor the survival of the transfected cells. CC2D1A is a signaling scaffold which modulates protein kinase A (PKA)/CREB activity by repressing phosphodiesterase PDE4D, an important node for learning and memory. To test PDE4D interaction, we immunoprecipitated Myc from cells co-transfected with *PDE4D5*-Myc and mutant (or WT) *CC2D1A* and quantified the amount of CC2D1A pulldown. Only V449M reduced PDE4D5 binding of CC2D1A. However, P319L, T580I and R886H showed a significant decrease in cAMP levels in competitive ELISA, suggesting that PDE4D activity may be affected by these variants even without disrupting binding. Next, we used luciferase assay to test their effect on CREB activation as a read out of the signaling activity. Compared to the WT, five VUSs G441V, V449M, P319L, T580I and R886H led to significantly blunted response to forskolin induced CREB activation. Since signaling activity informs about gene function, this luciferase assay can be scaled up to functionally annotate ~250 *CC2D1A* VUSs currently listed in Clinvar. CREB activation is also a common downstream target of multiple NDD signaling pathways, so our paradigm can be customized for testing VUSs in other NDD genes. VUSs disrupting PDE4D activity can be prioritized for testing against PDE4D inhibitors such as Zatolmilast (in phase III clinical trial for Fragile X syndrome) to check if the deficits are rescued. Prioritized VUSs can then be introduced into iPSCs via CRISPR and their molecular effects can be modeled via differentiation of such iPSCs into induced neurons.

Session Title: Molecular Effects of Genetic Variation Poster Session III

PB1014 A novel mutation in *ITGB2* in Pakistani families causes Leukocyte Adhesion Deficiency Type I

Authors:

S. Faraz, S. Ahmed, T. Shamsi, S. Munzir; Natl. Inst. of Blood Diseases and Bone Marrow Transplantation, Karachi, Pakistan

Leukocyte adhesion deficiency I (LADI) is a type of primary immunodeficiency characterized by recurrent bacterial infections, impaired pus formation, and delayed wound healing. The genetic cause of this disorder has been linked to variations in the integrin subunit beta 2 (*ITGB2)* gene, which encodes the β-2 integrin subunit. Individuals affected by LAD1 usually display symptoms during the initial stages of their life, experiencing repeated and severe infections that can lead to death before the age of two if they do not undergo a hematopoietic stem cell transplant (HSCT). Nevertheless, detecting LAD1 is a challenging task, and even with extensive antibiotic treatment, a significant number of patients do not survive beyond their early years. To investigate the genetic basis of LADI, we performed Next-generation sequencing on two patients who were born of two consanguineous Pakistani families, which identified one novel homozygous mutation (c.778del, p. Val260CysfsTer20) in exon 7 and one previously reported (c. 186C>A, p. Cys62Ter) mutation in exon 4 of *ITGB2* gene. The patients were observed to display various clinical characteristics such as skin abscesses, delayed separation of the umbilical cord, mild omphalitis, and an increase in the number of neutrophils in their blood. These cases highlight the importance of a multidisciplinary approach to recognizing clues in patients with uncommon manifestations of a rare disease. By employing this approach, a comprehensive diagnostic evaluation for primary immunodeficiency disorders can be initiated, resulting in an enhanced understanding of the disease, appropriate patient counseling, and equipping clinicians with better tools to handle potential complications.

Session Title: Molecular Effects of Genetic Variation Poster Session I

PB1015 A novel streamlined pipeline for studying RNA editing: detect the effect of RNA editing QTLs in the non-coding genome.

Authors:

C. Dai, C. Buen Abad Najar, B. Fair, Y. Li; Univ. of Chicago, Chicago, IL

Adenosine Deaminases Acting on RNA (ADAR) mediated Adenosine-to-Inosine (A-to-I) RNA editing is a post-transcriptional modification that impacts RNA processing, such as splicing and translation. This modification has been linked to several neurodegenerative and autoimmune diseases, highlighting its potential in gene regulation and cellular functions. However, most studies to date have focused on the coding regions of the genome, overlooking non-coding regions. To address this, we developed a streamlined pipeline for identifying RNA editing from RNA-seq data. Our pipeline requires minimum configuration and accelerates researchers' analyses.

We applied our novel pipeline to the GEUVADIS total RNA-seq dataset, which comprises 89 Yoruba lymphoblastoid cell lines (LCLs), and our lab-generated chromatin-enriched RNA-seq dataset on 87 Yoruba LCLs. This enabled a comprehensive identification and comparison of editing sites in different genomic contexts.

A key highlight of our study is using chromatin-enriched RNA-seq to identify RNA editing in non-coding regions. We also developed a novel Hidden Markov Model based method that identified 7,097 non-coding RNAs (ncRNA) in LCLs. This matched ncRNA annotation approach improved the identification accuracy of editing in ncRNA. We identified 1,121,568 editing sites in the chromatin-enriched RNA-seq dataset, including over 750K intronic sites and over 150K editing sites in ncRNAs, providing an unprecedented view of the RNA editing landscape in the non-coding genome.

Our analyses revealed A-to-I editing occurs faster than splicing, with over 75% of the 42,135 editing sites shared by GEUVADIS total RNA-seq and chromatin-enriched RNA-seq data showing insignificant differences in editing levels. This finding suggests the potential role of RNA editing in modulating the transcriptome in response to swift, dynamic changes in the cellular context. However, despite the extensive editing detected, we found limited effect of RNA editing QTLs in the non-coding regions. We identified 12,404 edQTLs in chromatin-enriched RNA-seq data, with limited colocalization signals - 144 with protein-coding gene eQTLs, 201 with splicing QTLs, and 206 with ncRNA eQTLs. Our results suggest that the direct influence of non-coding RNA editing on gene expression and phenotype may be limited and nuanced.

Overall, our work presents an important methodological advancement, providing researchers with a robust and efficient tool for identifying RNA editing, and offers insights into the roles of RNA editing in non-coding regions.

Session Title: Molecular Effects of Genetic Variation Poster Session II

PB1016 A pooled-based screening strategy for re-classification of NPR2 missense variants to support the expansion of Vosoritide for genetic causes of short stature

Authors:

S. Covarrubias, C. Bauer, D. Shanghavi, D. Ruhrmund, D. Zhou, N. Marom, Y-S. Tang, S. Froelich; Biomarin Pharmaceuticals, San Rafael, CA

NPR2 is a bidirectional therapeutic target that is associated with various forms of short and tall stature. NPR2 possesses guanylyl cyclase activity that leads to synthesis of cyclic guanosine monophosphate (cGMP), and down-regulation of this pathway is responsible for short stature phenotypes. A C-type natriuretic peptide analog (CNP) is now approved by the FDA for Achondroplasia - a common form of short stature. Binding of CNP to its receptor (NPR2), triggers endochondral and skeletal growth via cGMP production. Loss-of-function mutations in NPR2 are responsible for dwarfism in mice and result in a lack or lowered intracellular cGMP response to CNP in cultured chondrocytes. While there is no treatment yet for other forms of idiopathic short stature (ISS), our goal is to treat any form of ISS indication with CNP. High throughput characterization of NPR2 variants will enable us to better predict novel variants and for those which occur more commonly could benefit diagnosis and clinical trial enrollment for eligible patients. Here we develop a pooled-based screening platform which allows us to screen ~160 variants in parallel using a cGMP-GFP-on reporter and applying a flow-based screening strategy. This new screening strategy represents a significant increase in bandwidth compared to previous methods. The screen identifies both loss-of-function (LoFs) and gain-of-function (GoF) variants and is consistent with previous arrayed-based screening data. Our data allows for the reclassification of ~31 VUS as LoF and 7 VUS as GoF variants and also allowes for re-classification of ~123 VUS as neutral. We show that these variant phenotypes enhance prediction of adult height when combined with polygenic scores and augment statistical power to detect phenotypic associations. Additionally, these data provide insights into the biological mechanisms of C-Type Natriuretic Peptide (CNP) signaling through NPR2 in skeletal development.

Session Title: Molecular Effects of Genetic Variation Poster Session III

PB1017 A proactive missense variant atlas for the Autoimmune Regulator.

Authors:

A. Axakova[1], A. Berger[2], W. v. Loggerenberg[1], N. Kishore[1], M. Gebbia[1], M. Ding[1], S. Douville[1], M. Panchalingam[1], A. Cote[1], D. Tabet[1], J. Weile[1], R. Li[1], S. Kathirgamanathan[1], E. Bratland[2], S. Johansson[3], F. P. Roth[1]; [1]Univ. of Toronto, Toronto, ON, Canada, [2]Univ. of Bergen, Bergen, Norway, [3]Haukeland Univ Hosp, Bergen, Norway

Pathogenic variants in the Autoimmune Regulator (*AIRE*) gene cause Autoimmune Polyendocrine Syndrome Type 1 (APS-1), a rare primary immunodeficiency disease with symptoms including chronic mucocutaneous candidiasis, hypoparathyroidism, and Addison's disease. AIRE increases the expression of tissue-specific genes in the niche of developing T cells, thus triggering the elimination of emerging self-reactive T-cells and preventing autoimmunity. APS-1 diagnosis by *AIRE* sequencing is increasingly common, and earlier diagnosis can benefit patients. However, two thirds of clinical variants reported are missense, and more than half of reported clinical missense variants found to be "of unknown significance". Testing variant function in cell-based assays can provide strong evidence for variant classification, but these tests are carried out reactively, often months or years after a variant's first observation within a patient. By contrast, proactively assessing the function of all possible missense variants could provide immediate evidence for genetic diagnosis, even for never-before-seen variants. Here we describe the first saturated sequence-function map of AIRE missense variants, using an insulin promoter-driven reporter assay. Our sequence-function map of AIRE, which agrees with biochemical expectations and current pathogenicity annotations, promises to provide proactive evidence with potential to improve patient outcomes by providing more rapid and definitive genetic diagnoses of APS-1.

Session Title: Molecular Effects of Genetic Variation Poster Session I

PB1018 † A single cell CRISPRi enhancer screen in microglia identifies target genes for Alzheimer's disease GWAS loci.

Authors:

A. Chesi[1], S. Laub[1], N. Tulina[1], S. Ramachandran[1], J. Pippin[2], M. Conery[1], E. Burton[3], L-S. Wang[1], G. Schellenberg[1], S. Grant[4]; [1]Univ. of Pennsylvania, Philadelphia, PA, [2]The Children's Hosp. of Philadelphia, Philadelphia, PA, [3]The Univ. of Pennsylvania, Philadelphia, PA, [4]The Children's Hosp. of Philadelphia/UPenn, Philadelphia, PA

GWAS of Alzheimer's disease (AD) have recently uncovered ~ 70 genetic loci associated with this devastating disorder. The major challenge of the post-GWAS era is mapping genetic signals to their corresponding effector genes, which is required to develop effective drug therapies. We developed a genomic approach based on high-resolution Capture C, ATAC-seq and RNA-seq to identify candidate causal variants and effector genes in brain relevant human cells (microglia, neurons, and astrocytes). We prioritized 74 candidate regulatory regions containing AD-associated SNPs by requiring them to reside in open chromatin and contact an open gene promoter. These candidate regions contacted 85 putative effector genes. To assess their effect on expression of their candidate effector genes in a high-throughput manner, we performed a pooled enhancer-targeting CRISPRi screen followed by scRNA-seq in the human microglia cell line HMC3. We performed quality control (removal of low-quality cells, doublets, and cell cycle effects) and analysis of ~ 24,000 cells using Seurat and Mixscape. Differential gene expression analyses between cells containing a single perturbation and cells containing a non-targeting guide using MAST yielded 9 hits (9 enhancer regions targeting 8 effector genes). We validated one of them, an enhancer containing three AD-associated SNPs located in an intron of *TSPAN14*, via bulk CRISPRi and luciferase assays in HMC3. Transcriptomic analyses revealed 15 downregulated and 58 upregulated genes in the enhancer knock-down condition, corresponding to several altered cellular pathways, including 'interferon alpha and beta signaling' and other pro-inflammatory pathways (upregulated), and 'SUMOylation of DNA replication proteins' and pathways related to nuclear import and the nuclear pore complex (downregulated). Further functional investigations are required to clarify the role of this gene and these cellular pathways in AD pathogenesis. We plan to expand our screens to other cell types, and we anticipate this methodology to be portable to many neurodegenerative or neurodevelopmental disease contexts.

Session Title: Molecular Effects of Genetic Variation Poster Session II

PB1019 A VUS re-analysis: Understanding the likely disease-mechanism of a de novo missense variant in RALA using public bioinformatic tools

Authors:

L. van de Wiel[1], C. Reuter[2], M. Neu[1], S. Marwaha[1], D. Bonner[2], J. Kohler[2], H. R. Mendez[1], E. Kravets[1], Undiagnosed Diseases Network, E. A. Ashley[1], S. B. Montgomery[1], J. A. Bernstein[1], M. T. Wheeler[1]; [1]Stanford Univ., Palo Alto, CA, [2]Stanford Hlth.Care, Palo Alto, CA

The interpretation of missense variation in the context of Mendelian disease genes remains a major challenge. The advent of tools specializing in data-integration from vast resources has advanced our understanding of the functional impacts of genetic variants. We present a case study utilizing these tools that demonstrates that reanalysis of putative disease missense variants can lead to a better understanding and diagnosis of genetic variation.
We evaluated a 12 yo patient assigned female sex at birth who displayed developmental delay, bilateral ptosis, feeding issues, dysmorphic features, congenital cardiac defect (PDA), recurrent pancreatitis and abnormal movements from 3 months old. Genome sequencing identified a de novo missense in RALA NM_005402.4:c.(46A>G), NP_005393.2:p.(Lys16Glu). RALA missense variants may cause AD Hiatt-Neu-Cooper syndrome [OMIM:619311]. Despite phenotypic overlap with this syndrome, the identified variant was not observed in gnomAD or ClinVar.
Utilizing HOPE (www3.cmbi.umcn.nl/hope/) - a tool that elucidates the molecular effects of single missense mutations on protein 3D structure - we analyzed p.Lys16Glu. HOPE indicated that p.Lys16's positive charge, forming a salt bridge with p.Arg65 and p.Glu87 (PDB:6P0I), is disrupted by the variant residue's negative charge. p.Lys16 is 100% conserved (HSSP) and in the Ras Family domain (PF00071). We then consulted MetaDome (www.metadome.app), which annotates pathogenic variants at equivalent homologous protein domain positions, and found p.Lys16 is equivalent to 124 human codons in homologous Ras Family domains. Assessing ClinVar pathogenic information on these codons identified three pathogenic variants in KRAS; an identical p.Lys5Glu (#12596) and two structurally similar changes: p.Lys5Asn (#12594) and p.Lys5Gln (#2010214), with literature indicating structurally similar changes affecting RAS function by increasing activation and downstream signaling. Supported by this evidence, the sequencing lab reclassified the variant from VUS to likely pathogenic. Likely, the de novo RALA variant is the molecular diagnosis for this patient. This case exemplifies how freely available data-integration tools can collaboratively unveil the functional impacts of genetic variants and disease etiology.

Session Title: Molecular Effects of Genetic Variation Poster Session III

PB1020 † Aberrant expression prediction in 48 human tissues

Authors:

F. R. Hölzlwimmer[1], J. Lindner[1], N. Wagner[1,2], V. Yepez[1], F. Casale[3,1], J. Gagneur[1,2,3]; [1]Technical Univ. of Munich, Munich, Germany, [2]Helmholtz Association - Munich Sch. for Data Sci., Munich, Germany, [3]Helmholtz Zentrum München, Munich, Germany

Identifying the functional and phenotypic impact of genetic variants is a major challenge. Although certain types of genetic variants are associated with aberrant gene expression, the actual impact of these variants in different tissues often remains unknown. Measuring aberrant gene expression in affected tissues can aid in identifying functional, large-effect variants. However, this approach is limited to clinically accessible tissues such as skin or body fluids. Here we set out to predict rare variants associated with aberrant underexpression across 48 human tissues. To achieve this, we established the first systematic benchmark for aberrant underexpression using OUTRIDER, an aberrant expression caller using RNA-seq count data, in 11,096 GTEx samples. We assessed and developed predictors that use DNA sequence and optionally RNA sequencing data from clinically accessible tissues. Notably, existing tools such as CADD (0.88% auPRC) and LOFTEE (1.43% auPRC) exhibited poor performance in predicting aberrantly underexpressed genes.

To address this issue, we developed AbExp-DNA, a nonlinear model to predict tissue-specific aberrant underexpression.

We find that weighting variant consequences by the proportion of affected isoforms per tissue as well as incorporating the tissue-specific gene dispersion provides valuable information to increase the predictive accuracy of AbExp-DNA. Including tissue-specific aberrant splicing annotations predicted with AbSplice and optionally transcript ablations caused by structural variants further increases the performance of AbExp-DNA. In total, AbExp-DNA achieves an auPRC of 9.84%, outperforming LOFTEE and CADD by up to 7-fold.

We demonstrate how AbExp-DNA can be used in rare variant gene association studies (RVAS) as well as phenotype prediction on the UK Biobank 200k WES dataset with 40 blood traits. Using AbExp-DNA, we identify 60% more trait-associated genes compared to a LOFTEE-based burden test. In addition, AbExp-DNA significantly improves phenotype prediction over LOFTEE in 45% of the traits.

Finally, we show that gene expression measurements from clinically accessible tissues can be incorporated to yield another two-fold enhancement in tissue-specific aberrant expression prediction performance. While such measurements are not directly transferable to other tissues, combining gene expression measurements from fibroblasts with AbExp-DNA scores allows AbExp-RNA to reach an overall auPRC of 20.61%.

By providing a first method benchmark of tissue-specific expression outliers, this work is relevant for rare variant association studies as well as DNA- and RNA-based rare disease diagnostics.

Session Title: Molecular Effects of Genetic Variation Poster Session I

PB1021 Aberrant splicing prediction across human tissues.

Authors:

N. Wagner[1], M. Çelik[2], F. Hölzlwimmer[1], C. Mertes[1], H. Prokisch[3], V. Yépez[1], J. Gagneur[1]; [1]Technical Univ. of Munich, Garching bei München, Germany, [2]Univ. of California Irvine, Irvine, CA, [3]Helmholtz Zentrum Muenchen, Neuherberg, Germany

Aberrant splicing is a major cause of genetic disorders but its direct detection in transcriptomes is limited to clinically accessible tissues such as skin or body fluids. While DNA-based machine learning models allow prioritizing rare variants for affecting splicing, their performance on predicting tissue-specific aberrant splicing remains unassessed. Here, we generated the first aberrant splicing benchmark dataset, by calling aberrant splicing events on 16,213 RNA-seq samples from 49 human tissues from the GTEx dataset, comprising 8.8 million rare variants in paired genotype data from 946 individuals. At 20% recall, state-of-the-art DNA-based models SpliceAI and MMSplice achieve maximum 12% precision. We observed that many false predictions originated from inaccurate genome annotations. We constructed a tissue-specific splicing map (SpliceMap) by mapping and quantifying tissue-specific splice site usage transcriptome-wide and modeling isoform competition. Using SpliceMap together with SpliceAI and MMSplice in a combined model that we call AbSplice increased precision by three-fold at the same recall. These results replicated in two independent rare disease cohorts. Integrating RNA-sequencing data of clinically accessible tissues brought precision to 60%. After publication of the method [1], we have applied AbSplice to large rare disease (>23,000 WGS from Solve-RD European project) and cancer (>4,000 WGS from Munich Leukemia Labor) cohorts. In the Solve-RD cohort, the scores are being used to validate splicing outliers from RNA-seq data, as well as suggesting disease-causal variants. In the Munich Leukemia Labor cohort, we found an enrichment of genes with high AbSplice scores among known tumor-suppressor genes and recaptured splicing aberrations that were not detected from RNA due to nonsense-mediated decay. We provide precomputed AbSplice scores for all possible single-nucleotide variants genome-wide, publicly available software to score variants including indels directly from VCF files as well as a web interface for fast lookup of variant scores. Altogether, our results substantially contribute to non-coding loss-of-function variant identification and to genetic diagnostics design and analytics.
[1] https://www.nature.com/articles/s41588-023-01373-3

Session Title: Molecular Effects of Genetic Variation Poster Session II

PB1022 African-specific, Alzheimer's disease-associated deletion in ABCA7 alters the functionality of CNS cells.

Authors:

Y. Nam[1], B. DeRosa[2], L. D. Adams[3], T. Starks[4], R. Laux[5], M. Cuccaro[6], A. Caban-Holt[7], C. Reitz[8], G. Beecham[9], J. L. Haines[10], G. Byrd[7], M. A. Pericak-Vance[2], J. Young[11], J. Vance[11], D. M. Dykxhoorn[11]; [1]Univ. of Miami, MIAMI, FL, [2]Univ. of Miami Miller Sch. of Med., Miami, FL, [3]John P. Hussman Inst. for Human Genomics, Univ. of Miami Miller Sch. of Med., MIAMI, FL, [4]Wake Forest Sch. of Med., Winston-Salem, NC, [5]Dept. of Population and Quantitative Hlth.Sci., Case Western Reserve Univ., Cleveland, OH, [6]John P. Hussman Inst. for Human Genomics, Miami, FL, [7]Maya Angelou Ctr. for Hlth.Equity, Wake Forest Univ., Winston-Salem, NC, USA, Winston-Salem, NC, [8]Columbia Univ, New York, NY, [9]Univ. of Miami, Miami, FL, [10]Case Western Reserve Univ, Cleveland, OH, [11]John P. Hussman Inst. for Human Genomics, Univ. of Miami Miller Sch. of Med., Miami, FL

Background: We previously identified a 44-base pair deletion in (ATP-binding cassette sub-family A member 7) (*ABCA7*) that is significantly associated with Alzheimer's disease (AD) in African Americans (AAs) and produces a frameshift mutation resulting in a truncated protein (p.Arg578Alafs). Previous analysis in iPSC derived cells suggested that the ABCA7 deletion alters APP processing producing increased levels of aggregation-prone Amyloid beta (Aβ) species, such as Aβ42 and Aβ40. While we have shown the mutant mRNA is present in microglia, whether it is translated into a stable protein is not known. This question has been hampered by the lack of antibodies capable of recognizing the N-terminus of endogenous ABCA7. In microglia, ABCA7 is primarily involved in the transport of lipid and amyloid-β (Aβ) clearance by phagocytosis while in neurons, ABCA7 functions in the lipid transport of and APP processing. Methods: We analyzed scRNAseq data from prefrontal cortex from AA individuals to determine *ABCA7* cell type pattern of expression. To elucidate the impact of the AA-specific deletion in protein expression and function, a recombinant version of the ABCA7 gene with and without the deletion and bearing an N-terminal flag tag was transfected into HEK cells. The transduced cells were analyzed by immunoblot and immunocytochemistry (ICC). Induced pluripotent stem cell (iPSC) lines from three unrelated AA individuals with AD who are homozygous or heterozygous for the deletion, as well as age-matched AA controls were produced and differentiated into microglia. Results: Our results showed that the ABCA7 truncated protein of ABCA7 is expressed and stable in this transfected model. Expression of *ABCA7* is highest in microglia, suggesting that iPSC derived microglia cells are a proper platform to test functionality of the truncated protein. Discussion: While the transfected cells demonstrate that the truncated protein is stable in this model, iPSC models are needed to confirm this initial finding. Our scRNAseq data indicates high expression of ABCA7 in microglia, suggesting functional relevance for microglia function that will be tested in our iPSC model.

Session Title: Molecular Effects of Genetic Variation Poster Session III

PB1023 † Allele- and tissue-specific expression across 11 tissues in baboons (*Papio anubis*)

Authors:

R. Ramasamy[1], M. Raveendran[2], A. Harris[3], J. Rogers[3], S. Panda[1]; [1]Salk Inst. for Biological Studies, La Jolla, CA, [2]BCM, Houston, TX, [3]Baylor Coll. of Med., Houston, TX

The preferential expression of one allele of a gene over the other in a diploid organism has been shown to have a profound impact on numerous biological processes. We collected 11 different tissue types from 12 wild-caught olive baboons (*Papio anubis*) and sequenced their genomes using both Illumina PE 150 reads and Oxford Nanopore long reads achieving a minimum sequencing depth of 30X and 15X, respectively. We discovered 15.8 million heterozygous biallelic single nucleotide variants (SNVs) across 12 baboons with a median of 4.7 million SNVs per baboon. Across the 12 baboons, we were able to phase 96.5 to 98.5 percent of all protein-coding genes containing biallelic heterozygous SNVs into single haplotype blocks. We sequenced the complementary DNA produced from total RNA at a median depth of 25 million Illumina paired-end 150-bp reads per sample. We investigated allele-specific expression across 187,476 unique biallelic SNVs located in untranslated and coding regions of protein-coding genes with each sample containing between 12,953 and 35,507 SNVs. Between 6.9% to 14.6% of these SNVs showed allele-specific expression characteristics across 9,753 protein-coding genes. The number of SNVs with ASE patterns was slightly influenced by gene expression heterogeneity of the tissue (r=0.36). Of the genes, 7,402 were found to have ASE characteristics in more than one animal, including 109 genes that displayed ASE in all 12 animals. We found 2,686 genes with ASE patterns that were specific to one tissue type. Among the 11 tissue types we studied, the lungs had the largest number of allelic imbalanced genes specific to that tissue type.

Session Title: Molecular Effects of Genetic Variation Poster Session I

PB1024 An analysis of how known SNPs affect ramp sequences in the human genome and their implications for human disease

Authors:

J. Wride[1], A. Nitz[1], M. Cloward[2], P. Ridge[2]; [1]Brigham Young Univ., Sandy, UT, [2]Brigham Young Univ., Provo, UT

Codon bias is the unequal distribution of synonymous codons across a gene or genome. Codon choice is subject to selection and is non-random. Codon choice is highly correlated with the most abundant tRNA. This is energetically desirable because the anticodons more efficiently bind to the codon during translation, unlike codons that do not bind in the wobble position. Codons may be labeled as efficient (i.e., their cognate tRNA is highly expressed) and translated quickly, or inefficient and translated more slowly. Groups of consecutive, inefficient codons appear in many genes. A ramp sequence is 30-50 consecutive inefficient codons at the 5' end of the coding region. We previously demonstrated that the presence or absence of a ramp sequence (i.e., ramp status) in a gene is highly conserved. While the exact role of ramp sequences is unknown, a common hypothesis is that the stretch of slowly translated codons spaces ribosomes along the transcript to reduce the number of ribosomal collisions. Ramp sequences are more common in highly expressed genes, and when an efficient (even synonymous) codon is substituted in the ramp sequence, the ramp sequence is destroyed and total protein reduced. This project aimed to determine how many known human SNPs change a gene's ramp status.

We compiled a list of 14 million SNPs in coding regions. We used our published software, ExtRamps, to compare the wild-type transcript to the mutant transcript for each SNP to determine if a SNP destroys an existing, or creates a new, ramp sequence. 5,133 total SNPs are located in 2,865 of the 5,377 human genes with ramp sequences. Most of these SNPs do not change ramp status. However, we identified 879 SNPs in 683 genes that destroy an existing ramp sequence and 897 SNPs in 761 genes that create a ramp sequence.

Changes in ramp status could have broad implications for understanding the functional effects of SNPs. While many synonymous SNPs are likely benign, we have few tools for predicting when synonymous changes could be functional. Moreover, 500,000 GWAS SNPs have been published, most of which have no known function. Changes to ramp sequences provide a plausible explanation for effects caused by synonymous changes.

Session Title: Molecular Effects of Genetic Variation Poster Session II

PB1025 † An atlas of *trans*- genetic effects across human tissues.

Authors:

J. Li, X. Liu; Univ. of Chicago, Chicago, IL

A deep understanding of the genetic basis of gene regulation and complex traits critically relies on comprehensive, high-quality maps of *cis*- and *trans*- regulatory effects (Liu et al. 2019 *Cell*; The GTEx Consortium 2020 *Science*; Võsa et al. 2021 *Nat Genet*). While most studies focus on mapping *cis*- genetic effects, there is a scarcity of high quality *trans*-eQTLs, due to their small effect sizes and statistical challenges. In this study, we aim to build the most comprehensive maps of *trans*-eQTLs in 22 human tissues and cell types, by using an ensemble of powerful statistical methods, including *trans*-PCO (Wang et al. 2022 *bioRxiv*), GBAT (Liu et al. 2020 *Genome Biol*) and the standard univariate method. We first applied *trans*-PCO, which is a method to identify *trans*-eQTLs of co-expressed gene modules, to the GTEx dataset. We found a total of 6,570 significant *trans* eSNP-module pairs at 10% FDR in 22 GTEx tissues and cell types, which are highly replicable across datasets. For example, 73.2% of *trans*-eQTLs in GTEx whole blood (N = 670) were replicated in another blood dataset (DGN, N = 913; Bonferroni corrected $p < 0.05$). Among all the *trans*-eSNPs, 64.9% of them overlapped with either *cis*-eQTLs or *cis*-sQTLs in the corresponding tissue, suggesting that trans regulations are primarily mediated through *cis*- regulation. To investigate how the *trans*-eQTL maps can explain mechanisms of diseases, we performed colocalization analyses of *trans*-eQTLs and genome-wide association (GWAS) loci from 37 complex traits, and identified a total of 533 significant colocalization signals (PP4 > 0.5). We found that *trans*-eQTLs tend to colocalize with GWAS loci in trait relevant tissues. For example, *trans*-eQTLs in the whole blood had the highest proportion of colocalization (62%) among 29 blood traits and 3 autoimmune diseases, while other tissues showed a much lower proportion of colocalization (< 30%). In addition, the gene modules associated with the colocalized *trans*-eQTLs showed relevant biological functions to the corresponding traits. For example, we identified a colocalization between a GWAS locus of platelet count near *ARHGEF3* and a *trans*-eQTL on chromosome 3 (PP4 = 0.67). The gene module associated with the colocalized *trans*-eQTL not only exhibited enriched functions of platelet activation and hemostasis, but also had a strong heritability enrichment for platelet count (3.1 fold; Bonferroni corrected $p = 1.6 \times 10^{-5}$). In conclusion, our study will produce comprehensive maps of *trans*-eQTLs in human tissues and cell types, which will characterize *trans*- regulatory effects and mechanism, as well as shed light on detailed molecular mechanisms of complex traits.

Session Title: Molecular Effects of Genetic Variation Poster Session III

PB1026 An *in silico* analysis of*rpoB*mutations to affect*Chlamydia pneumoniae*sensitivity to rifamycin

Authors:

F. Radouani[1], I. Benamri[1,2], A. Moussa[3]; [1]Inst. Pasteur du Maroc, Casablanca, Morocco, [2]Natl. Sch. of Applied Sci., Abdelmalek Essaâdi Univ., Tangier, Morocco, Tangier, Morocco, [3]Natl. Sch. of Applied Sci., Abdelmalek Essaâdi Univ., Tangier, Morocco

Chlamydia pneumoniae is a common cause of human respiratory diseases and most commonly manifests as pneumoniae and bronchitis. Furthermore, Chlamydia pneumoniae associated with other acute and chronic respiratory conditions such as: chronic obstructive pulmonary disease, asthma, and lung cancer. In addition, it is associated with atherosclerotic cardiovascular disease. Antibiotics are the commonly used drugs to treat chlamydiae infections. However, when overused or mise used this may lead to strains' resistance to antibiotics, this phenomenon represents a real health problem worldwide. Numerous studies showed the association of *Chlamydia pneumoniae* resistance with mutations in different genes. The aim of this study is to perform an *in silico* analysis of *C. pneumoniae rpoB* encoded proteins using numerous bioinformatics tools and to identify the functional and structural related effects of the mutations and consequently their impact on the bacteria sensitivity to antibiotics.The analysis revealed that the prediction of the damaging impact related to the mutations in *rpoB* encoded proteins showed eight mutations: R421S, F450S, S454F, L456I, D461E, S476F, L478S, and S519Y with big deleterious effects, and they are located in a highly conserved regions decreasing the protein's stability. Furthermore, the structures analysis showed that the mutations models had a deviation compared to the wild type. Moreover, the prediction of protein-protein network indicated that *rpoB* wild type interacts strongly with 10 proteins of *C. pneumoniae*, which are playing different roles at different levels.The present study revealed that the mutations in the encoded proteins can affect their functions and structures, in addition to their interactions with other proteins which impact the bacteria sensitivity to antibiotics. Consequently, the information revealed through this in silico analysis would be useful for deeper exploration to understand the mechanisms of *C. pneumoniae* resistance and enable managing the infection to avoid its complications. We recommend further investigations and perform deeper experimental analysis with collaboration between bioinformaticians, physicians, biologists, pharmacists, and chemistry and biochemistry scientists.

Session Title: Molecular Effects of Genetic Variation Poster Session I

PB1027 An integrative strategy to link regulatory elements from disease GWAS to genes using single-cell multiome data.

Authors:

E. Dorans[1,2], K. A. Jagadeesh[1,3], K. K. Dey[4], A. L. Price[1,3,5]; [1]Dept. of Epidemiology, Harvard Sch. of Publ. Hlth., Boston, MA, [2]PhD Program in Biological and BioMed. Sci., Harvard Med. Sch., Boston, MA, [3]Program in Med. and Population Genetics, Broad Inst. of MIT and Harvard, Cambridge, MA, [4]Sloan Kettering Inst., Mem. Sloan Kettering Cancer Ctr., New York, NY, [5]Dept. of Biostatistics, Harvard Sch. of Publ. Hlth., Boston, MA

Single-cell paired RNA and ATAC multiome data have shown great promise in linking regulatory elements to genes through the synchronized activity of ATAC-seq peaks and expression of nearby genes. However, existing methods (e.g. Signac, Cicero, ArchR, SCENT) differ in how they account for biological and technical noise in single-cell data as well as in their modeling assumptions; as a result, we observe low concordance in their linking scores (e.g. $r = 0.20$ for Signac vs. Cicero) and the subset of candidate links that they are able to score.

Here, we propose pgBoost, an integrative modeling framework that learns a non-linear combination of existing linking strategies to assign a probabilistic score to each peak-gene pair consisting of an ATAC-Seq peak and a nearby gene. pgBoost optimizes a gradient boosting classification task using peak-gene linking scores from Signac, Cicero, ArchR, and SCENT as features and a training set comprised of fine-mapped eSNP-eGene pairs with posterior probability (PIP) > 0.2 as positives and eSNP-eGene pairs with PIP < 0.0001 as negatives (taking the within-peak eSNP with maximum PIP for each peak-gene link). To avoid overfitting, we train the model holding out one chromosome at a time and predict peak-gene links on held-out chromosomes.

We applied pgBoost to multiome data from 10,000 PBMC cells (10x Genomics) using fine-mapped eQTL from GTEx whole blood for training. We reached 3 main conclusions. First, pgBoost attained a higher AUPRC on a held-out test set of fine-mapped eSNP-eGene pairs when trained on all 4 linking scores (0.72) than when trained on individual scores (0.52-0.64). Second, pgBoost attained a higher enrichment for the held-out test set of fine-mapped eSNP-eGene pairs (17.4±0.6x) than the constituent methods (3.0-14x); contributing to this improvement was the higher proportion of within-peak eSNPs scored by pgBoost (99%) than the constituent methods (36-97%). Third, when applied to a "silver-standard" set of 1,522 non-coding SNP-gene pairs obtained from fine-mapped variants for 12 blood-related diseases and traits with a unique fine-mapped coding variant within a 2Mb window (Weeks et al. medrxiv), pgBoost attained a SNP-gene link enrichment of 7.6±0.4x, compared to 2.3-10x for the constituent methods. Our results suggest that an optimal non-linear combination of existing peak-gene linking methods improves power to link regulatory variants to genes, supported by benchmarking analyses using both expression QTL and disease GWAS data.

Session Title: Molecular Effects of Genetic Variation Poster Session II

PB1028 Ancestry-specific changes in the genomic regulatory architecture of iPSC derived oligodendrocyte enriched neural spheroids in the light of Alzheimer's disease.

Authors:

A. Ramirez[1], J. J. Shepherd[1], L. E. Coombs[1], S. Simon[1], S. Moura[1], B. DeRosa[2], F. Rajabli[1], P. Whitehead[1], L. Adams[1], T. Starks[3], P. Mena[1], M. Illanes-Manrique[4], S. Tejada[1], G. Byrd[5], M. Cornejo-Olivas[6], B. Feliciano-Astacio[7], L. Wang[1], F. Jin[8], W. Xu[9], A. Griswold[1], J. Young[1], D. Dykxhoorn[1], J. Vance[1]; [1]Univ. of Miami, Miami, FL, [2]Univ. of Miami Miller Sch. of Med., Miami, FL, [3]Wake Forest Sch. of Med., Winston-Salem, NC, [4]Global Brain Hlth.Inst. (GBHI), San Francisco, CA, [5]NC A&T State Univ, Greensboro, NC, [6]Inst. Natl. de Ciencias Neurologicas, Lima, Peru, [7]Univ. Central del Caribe, Bayamon, PR, [8]CASE WESTERN RESERVE Univ., SOLON, OH, OH, [9]Case Western Resreve Univ, Cleveland, OH

BACKGROUND The impact of genetic risk variants associated with developing late onset Alzheimer's disease (AD) can be influenced by a person's ancestry via changes in the genomic regulatory architecture (GRA). To comprehensively interpret the variability associated with AD risk genes across populations, it is crucial to have a global understanding of GRA in the context of AD. Most of the currently available GRA data has been generated from individuals of European ancestry. In this study, we aimed to establish ancestry-dependent differences in the GRA using induced pluripotent cell (iPSC) derived neural spheroids with African, Amerindian, and European ancestral backgrounds. Since GRA is specific to different cell types, we focused on the oligodendroglia generated from the spheroids given the abundance of these cells in the central nervous system and the limited attention they have received in AD research. **METHODS:** We obtained cells from individuals with AD or without cognitive impairment, with >85% global ancestry of each ancestral background. These cells were reprogrammed into iPSC and then differentiated into oligodendrocyte (OL) enriched neural spheroids using a protocol that promotes oligodendroglia development. On day 76 of differentiation, we collected and lysed the cells to isolate nuclei for multiomic profiling. This involved analyzing chromatin accessibility and transcriptome using Single Cell ATAC and Single Cell RNA-seq techniques. In addition, we investigated chromatin interactions using Hi-C analysis. **RESULTS:** We identified various stages of OL lineage cells, ranging from cycling progenitors and OL precursor cells to fully mature myelinating OLs. We compared the clusters involving these cell types across ancestries, distinguishing cases from controls and considering APOE genotypes. Preliminary data shows ancestry-dependent differential gene expression of several genes in mature oligodendrocytes including AD GWAS hits such as *APOE, SORL1*, *CLU*, *PICALM*, *ANK1* and *PLCG2*. Analyses of other oligodendroglia clusters, astrocytes and neurons are still ongoing. **CONCLUSION:** Our findings provide ancestry-specific insights into the chromatin structure and gene regulation of oligodendrocytes in the context of AD. This comprehensive understanding contributes to our knowledge of a previously overlooked cell lineage in AD. These results expand the available functional resources for gene identification studies in African American and Hispanic/Latino populations.

Session Title: Molecular Effects of Genetic Variation Poster Session III

PB1029 Annotation Query (AnnoQ): An Integrated and Interactive Platform for Comprehensive Genetic Variant Annotation on a Large Scale

Authors:

B. Queme[1], T. Mushayahama[1], D. Ebert[1], A. Muruganujan[1], P. D. Thomas[1], H. Mi[2]; [1]Univ. of Southern California, Los Angeles, CA, [2]Univ. of Southern California, Irvine, CA

The Annotation Query (AnnoQ) system (http://annoq.org/) serves as a comprehensive, dynamic platform providing the most recent functional annotations for human genetic variants. Supported by an extensive annotation database, it encompasses approximately 39 million human variants from the Haplotype Reference Consortium (HRC), all pre-annotated with over 600 distinct annotation types sourced from various resources. AnnoQ stands out due to its user-friendly and efficient interactive query tool. Unlike traditional annotation tools, which require the user to set up and execute queries, AnnoQ has already pre-annotated all possible variants, thus operating as an intuitive search tool. With its advanced search framework and API, it facilitates efficient management of large datasets, serving the needs of researchers from diverse computational and bioinformatics backgrounds. Moreover, the system features an interactive interface that allows users to conduct quick queries and review annotation results before initiating extensive analyses. AnnoQ ensures that researchers have access to the latest functional annotations by maintaining strong collaborations with renowned resources such as the Gene Ontology Consortium, Reactome, and PANTHER. By incorporating updates from these partners consistently, it guarantees that the annotation information available is always current. An additional noteworthy feature is that AnnoQ extends functional annotations to variants located in the non-coding regions. By utilizing PEREGRINE, a comprehensive genome-wide enhance-gene link database, AnnoQ associates variants in non-coding enhancer regions with their target genes, functions, and pathways. This inclusion expands the annotation scope, covering non-coding regions as well. The AnnoQ system has versatile search capabilities, thanks to its optimized Elasticsearch framework. It supports real-time complex searches, an array of search types including chromosome range, gene product IDs, variants present in VCF files, rsID lists, and free text search. To top it off, AnnoQ offers an API that enables programmatic access to annotated data. Users can avail software packages tailored for popular programming languages like R, which allows them to integrate annotation queries directly within their scripts. In conclusion, AnnoQ operates as a thorough annotation platform that caters to the diverse needs of the research community. By providing detailed and up-to-date functional annotations for human genetic variants, AnnoQ enables researchers to efficiently explore and analyze genomic data.

Session Title: Molecular Effects of Genetic Variation Poster Session I

PB1030 Application of deep mutational scanning data for MLH1 variant interpretation

Authors:

A. Scott[1], F. Hernandez[2], S. Belhadj[2], A. Marsh[2], M. Varga[2], X. Jia[3], G. Clark[3], S. Jayakody[3], B. Burugula[3], R. Karam[2], J. Kitzman[3]; [1]Univ. of Michigan Hlth.Systems, Ann Arbor, MI, [2]Ambry Genetics, Aliso Viejo, CA, [3]Univ. of Michigan, Ann Arbor, MI

Pathogenic germline variants in the DNA mismatch repair (MMR) pathway cause Lynch Syndrome (LS), a hereditary cancer predisposition syndrome affecting more than 1 in every 300 individuals worldwide. Effectiveness of LS genetic counseling is limited by the prevalence of variants of uncertain significance (VUS), which comprise the majority of missense variants identified by clinical genetic testing. We have established deep mutational scanning (DMS) as a scalable means for functional testing to support accurate variant interpretation in LS (Jia et al, *AJHG*, 2021; Scott et al, *Genome Biol*, 2022), which we now apply to the key gene *MLH1*. We overlaid the results of an *MLH1* DMS on clinical databases comprising >15,000 individuals with MMR gene variants from a clinical genetic testing laboratory. In order to determine their applicability to patients, we first applied these results to *MLH1* germline missense variants previously classified as Pathogenic (N=23) or Benign (N=27). All variants which exhibited normal function in this screen had a benign classification, excluding one variant (c.1517T>C; p.V506A) for which the measured effect was intermediate. Conversely, most variants with abnormal function in our DMS data were previously classified as pathogenic or likely pathogenic, such that this function map provides strong evidence under the OddsPath framework (ClinGen Sequence Variant Interpretation Working Group, Tavtigian et al, *Genet Med* 2018). This cohort also included 590 VUS missense variants in *MLH1,* of which a majority (78%) scored in the normal range, consistent with incidentally discovered, benign rare variants unrelated to individual cancer history. By contrast, 12.4% of the clinical missense VUSs exhibited loss of function at the protein level and 5% were predicted to disrupt splicing: reclassification using the functional evidence for these 80 VUSs is ongoing. Saturation-scale functional testing via DMS can provide badly needed functional evidence to improve the actionability of genetic testing.

Session Title: Molecular Effects of Genetic Variation Poster Session II

PB1031 Assessing the functional impact of *ASL* missense variants and genetic interactions between them using high throughput yeast assays

Authors:

M. Tang[1], G. Cromie[1], R. Lo[1], M. Timour[1], H. Morizono[2,3], L. Caldovic[2,3], N. Ah Mew[2,3], A. Gropman[2,3,4,5], A. Dudley[1]; [1]Pacific Northwest Res. Inst., Seattle, WA, [2]Ctr. for Genetic Med. Res., Children's Natl. Res. Inst., Children's Natl. Hosp., Washington, DC, [3]Dept. of Genomics and Precision Med., Sch. of Med. and Hlth.Sci., The George Washington Univ., Washington, DC, [4]Dept. of Neurology, Div. of Neurogenetics and Neurodevelopmental Disabilities, Children's Natl. Hosp., Washington, Washington, DC, [5]Ctr. for NeuroSci. Res., Children's Natl. Res. Inst., Children's Natl. Hosp., Washington, DC

Argininosuccinate lyase (ASL) deficiency is the second most common urea cycle disorder. Clinical presentation of ASL deficiency is variable with onset ranging from the neonatal period to later stages of life. Compared to other urea cycle disorders, individuals with ASL deficiency also experience a higher rate of neurological complications, even in the absence of hyperammonemic events. While disease severity generally correlates with enzyme activity, cases of asymptomatic individuals with undetectable ASL activity and affected individuals with high ASL activity exhibiting severe neurocognitive deficits have been observed. Clinical and genetic heterogeneity of the disease is further confounded by phenomenon of intragenic complementation, in which the combination of two variants of the same protein have activity significantly greater than what is expected based on each allele's individual activity. To better understand the molecular basis of the phenotypic variability in ASL deficiency, we developed and performed high throughput yeast assays to assess the functional impact of all single nucleotide variant (SNV)-accessible missense variants in *ASL*. The assay leverages the ability of the human *ASL* to functionally replace deletion of its yeast ortholog, *ARG4*, enabling yeast strains to grow on arginine-deficient medium. Colony growth of yeast strains harboring *ASL* variants as the sole source of the enzyme provides a quantitative readout of variant enzyme function. We estimated the functional impact of nearly 2,400 *ASL* missense variants (88% of all possible SNV-accessible variants). Approximately 20% of the tested variants displayed null growth, with 65 variants mapping to active and substrate binding sites and 111 to tetramer interface residues. We identified 25 amino acid positions beyond the active sites and tetramer interfaces that exhibited null growth in at least half of the variants tested, suggesting their importance for ASL function. In addition to measuring the functional impact of individual *ASL* missense variants, we will present results of compound heterozygous genotypes to determine the extent to which intragenic complementation can occur. Our high-throughput yeast-based assays provide a comprehensive functional assessment of *ASL* missense variants and allow us to test different *ASL* allelic combinations. Our analyses of the effects of individual mutations as well as compound heterozygotes will help gain a better understanding of disease manifestation and genotype-phenotype correlation in ASL.

Session Title: Molecular Effects of Genetic Variation Poster Session III

PB1032 Assessing the relative power of constraint metrics in variant prioritisation for rare-variant association testing

Authors:

B. Hill[1], N. Baya[1], D. Palmer[2], C. Lindgren[3]; [1]Univ. of Oxford, Oxford, United Kingdom, [2]Univ. of Oxford, Big Data Inst., Oxford, United Kingdom, [3]Big Data Inst., Oxford Univ., Oxford, United Kingdom

Recent rare-variant association studies have unveiled numerous gene-phenotype associations within the human exome. These associations are enabled by set-based testing applied to prioritised collections of variants with evidence of pathogenicity. As such, our power to detect significant associations is limited by our ability to accurately predict such damaging variations. Additionally, within recent years machine learning models of the distribution of aligned protein sequence variation and utilising evolutionary constraint has enabled large advances in protein variant pathogenicity prediction.

We aim to map missense variants in the UK Biobank onto their canonical protein sequences to then apply protein variant effect predictors such as EVE and Tranception to classify damaging missense variants.

First, we evaluated the performance of these classifiers by comparing the number of significant burden tests across selected phenotypes in the UK Biobank 450K exome sequence dataset. Second, we compared the overlapping annotation sets with the existing ClinVar pathogenic labeled missense variants. During this analysis, we controlled score thresholds to fix the annotation count across the analysed prediction metrics.

When compared to the state-of-the-art annotations, EVE and Tranception successfully replicated almost all existing associations and increased the total number of associations across a wide range of phenotypes. The precision of both EVE and Tranception, when compared against ClinVar, was 68% and 57% respectively, significantly outperforming the 47% precision achieved by using PP3 moderate and/or SpliceAI > 0.2.

Our findings demonstrate that the use of protein fitness scores can significantly improve the prediction accuracy of damaging missense variants, enabling more powerful gene association tests. We anticipate that the continued development of methods integrating protein structural prediction into variant classification will further enhance accuracy and boost our ability to detect gene-phenotype associations.

Session Title: Molecular Effects of Genetic Variation Poster Session I

PB1033 *ATXN2* knockdown in zebrafish leads to phenotypic features supporting a role for Ataxin-2 in primary open-angle glaucoma pathogenesis

Authors:

S. Rong[1], A. Larson[1], J. Wiggs[2]; [1]Mass Eye and Ear, Mass Gen. Brigham, Harvard Med. Sch., Boston, MA, [2]Harvard Med Sch, MEEI, Boston, MA

Purpose: Primary open-angle glaucoma (POAG) is the leading cause of irreversible blindness globally. Genome-wide association studies have identified significant POAG association at a chromosome 12 locus, which includes *ATXN2,* a gene known to contribute to other neurodegenerative disorders including spinal cerebellar ataxia 2 and amyotrophic lateral sclerosis. To provide additional support for a role for *ATXN2* in glaucoma, we created and characterized an atxn2-knockdown (KD) zebrafish line. **Methods**: We used in situ hybridization to locate the expression of *atxn2* in zebrafish embryos. Subsequently, *atxn2* was knocked down using a CRISPR/Cas9 system and a translation-blocking morpholino (TBMO). We further purified a line carrying an *atxn2* 58-bp deletion (c.58_115del) from the F0 edited fish. Eye size, retinal ganglion cell (RGC) count, visual motor response (VMR), and intraocular pressure (IOP) were compared between the KD fish and controls at different developmental stages. We used the KD fish for a complementation assay to test a selected group of rare (MAF < 1%) ATXN2 coding variants (CADD >15) with at least nominal evidence for POAG association ($p< 0.05$) for the rescue of eye size and VMR at 5dpf **Results**: In situ hybridization showed atxn2 expression in the RGCs in developing zebrafish ($P<0.05$). The *atxn2*-KD fish line embryos exhibited a significant reduction in eye size ($P\leq0.011$), RGC counts ($P\leq0.004$), and VMR ($P\leq0.009$) at 5dpf. These phenotypes were successfully replicated in the TBMO morphants ($P\leq0.027$). Moreover, IOP was also elevated in the adult *atxn2*-KD fish line ($P=0.01$). Complementation assays using the KD line demonstrated that selected rare, POAG-associated *ATXN2* missense variants reduced eye size ($P<0.001$) and resulted in impaired VMR ($P<0.05$). **Conclusions**: This study identifies glaucoma-related phenotypic features in an *ATXN2* knockdown zebrafish model supporting a role for *ATXN2* in glaucoma pathogenesis.

Session Title: Molecular Effects of Genetic Variation Poster Session II

PB1034 Australia's largest primary ciliary dyskinesia cohort: exploring the genetic background

Authors:

E. Maas[1], M. Marshall[2], P. Leo[3], A. Cook[4], K. Gantert[5], L. Morgan[6], A. Change[4], E. Duncan[7], A. McIneryney-Leo[8]; [1]The Univ. of Queensland, Brisbane, Australia, [2]Australian Translational Genomics Ctr., Queensland Univ. of Technology,, Brisbane, Australia, [3]Australian Translational Genomics Ctr., Queensland Univ. of Technology, Brisbane, Australia, [4]Australian Ctr. for Hlth.Services Innovation, Queensland Univ. of Technology, Brisbane, Australia, [5]Dept. of Hlth.and Ageing, Concord Repatriation Gen. Hosp.,, Sydney, Australia, [6]Dept. of Hlth.and Ageing, Concord Repatriation Gen. Hosp., Sydney, Australia, [7]Dept. of Endocrinology, Guy's and St Thomas' NHS Fndn. Trust, London, United Kingdom, [8]Frazer Inst., The Univ. of Queensland, Dermatology Res. Ctr., Brisbane, Australia

Background: Primary ciliary dyskinesia (PCD) is a genetically heterogenous, autosomal/X-linked recessive disease. ~50% PCD patients have homozygous/compound heterozygous (CompHet) pathogenic variants in one of 72 known PCD genes.
Methods: Australia's largest PCD cohort to date (n=56 unrelated; n=4 siblings) underwent whole exome sequencing (WES), to identify causal variants using a virtual gene panel comprising known PCD and ciliopathy genes (316 genes). Filtering WES data retained good quality, nonsynonymous, rare variants (minor allele frequency <0.005).
Results: 21/60 individuals carried n=35 rare, pathogenic/likely pathogenic/variants-of-unknown-significance in 11 PCD genes (n=26/35 were pathogenic/likely pathogenic, n=21/35 were loss-of-function (LOF)). Of these, n=3 homozygotes (*RPGR*-novel, *DYX1C1* and *OFD1*-novel), and n=17 CompHet for which n=7 carried 1x novel and 1x previously reported variant, and n=2 carried novel LOF variants. While *DNAH5* accounts for the greatest number of reported international cases in literature, *DNAH11* was greatest for this Australian cohort (n=7, two were siblings). Eleven individuals carried a single rare LOF and/or pathogenic/likely pathogenic variant in a known PCD gene, but a second variant could not be identified. Expanding the filtering to include known ciliopathy genes (not associated with PCD), identified n=21 rare, variants of likely functional significance across nine genes in 10/39 individuals. *CFTR* was the most frequently mutated gene in this subset (n=7 variants across n=3 individuals) though there were no cases of previously reported biallelic *CFTR* variants.
Conclusion: This study of an Australian PCD population identified novel/rare variants in known PCD and ciliopathy genes. Further elucidation of genes involved in disease pathogenesis may inform the development of targeted therapies.

Session Title: Molecular Effects of Genetic Variation Poster Session III

PB1035 Automated High-throughput System to classify pathogenic*SLC6A1*variants.

Authors:

A. Seminario[1], M. Trinidad[2], A. Wuster[3], G. Berguig[2], W. Wallace[4], J. LeBowitz[2], K. Estrada[5], S. Froelich[6]; [1]BioMarin Pharmaceutical, Greenbrae, CA, [2]BioMarin, San Rafael, CA, [3]BioMarin, San Anselmo, CA, [4]BioMarin Pharmaceuticals, San Rafael, CA, [5]BioMarin, South San Francisco, CA, [6]BIOMARIN, San Rafael, CA

Protein-coding variants in SLC6A1 are associated with early onset epilepsy, autism, and schizophrenia. SLC6A1 encodes the GABA transporter GAT1 and is expressed at the synapses of GABAergic neurons. The primary function of neuronal GAT1 is to re-uptake the inhibitory neurotransmitter GABA from the synaptic cleft to sustain its synaptic concentrations. Recently, variants in SLC6A1 have been associated with a spectrum of epilepsy syndromes and neurodevelopmental disorders. Pediatric patients present with seizures, neurodevelopmental delay, autism-like behaviors, and other neurological and developmental symptoms. The average age of onset of seizures is 3.7 years. De novo mutations in SLC6A1 have also been associated with schizophrenia. Schizophrenia patients with SLC6A1 mutations are typically diagnosed in early adulthood and do not present with the symptoms seen in pediatric patients.

Although advances in genomic sequencing have enabled aggregation of patient variants and observed clinical manifestations, the functional impact of many of the disease associated variants remains unknown. Experimental characterization remains a valuable method for validating the impact of variants. To speed up experimental characterization of these SLC6A1 missense-variants, we built a system composed of a MICROLAB STAR liquid-handling robot (Hamilton Robotics, Reno, NV), a 96-well Shuttle nucleofector (Lonza, Basel, Switzerland). Both instruments were controlled by a single PC through Hamilton Venus software. The use of this automated system allowed researchers to screen 180+ rare SLC6A1 missense-variants in four months rather than one year.

Samples were analyzed downstream by mass spectrometry to assess GAT1 transporter activity. Our results find this method to be highly concordant with variant activities previously reported in the literature. These results enable us to classify variants based on activity to generate a more accurate estimate of the prevalence of SLC6A1-associated disorders, and to better understand the molecular biology of SLC6A1-associated conditions. The acquired knowledge will help guide therapeutic decisions and enable the development of targeted therapies that enhance transporter function and improve symptoms.

Session Title: Molecular Effects of Genetic Variation Poster Session I

PB1036 Automatic detection of complex genome structural variation across human populations and PsychENCODE brains implicate their associations with neuropsychiatric disorders

Authors:

H. Guo[1], B. Zhou[1], J. Arthur[1], T. Kim[2], Y. Huang[1], R. Pattni[1], H. Lee[1], H. Ji[1], G. Song[2], D. Palejev[3], X. Zhu[4], W. Wong[1], A. Urban[1]; [1]Stanford Univ., Stanford, CA, [2]Pusan Natl. Univ., Busan, Korea, Republic of, [3]Bulgarian Academy of Sci., Sofia, Bulgaria, [4]The Pennsylvania State Univ., University Park, PA

Complex structural variation (cxSVs) are major sources of natural human genetic variation, but their genome biology is not well characterized due to difficulties in detecting them from individual genomes. To address this, we developed Automated Reconstruction of Complex Variations (ARC-SV) which combines probabilistic modeling with machine learning to enable the accurate detection of cxSVs from standard whole-genome sequencing (WGS) at unprecedented accuracy, allowing for cxSVs detection on a population-scale. Applying ARC-SV to 4,262 human genomes from all super-populations, we identified and catalogued 8,493 cxSVs belonging to more than 12 subclasses. Overall, cxSVs are significantly enriched in regions prone to recombination and germline *de novo* mutations. Interestingly, rare cxSVs show significant enrichment in genes of neural processes and loci undergoing rapid evolution (including recently evolved cis-regulatory regions relevant to human corticogenesis). Furthermore, by integrating ARC-SV with single-cell multiome analysis for 40 PsychENCODE brains, we found strong colocalized effects of cxSVs on gene expression and on psychiatric disorders. We also identified significant associations of cxSV with gene expression and with chromatin accessibility. Our study paves the way for the automatic integration of cxSVs detection for future population-scale WGS studies and provides valuable insights into the genome biology of cxSVs in human disease.

Session Title: Molecular Effects of Genetic Variation Poster Session II

PB1037 Body mass index has a broad causal effect on fecal 16S microbiome and plasma metabolome variation.

Authors:

D. Hughes[1], L. J. Corbin[2], C. Hatcher[1], K. H. Wade[1], J. Raes[3], N. J. Timpson[4]; [1]Univ. of Bristol, Bristol, United Kingdom, [2]Univ Bristol, Bristol, United Kingdom, [3]KU Leuven–Univ. of Leuven, Leuven, Belgium, [4]Bristol Univ., Bristol, United Kingdom

Variation in the accumulation of body fat, body composition, and obesity can all be proxied by body mass index (BMI) which itself is a recognized risk factor for numerous health outcomes including life expectancy, various cancers, and cardiometabolic diseases. What remains unclear is the complete picture of how BMI influences health. Two possible factors are variation in the microbiome and metabolome, which previous research has demonstrated to be strongly correlated with BMI. Here, we use genetic contributions to BMI from the Flemish Gut Flora Project (n=2257) to estimate causal associations between BMI and (1) gut flora and (2) the plasma metabolome. Undertaking one-sample Mendelian randomization (MR) analysis, using a polygenic score reliably associated with BMI, yielded evidence from total and sex-specific sub-analysis suggesting the presence of causal effects of BMI on gut flora and the metabolome. In total, 1058 metabolites from the Metabolon platform and 215 microbiome traits (MTs), including three alpha-diversity, one beta-diversity, 115 presence (vs. absence), 40 zero-truncated abundance, and 51 abundance traits, were included in all analyses. After correcting for multiple testing, 495 metabolites and 67 MTs were associated with a standard deviation change in BMI (~4.5kg/m2) in (generalized) linear modeling. MR-derived effect estimates were strongly correlated with estimates from (generalized) linear models. After correcting for multiple testing, 55 metabolites and 14 MTs retained evidence for a strong causal association with BMI in the total population. This includes 24 lipids and 13 amino acids such as alanine, valine, and isoleucine, and it includes genera Sporobacter and Barnesiella, family Porphyromonadaceae, beta-diversity MDS1, two alpha diversity metrics, and the enterotype Bact2. These 55 metabolites and 14 MTs yield a BMI driven network of 327 metabolite-MT connections including all 14 MTs and 44 metabolites. The strongest MT-metabolite correlation is between beta-diversity (MDS1) and 3-phenyl-propionate - a xenobiotic - with a Spearman's rho of 0.54. We might conclude that microbiome diversity drives variation in 3-phenyl-propionate but at present it is equally likely that BMI is having independent but correlated effects on microbiome and metabolome variation. Overall, results support a conclusion that the plasma metabolome and gut microbiome variation can be causally influenced by variation in BMI and that while BMI driven microbiome and metabolome variation is intercorrelated caution is warranted when making further inferences of causality.

Session Title: Molecular Effects of Genetic Variation Poster Session III

PB1038 Brain development mutations in the beta-tubulin TUBB result in defective ciliogenesis

Authors:

A. Mollica[1], S. Omer[2], S. L. Evagelou[1], S. Naumenko[3], S. Walker[4], L. Li[1], A. Teeling[1], K. Lindsay[1], S. Erwood[1], A. Vig[1], R. M. Vernon[1], J. D. Forman-Kay[1], M. Shroff[1], R. E. Harrison[2], R. D. Cohn[1], E. A. Ivakine[1]; [1]The Hosp. for Sick Children, Toronto, ON, Canada, [2]Univ. of Toronto Scarborough, Toronto, ON, Canada, [3]Harvard Chan Sch. of Publ. Hlth., Boston, MA, [4]Genomics England, London, United Kingdom

Tubulinopathies and neurodevelopmental ciliopathies are two groups of genetic disorders characterized by abnormal brain development resulting in structural brain malformations. Tubulinopathies are caused by dominant missense mutations in genes encoding for tubulins, the building blocks of microtubules. Neurodevelopmental ciliopathies are mostly recessive disorders caused by defects in the function of the primary cilium, a sensory organelle that modulates signaling pathways important for brain development. Though more than 40 genes have been associated with neurodevelopmental ciliopathies, many patients still do not have an identified genetic etiology. Here, we present a novel *de novo* heterozygous missense variant in Tubulin Beta Class I (*TUBB*) identified through whole-genome sequencing analysis in a patient with both ciliopathy and tubulinopathy brain features. While microtubules are fundamental to primary cilia formation and function, no association between mutations in tubulin genes and neurodevelopmental ciliopathies has been found to date. Using patient-derived cells and gene-edited isogenic cell lines, we show that the identified variant impairs the early stages of cilia formation by altering microtubule dynamics and structure. Furthermore, we demonstrate that the disease mechanism is not haploinsufficiency and that other patient mutations in *TUBB* affect cilia formation *in vitro*, putting forward defective ciliogenesis as a contributing pathogenic factor in a subset of tubulinopathy patients.

Session Title: Molecular Effects of Genetic Variation Poster Session I

PB1039 CACNA1E (c.5365-2A>G) de novo splicing variant as a potential mechanism of pathogenicity: a novel in silico study.

Authors:

J. Lopez Rivera[1], C. Estrada[1], T. Castro[1], O. Londoño[1], M. Arcos[1], M. Chaparro[1], P. Rueda[1], L. Rios[1], D. Rodriguez[1], D. Moreno[1], C. Rucinski[2], G. Ardila-Patino[3], J. Vélez[4], O. Vidal[4]; [1]Clínica Colsanitas, Bogotá, Colombia, [2]Univ. Natl. de Colombia, Bogota, Colombia, [3]Clínica Colsanitas, Bogota, Colombia, [4]Univ. del Norte, Barranquilla, Colombia

Missense variants in CACNA1E gene have been described as causative for autosomal dominant Developmental and epileptic encephalopathy-69, characterized by early-onset seizures, hypotonia, and severe impaired development. We report a novel likely pathogenic de novo variant at the splicing site (c.5365-2A>G) for an 8-year-old patient with a behavioral disorder without seizures. Recent literature has reported several patients with similar phenotypes with splicing site variants, which suggests a pathogenicity mechanism, however it has not been reported in reference databases. Therefore, we performed an in silico study to determine the effect of this variant in the resulting protein and compare it with previous studies for missense variants. Our results suggest a different protein effect that could be associated to the wide expressivity and possible incomplete penetrance of the phenotype given by each type of variant. However, further functional studies are needed to support this hypothesis.

Session Title: Molecular Effects of Genetic Variation Poster Session II

PB1040 Cas9-targeted Nanopore Sequencing to Quantify Native DNA Methylation and Genomic Instability of SVA Hexamer in X-Linked Dystonia Parkinsonism Ventral Forebrain Organoids.

Authors:

T. Evans[1,2], A. P. R. Lorenzetti[1], L. D'Ignazio[1,2], A. Gershman[3], T. Sawada[1], B. Sheehan[1], W. Timp[4], C. Bragg[5,6], A. Paquola[1,2], J. Erwin[1,2,7]; [1]Lieber Inst. for Brain Dev., Baltimore, MD, [2]Dept of Neurology, Johns Hopkins Univ. Sch. of Med., Baltimore, MD, [3]Dept. of Molecular Biology and Genetics, Johns Hopkins Univ. Sch. of Med., Baltimore, MD, [4]Dept of BioMed. Engineering, Johns Hopkins Univ., Baltimore, MD, [5]Dept of Neurology, Massachusetts Gen. Hosp. and Harvard Med. Sch., Boston, MA, [6]The Collaborative Ctr. for X-linked Dystonia-Parkinsonism, Massachusetts Gen. Hosp., Boston, MA, [7]Dept of NeuroSci., Johns Hopkins Univ. Sch. of Med., Baltimore, MD

X-linked Dystonia Parkinsonism is a rare, inherited neurodegenerative condition caused by insertion of SINE-VNTR-Alu (SVA) element into intron 32 of TATA-box binding protein associated factor 1 (*TAF1*). Present in ~2,700 copies in the human genome, SVAs contain a canonical hexameric repeat (CCCTCT)n and similar to Huntington's Disease, the length of this hexameric sequence correlates with disease of onset in XDP patient brains. Previous studies of neurodevelopmental models have focused on the impact of SVA insertion on the differential transcription of *TAF1* in XDP cells, however, *TAF1* transcriptional differences between XDP and healthy are negligible in mature neurons. Thus, there is a need to better understand the molecular pathology in neurons, including expansion of the XDP SVA hexamer. Using a ventral forebrain organoid (VFO) model derived from XDP patient iPSCs, we performed Cas9-targeted Nanopore sequencing on the XDP SVA locus, along with four autosomal SVA_F, two SVA_D, *FMR1* (a gene associated with repeat expansion disorder), and *TP53* as a control. Combining both long-read and PCR-based methods, we evaluated the instability of the XDP SVA locus and identified both expansion and retraction of the XDP SVA hexameric sequence. Additionally, we characterized the methylation status and instability of the XDP SVA and six autosomal SVA loci to contextualize our findings. Future work will focus on understanding the source of the genomic instability at the XDP SVA locus and evaluating this phenomenon in SVAs across the human genome.

Session Title: Molecular Effects of Genetic Variation Poster Session III

PB1041 Cataloguing the landscape of short tandem repeat (STR) variants and their regulatory effects in 2,040 individuals.

Authors:

H. A. Tanudisastro[1,2,3,4], A. S. E. Cuomo[1,2,3,5,6], M. Welland[1,2,3], M. Franklin[1,2,3], A. W. Hewitt[7,8,9], G. A. Figtree[10,11], K. M. de Lange[1,2,3], J. E. Powell[5,6], D. G. MacArthur[1,2,3]; [1]Ctr. for Population Genomics, Garvan Inst. of Med. Res., Sydney, NSW, Australia, [2]Ctr. for Population Genomics, Murdoch Children's Res. Inst., Melbourne, VIC, Australia, [3]Faculty of Med. and Hlth., Univ. of New South Wales, Sydney, NSW, Australia, [4]Faculty of Med. and Hlth., Univ. of Sydney, Sydney, NSW, Australia, [5]Garvan-Weizmann Ctr. for Cellular Genomics, Garvan Inst. of Med. Res., Sydney, NSW, Australia, [6]UNSW Cellular Genomics Futures Inst., Univ. of New South Wales, Sydney, NSW, Australia, [7]Menzies Inst. for Med. Res., Univ. of Tasmania, Hobart, TAS, Australia, [8]Dept. of Ophthalmology, Royal Hobart Hosp., Hobart, TAS, Australia, [9]Ctr. for Eye Res. Australia, Univ. of Melbourne, Melbourne, VIC, Australia, [10]Charles Perkins Ctr., Univ. of Sydney, Sydney, NSW, Australia, [11]Kolling Inst. of Med. Res., Royal North Shore Hosp., Sydney, NSW, Australia

Short tandem repeats (STRs) are variations in the genome defined by repeating copies of a small (2-6bp) motif. There are over one million known STR loci, comprising approximately 3-5% of the human genome. Variation in STR repeat length modulates gene expression and can influence biological function. STR variants cause over 60 Mendelian disorders, including Huntington's disease and Fragile X Syndrome, and can influence complex traits such as dyslipidemia.

Despite their biological importance, STRs remain poorly understood due to limitations in discovery and genotyping with short-read sequencing technology. Repeat-containing reads are also difficult to align to the reference genome precisely. These challenges mean that STRs are often excluded from whole genome sequencing (WGS) pipelines and therefore represent a crucial gap in our understanding of genetic variation and their downstream biological impact on gene expression.

We genotyped 164,847 STR loci in a cohort of 1,055 Australian adults for whom both WGS and single-cell RNA sequencing (scRNA-seq) data from approximately 1,000 peripheral blood mononuclear cells are available for each individual. We have completed single nucleotide polymorphism (SNP) genotyping and are generating STR calls for a separate cohort of 985 Australians. The final cohort of 2,040 individuals, phase 1 of the TenK10K program, will be the largest cohort of paired human WGS and scRNA-seq data generated to date, enabling not only the characterisation of STR variation genome-wide but also novel analysis of the effect of STRs on gene expression using single-cell transcriptomic data.

After filtering the STR catalogue and performing stringent QC on the variant calls from ExpansionHunter, we identified 157,551 loci that are confidently variable in the first 1,055 individuals. The motif distribution of the STR loci is mainly composed of dinucleotides (60%) and tetranucleotides (23%). The mean STR length genotyped was 26bp (max: 505bp, min: 2bp), and the mean number of distinct alleles genotyped per locus was 6.8.

This robust catalogue of STR variation will now be combined with the paired scRNA-seq data to identify gene expression-modulating STRs in a cell type-specific context.

Session Title: Molecular Effects of Genetic Variation Poster Session I

PB1042 *CD33* SNP rs2455069: A hypothesis that arginine/glycine substitution affects CD33 functional role in microglial activation.

Authors:

E. Vitale[1], A. Angiolillo[2], A. Di Costanzo[2], A. Donizetti[3]; [1]Inst. of Biochemistry and Cell Biology, Natl. Res. Council (CNR), Napoli, Italy, [2]Ctr. for Res. and Training in Med. of Aging, Dept. of Med. and Hlth.Sci. "V. Tiberio", Univ. of Molise, Campobasso, Italy, [3]Dept. of Biology, Univ. of Naples Federico II, Napoli, Italy

Modifications in the CD33 gene identified by genome-wide association studies (GWAS) make it one of the highest ranked risk factors for Alzheimer Disease (AD). CD33 is a sialic acid-binding immunoglobulin (Ig)-type lectin (Siglec) receptor expressed in myeloid cells, including microglial cells. Its binding to sialic acid-modified glycoproteins and glycolipids leads to inhibition of cellular functions such as phagocytosis. We have identified rs2455069-A>G (R69G), a Single-Nucleotide-Polymorphism (SNP) in CD33 exon 2 and demonstrated that patients carrying this SNP had an increased risk of AD. However, beyond an association with a few reported cases, the function, and potential mechanisms of most of the CD33 SNPs in promoting AD remains elusive. To try to attribute a functional role to this polymorphism, we demonstrated, using bioinformatics analyses, that the CD33-R69G variant protein has an increased binding affinity to sialic acid, and consequently, a more efficient binding to sialic acid-containing gangliosides at lower concentrations than the wild-type variant. We hypothesize that the binding of the CD33-R69G form to sialic acid could have long-term cumulative effects and may explain a form of late-onset of Alzheimer's disease, by an inefficient removal of β-amyloid peptide. To test our hypothesis for a role for the CD33 variant and identify a potential mechanism for its causative effects in LOAD, we plan to generate a human cell-based system, similar to one reported by Wißfeld et al. [Sci Rep. 2021], by using the Flp-In™-293 Cell Line and Flp-In system. The system will be based on fusing the extracellular CD33 domain to TYROBP/DAP12 to monitor the activation of the CD33 by sialic acid binding. As redout of the activation, we will analyze the level of pSYK and calcium. We will evaluate differences in CD33 signaling resulting from an altered binding capacity of CD33-R69G for gangliosides. In addition, we will approach a putative therapeutic strategy, using the antagonistic effects of the anti-CD33 antibody in combination with different concentrations of gangliosides as proof of principle of an existing interrelation between anti-CD33 therapeutic efficacy and ganglioside metabolism. This strategy will test, for the first time, whether this treatment approach can be developed for addressing late onset Alzheimer's disease.

Session Title: Molecular Effects of Genetic Variation Poster Session II

PB1043 Challenges and solutions in variant analysis and therapy testing in the retinal dystrophies.

Authors:

R. Jamieson, T. Loi, J. Grigg, A. Cheng, S. Eamegdool; Eye Genetics Res. Unit, Children's Med. Res. Inst., Sydney Children's Hosp. Network, Save Sight Inst., Univ. of Sydney, Sydney, Australia

Background: The increasing number of genetic therapies and clinical trials for patients with inherited retinal disorders (IRDs) is a tremendous step forward. Eligibility for these therapies requires a clearcut genetic diagnosis. Genomic investigation may lead to variants of uncertain significance (VUS) using ACMG criteria. We have developed a pipeline of investigation using patient-derived or genome edited induced pluripotent stem cells (iPSCs) differentiated to retinal organoids (iROs) or retinal pigment epithelium (iRPE). Expression and omic studies were used in criteria assessment for VUS reclassification, as well as for novel therapy investigation. Methods: Patient-derived iPSC lines were created, and isogenic controls as well as known pathogenic or variants for investigation were generated. Guide RNAs were selected and homology directed repair templates were used to introduce the relevant variant sequence or create isogenic controls for VUS in *RPGRIP1* and *PDE6B.* Variant and control iPSC lines were differentiated to iROs. Organoids were examined for morphological markers of photoreceptor outer segment development and function. RT-PCR, transcriptomic and immunofluorescence studies were used to assess variations in expression and protein interactions. Results: Human iROs with VUS in *RPGRIP1* and *PDE6B* were successfully generated. Lines with VUS demonstrated abnormal morphology, protein expression, or transcriptomic abnormalities, compared with control lines. Abnormal localisation of outer segment photoreceptor proteins was a key finding. Transcriptomic analysis identified characteristic signature pathways in those with pathogenic variants, compared with controls. Conclusions: Novel VUS in IRD genes, *RPGRIP1* and *PDE6B*, showed markers of pathogenicity using patient-derived and genome edited iPSCs differentiated to iROs. Analyses of morphological features, protein trafficking, RNA studies and transcriptomic profiling, were useful as indicators for VUS reclassification. This approach is valuable in assessment of VUS and pre-clinical assessment of novel therapies.

Session Title: Molecular Effects of Genetic Variation Poster Session III

PB1044 Characterization of allele-specific expression in Alzheimer's Disease

Authors:

V. Subramaniam, Z. Wang, R. Vaza, M. Wang, K-l. Huang; Icahn Sch. of Med. at Mount Sinai, New York, NY

Alzheimer’s disease (AD) is the leading cause of dementia, accounting for 50-70% of cases worldwide. Previous studies profiling gene expression in heritable AD have identified several loci conferring disease risk. However, the mechanism of pathogenesis remains ambiguous. Allele-specific expression (ASE) is the phenomenon through which one allele is preferentially expressed through *cis*-acting regulators and may explain how risk variants lead to AD. We assessed ASE in AD using DNA- and RNA-sequencing data from the Mount Sinai Brain Bank (MSBB) and Religious Orders Study/ Memory and Aging Project (ROSMAP) databases. The databases included samples from various brain regions as well as clinical variables related to AD. Analysis of both sample sets showed ASE prevalence of 1-2% across the AD genome. As expected, imprinted genes carried more ASE variants compared to other genes. Many ASE variants were shared between all four brain regions. However, symptomatic AD individuals showed greater ASE in Brodman area 22 compared with Brodman 36, 44, and 10. At the chromosome level, 18 risk sample variants were identified through a linear mix model applied to each brain region. Chromosomes 6, 14, and 15 were enriched for ASE variants in both control and AD cohorts, while chromosome 7 was enriched for ASE variants exclusively in the AD cohort. At the chromosome band level, regions 14q32 and 15q11, which have previously demonstrated higher minor allele expression in autism spectrum disorder, were enriched for ASE in our AD sample. Finally, we observed that nearly all variants located at chromosome 9 q13, and chromosome 7 q32.2 demonstrated ASE in both AD and control cohorts . Overall, we have provided a comprehensive characterization of ASE in AD, Further analysis is necessary to understand the implications of these variants in AD pathogenesis and regulation.

Session Title: Molecular Effects of Genetic Variation Poster Session I

PB1045 Characterization of extracellular vesicles from fibroblast-derived osteoblast-like cells in patients with *SGMS2*-related osteoporosis.

Authors:

K. Määttä[1,2], S. Pihlström[1,2], R. Mäkitie[1,2,3], O. Makitie[1,2,4,5], M. Pekkinen[1,2,4]; [1]Inst. of Genetics, Folkhälsan Res. Ctr., Helsinki, Finland, [2]Res. Program for Clinical and Molecular Metabolism, Faculty of Med., Univ. of Helsinki, Helsinki, Finland, [3]Dept. of Otorhinolaryngology – Head and Neck Surgery, Helsinki Univ. Hosp. and Univ. of Helsinki, Helsinki, Finland, [4]Children´s Hosp., Helsinki Univ. Hosp. and Univ. of Helsinki, Helsinki, Finland, [5]Dept. of Molecular Med. and Surgery and Ctr. for Molecular Med., Karolinska Inst.t, Stockholm, Sweden

Pathogenic variants in *Sphingomyelin Synthase 2 (SGMS2)*, an essential gene in sphingomyelin metabolism, associate with a rare autosomal dominant skeletal disorder characterized by low bone mineral density, increased spinal and peripheral fractures and sclerotic doughnut-shaped lesions in cranial bones. Sphingomyelin metabolism sustains skeletal homeostasis through yet unknown molecular mechanism. Extracellular vesicles (EVs) are nanoparticles secreted by various cell types. EVs possess important roles in various biological processes, such as osteogenic differentiation, by functioning as cell-to-cell communicators. Aim of the study was to characterize EVs from pathogenic *SGMS2* p.Arg50* mutation carriers and healthy controls by using fibroblast-derived osteoblast-like cells. EVs were isolated from the media of the cell cultures of 4 patients and 4 healthy controls in both fibroblast-stage (day 0) and osteoblast-like stage (differentiation of 27 days) using ultracentrifugation-based method and characterized by size, concentration and the presence of protein markers using nanoparticle tracking analysis, immuno electron microscopy and dot blot method. In addition, the RNA-content of EVs will be analyzed by RNA sequencing. As a result, the average EV concentration increased in patients from 2.0E+11 (range 1.7-2.4E+11) particles/ml (day 0) to 5.1E+11 (range 3.7-6.6E+11) particles/ml (day 27) (p=0.03) and in controls from 2.0E+11 (range 1.7-2.3E+11) particles/ml (day 0) to 5.0E+11 (range 3.1-7.8E+11) particles/ml (day 27) (p=0.03). The average EV median size increased in patients from 109.9 nm (range 104.6-117.2 nm) (day 0) to 121.1 nm (range 112.0-129.3 nm) (day 27) (p=0.11) and in controls from 107.4 nm (range 98.5-119.8 nm) (day 0) to 121.9 nm (range 115.5-133.4 nm) (day 27) (p=0.11) indicating increased cargo content. No statistically significant difference in EV concentration and median size between patients and controls was observed at timepoints 0 and 27 days. Protein content-based EV characterization confirmed presence of markers CD9, CD63, CD81 and HSP70 and the absence of ER marker calnexin. RNA cargo analyses are ongoing. In conclusion, fibroblast-derived osteoblast-like cells can provide an alternative approach for osteogenic EV studies aiming to gain novel insights into the molecular mechanisms and therapeutic applications of EVs in skeletal fragility.

Session Title: Molecular Effects of Genetic Variation Poster Session II

PB1046 Characterizing genetic variation in the regulation of the ER stress response through computational and *cis*-eQTL analyses

Authors:

N. Russell, L. B. Jorde, C. Y. Chow; Univ. of Utah, Salt Lake City, UT

Misfolded proteins in the endoplasmic reticulum (ER) elicit the ER stress response, a large transcriptional response driven by three well-characterized transcription factors. This transcriptional response is variable across different genetic backgrounds. One mechanism by which genetic variation can lead to transcriptional variability in the ER stress response is through altered binding and activity of the three main transcription factors: XBP1, ATF6, and ATF4. This work attempts to better understand this mechanism by first creating a computational pipeline to identify potential binding sites throughout the human genome. We utilized GTEx datasets to identify *cis*- eQTLs that fall within predicted transcription factor binding sites (TFBSs). We also utilized the ClinVar database to compare the number of pathogenic versus benign variants at different positions of the binding motifs. Finally, we performed a *cis*- eQTL analysis on human cell lines experiencing ER stress to identify *cis*- eQTLs that regulate the variable ER stress response. The majority of these *cis*- eQTLs are unique to a given condition: control or ER stress. Some of these stress-specific *cis*- eQTLs fall within putative binding sites of the three main ER stress response transcription factors, providing a potential mechanism by which these *cis*- eQTLs might be impacting gene expression under ER stress conditions through altered TF binding. This study represents the first *cis*- eQTL analysis on human samples experiencing ER stress and is a vital step towards identifying the genetic components responsible for variation in the ER stress response.

Session Title: Molecular Effects of Genetic Variation Poster Session III

PB1047 Characterizing the genetic architecture of comprehensive inflammatory markers and their associations with obesity

Authors:

B. Truong[1], Z. Sun[2], Z. Wang[3], V. Sarsani[2], J. Hu[2], Z. Wang[2], L. Liang[4], J. Li[5]; [1]Broad Inst. of MIT and Harvard, Cambridge, MA, [2]Dept. of Epidemiology, Harvard T.H. Chan Sch. of Publ. Hlth., Boston, MA, [3]Dept. of Biostatistics, Harvard T.H. Chan Sch. of Publ. Hlth., Boston, MA, [4]Harvard Sch Publ. Hlth., Boston, MA, [5]Brigham & Women s Hosp., Harvard Med. Sch., Boston, MA

Systemic inflammation in obesity is postulated to play a pathogenic role in the development of obesity-related diseases including type 2 diabetes, cardiovascular disease, and cancers. However, the genetic architecture of inflammatory markers and their casual relationships with obesity is yet to be elucidated. To address this gap, we curated large-scale GWAS summary statistics for 138 cellular and molecular markers of inflammation by meta-analyzing data from several cohorts and recent proteomics GWAS (median N = 39,747, ranging from 2,538 to 516,121), and for obesity and adiposity traits (i.e., body mass index [BMI], and subcutaneous, visceral, liver, and pancreas fat volumes) from the UK Biobank (median N = 40,796, ranging from 21,705 to 337,000).
We identified 83 markers with significant heritability ranging from 2.7% to 50% (FDR < 5%). Hierarchical clustering analysis on pairwise genetic correlations of 83 significant markers identified 3 subclusters with averaged pairwise correlations greater than 0.7, for example, interleukins - TNFa/b - growth factors, blood cells - C-reactive protein, and PDL-1 - VCAM1 - TNFaR1/R2. In contrast, adiponectin, MMP2, and DNER showed negative genetic correlations with most other inflammatory markers. Pairwise bidirectional Mendelian Randomization (MR) identified 35 pairs of suggestive causations (FDR < 5%) between specific inflammatory markers (e.g. WBC - lymphocyte, and TRAILr1 on MMP16), highlighting the potential regulation of inflammatory pathways. Colocalization with tissue-specific eQTLs suggested that most of these inflammatory markers are expressed in multiple tissue types, especially in whole blood, adipose tissues, thyroid and nerve tibial tissues. Further examination of obesity/adiposity traits revealed significantly positive genetic correlations between leptin, CRP, interleukins, and TNF families with obesity/adiposity (rg ranged from 0.07 to 1), whereas adiponectin, MMP2, and DNER showed negative genetic correlations with obesity/adiposity (rg ranged from -0.03 to -1). Finally, MR highlighted the causality of obesity on inflammatory markers such as BMI on IL6, TNF receptor 2 and VEGF (effect = 0.16-0.3), waist-hip-ratio adjusting for BMI on adiponectin (effect = -0.4) and CCL25 (effect = 0.3).
Our study provided a comprehensive map of the genetic architecture of inflammatory markers and their shared genetics with obesity/adiposity. Given the pivotal role of inflammation in diseases and aging, a better understanding of the complex relationships between obesity and inflammatory signaling could inform novel therapeutics for the prevention of obesity comorbidities and complications.

Session Title: Molecular Effects of Genetic Variation Poster Session I

PB1048 Chromatin accessibility QTL mapped in 138 liver tissue donors identify functional variants and genes at GWAS loci.

Authors:

K. Currin[1], H. J. Perrin[1], A. S. Etheridge[1,2], S. Vadlamudi[1], A. E. Musser[1], G. K. Pandey[1], A. A. Alkhawaja[1], K. A. Broadaway[1], J. D. Rosen[1], A. Varshney[3], A. S. Chaudhry[4], P. J. Gallins[5], Y. Zhou[5], S. C. J. Parker[3], L. Raffield[1], F. Wright[5], E. Schuetz[4], F. Innocenti[2,6], K. L. Mohlke[1]; [1]Dept. of Genetics, Univ. of North Carolina, Chapel Hill, NC, [2]Eshelman Sch. of Pharmacy, Univ. of North Carolina, Chapel Hill, NC, [3]Dept. of Computational Med. & Bioinformatics, Univ. of Michigan, Ann Arbor, MI, [4]Dept. of Pharmaceutical Sci., St. Jude Children's Res. Hosp., Memphis, TN, [5]Bioinformatics Res. Ctr., North Carolina State Univ., Raleigh, NC, [6]Lineberger Comprehensive Cancer Ctr., Univ. of North Carolina, Chapel Hill, NC

The majority of GWAS loci contain variants within regulatory elements, but existing regulatory element maps for many tissues are based on a small number of donors and may not capture individual variability in activity. To identify individual differences in regulatory element activity in liver tissue, we mapped chromatin accessibility using ATAC-seq from 138 donors and identified 358,304 regulatory elements (peaks) present in at least 7 donors. Compared to peaks found in many donors, peaks found in fewer donors are less likely to overlap DNase-seq peaks from ENCODE or promoter and enhancer states from Roadmap liver, indicating that some peaks are not captured by existing datasets. To identify peaks that differ by genotype, we mapped chromatin accessibility QTL (caQTL) and identified 35,361 peaks with a caQTL (caPeaks, FDR<5%, FastQTL). CaQTL variants are more likely to disrupt motifs of specific transcription factors (TFs), including the liver key regulators HNF4A, FOXA1/2/3, and CEBPB. We identified 2,126 caQTL signals that contain multiple caPeaks (coordinated caPeaks), including 271 signals that show long-range coordination (> 100 kb). Signals with coordinated caPeaks are more likely to colocalize (coloc PP H4>0.8) with GTEx liver eQTL compared to signals with 1 caPeak (OR=4.1, p=8E-99), even after adjusting for transcription start site (TSS) proximity (OR=3.7, p=4E-85). We predicted 16,198 links between caPeaks and target genes using TSS proximity, eQTL colocalization, Hi-C, and coordination of distal and promoter peaks. The 9,019 unique target genes of caPeaks include 384 TFs, 550 drug response genes, and 130 liver-enriched genes defined by the Protein Atlas. Using summary statistics of 703 traits (UK Biobank, Neale lab), caPeaks are enriched for heritability of 21 traits (LDSC, FDR<5%), including liver enzymes, cholesterol, albumin, and C-reactive protein. CaQTL colocalized with dozens of GWAS signals for liver-relevant traits, including 113 for total cholesterol and 54 for the liver enzyme gamma-glutamyl transferase (GGT). To validate regulatory activity, we focused on a GGT GWAS signal colocalized with a caQTL signal containing 40 coordinated caPeaks that was linked to TENM2 by eQTL colocalization and promoter-distal caPeak coordination. We observed significant allelic differences in transcriptional activity in HepG2 reporter assays for rs7726117, which was the only variant at the locus in a caPeak (forward fold change (FC) 1.5, p<1E-4, reverse FC 1.3, p<1E-4). These results show the utility of increased sample size for identifying unannotated regulatory elements and for mapping caQTL to identify functional variants and genes at GWAS loci.

Session Title: Molecular Effects of Genetic Variation Poster Session II

PB1049 Chromatin interaction inferences based on the paired transcriptome and chromatin accessibility datasets identify candidate risk genes of autoimmunity.

Authors:

H. Hatano[1], A. Suzuki[2], M. Nakano[1], A. Oguchi[3], Y. Murakawa[3], C. Terao[4], K. Yamamoto[2], K. Ishigaki[1]; [1]Lab. for Human Immunogenetics, RIKEN Ctr. for Integrative Med. Sci., Yokohama, Kanagawa, Japan, [2]Lab. for Autoimmune Diseases, RIKEN Ctr. for Integrative Med. Sci., Yokohama, Kanagawa, Japan, [3]RIKEN-IFOM Joint Lab. for Cancer Genomics, RIKEN Ctr. for Integrative Med. Sci., Yokohama, Kanagawa, Japan, [4]Lab. for Statistical and Translational Genetics, RIKEN Ctr. for Integrative Med. Sci., Yokohama, Kanagawa, Japan

Genome-wide association studies (GWAS) have identified numerous risk variants associated with autoimmunity onset. Risk variants are frequently found within cell-type specific gene regulatory regions predicted to interact physically with the promotors of the risk gene via the chromatin interaction and regulate its expression level. Therefore, previous studies inferred the chromatin interactions between the cis-regulatory elements (CRE) harboring risk variants and the promotor regions of surrounding genes using functional genomic assays (e.g., HiC). However, such assays usually evaluated cell lines due to the requirement of large input material and failed to detect the chromatin interaction specific to primary immune cells. To overcome these limitations, we sought to infer functional chromatin interactions of primary immune cells by developing a novel experimental and analytical approach. We first obtained peripheral blood from 70 healthy donors and fine-sorted 26 distinct immune cell subsets by FACS and conducted RNA sequencing (RNA-seq) and Assay for Transposase-Accessible Chromatin sequencing (ATAC-seq). The canonical correlation analysis confirmed cell-type specific covariations between RNA- and ATAC-seq datasets, supporting the feasibility of the paired dataset integration. The stratified LD score regression found that the CREs detected by ATAC-seq explained a substantial fraction of the autoimmunity's heritability, such as T cell lineage subsets for rheumatoid arthritis (RA) and B cell lineage subsets for systemic lupus erythematosus (SLE). We then comprehensively tested the associations of each gene's expression level with the accessibility of all CREs within ±1 Mb of the gene's TSS. We evaluated the entire associations using the mashr software and confirmed that the CRE-gene associations preserved immunologically reasonable cell-type specific patterns. Among all 10,902,917 CRE-gene pairs, 29,277 associations (0.27%) were significant (local false sign rate<0.1) in at least one cell type, and 14,592 associations (0.13%) were specific to only one cell type. The significant CRE-gene pairs were enriched within the previously reported HiC datasets, supporting that our results reflect the physical chromatin interactions. Finally, we leveraged the CRE-gene associations to infer the autoimmunity risk genes. We overlayed the autoimmunity risk variants onto the CRE and found multiple candidate risk genes: e.g. 73 genes for RA and 77 genes for SLE. Together, our rich resource of primary immune cell-specific CRE-gene association database improved our understanding of autoimmunity pathology.

Session Title: Molecular Effects of Genetic Variation Poster Session III

PB1050 Clinical interpretation of partial duplication and deletion involving neurodevelopmental genes

Authors:

J. England[1], G. Huguet[2], K. Jizi[3], C. Poulain[4], E. Douard[3], Z. Saci[3], S. Jacquemont[5]; [1]Université de Montréal, Montréal, QC, Canada, [2]CHU Ste justine, Montréal, QC, Canada, [3]CHU Sainte-Justine, Montréal, QC, Canada, [4]Montréal Univ., Montreal, QC, Canada, [5]CHU Sainte justine Res. Ctr., Montreal, QC, Canada

Chromosomal microarrays are routinely used to detect copy number variants (CNV) in patients with neurodevelopmental disorders including intellectual disability and autism spectrum disorder and the diagnostic yield is around 20%. A gene fully deleted or duplicated has clear consequences with a significant decrease or increase in transcription respectively. These changes in known dosage sensitive neurodevelopmental (ND) genes are mostly observed *de novo.* However, CNVs involving only a portion of a known ND gene are frequently observed in the diagnostic setting and whether these variants are gene disrupting is difficult to establish, especially for duplications. Our aim is to quantify the deleteriousness of partial deletions and duplications involving known ND genes.
We used a list of 614 ND genes based on literature (McRae JF et al. 2015, Coe et al. 2019, Satterstrom et al. 2020), PanelApp and ClinGen. We studied transmission of CNVs in 8113 individuals with microarrays performed at the CHU Sainte-Justine diagnostic laboratory. We compared the proportion of inherited and *de novo* CNVs by Fisher's test. The analysis was replicated in the DECIPHER database.
Partial duplications of ND genes (n= 38) almost never occurred *de novo* (3%) and this was not significantly different from the frequency of *de novo* partial duplications in non-ND genes (1%), which is also the probability that a neutral CNV is observed *de novo*. On the other hand, the proportion of duplications fully including a ND gene observed *de novo* was significantly higher (43%, $p= 2.68 \times 10^{-5}$); also higher ($p=1.07 \times 10^{-10}$) than the complete duplications of non-ND genes (14%). On the other hand, the proportion of *de novo* complete (80%) and partial (71%) deletions involving ND genes were both very high and not significantly different.
Complete CNVs of ND genes are more often *de novo* which suggest that haploinsufficient genes are also triplosensitive. Complete and partial deletion of ND genes have a similar pattern of inheritance, we conclude that the mechanism of action is probably loss of function. Partial duplications of ND genes are rarely *de novo* (a rate similar to the probability of a neutral CNV occurring *de novo* and to the non-ND genes in our dataset). As opposed to partial deletions, partial duplications appear to be benign in the vast majority of cases. This could be explained by the fact that the duplicated segment is in tandem and do not disturb the function of the gene or the segment is located elsewhere in the genome and doesn't disrupt the function of a gene.
In conclusion, partial duplications can be interpreted as benign in most cases and clinician should pursue additional genetic testing for the phenotype of the patient.

Session Title: Molecular Effects of Genetic Variation Poster Session I

PB1051 Cluster analysis identifies genes with distinct patterns of loss of function variants

Authors:

R. Beaumont[1], C. Wright[2]; [1]Univ. of Exeter Med. Sch., Exeter, United Kingdom, [2]Univ. of Exeter, Exeter, United Kingdom

Background: Genetic variants that severely alter protein products (e.g. stop-gain, start-loss) are often associated with disease. For some genes, these predicted loss-of-function variants (pLoFs) are observed throughout the gene in population cohorts, while in others they occur only at specific locations. We hypothesised that for disease genes with incomplete penetrance, regions with pLoF variants present in a population cohort may be those where pLoFs are tolerated. To test this, we investigated whether pLoF location could explain instances of incomplete penetrance of variants expected to be pathogenic for Mendelian conditions.
Methods: We used exome sequence data in 454,773 individuals in UK Biobank (UKB) to investigate locations of pLoFs in a population cohort. We counted numbers of unique pLoF, missense, and synonymous variants in UKB in each quintile of the gene, and clustered the variants using Gaussian mixture models. We limited analyses to genes with coding sequence >1000bp and ≥5 variants of each type (11,050 genes). We compared locations of pLoFs in UKB with all theoretically possible pLoFs and pathogenic pLoFs from ClinVar, and performed simulations to estimate the false-positive rate of non-uniformly distributed variants.
Results: For most genes, synonymous, missense and all possible pLoF variants fell into clusters representing uniform variant distributions. Of these, pLoFs were not uniformly distributed for 625 genes. Notably, genes causing developmental disorders via haploinsufficiency were less likely to have uniform pLoF distribution than other genes ($P<0.001$). For two such genes, *DYRK1A* and *ODC1*, pathogenic pLoFs were located approximately uniformly across the gene, whilst likely benign pLoFs were clustered in specific locations, suggesting rescue via translation reinitiation, alternative splicing or other mechanisms. We also found that fluid intelligence was lower (-0.02 units) in individuals carrying pLoF variants in genes where pLoFs altered plasma protein levels than those where pLoFs did not alter protein levels (0.003 units).
Conclusions: Our results suggest potential benefits of localised constraint metrics and that location of pLoF variants should be considered when interpreting variants.

Session Title: Molecular Effects of Genetic Variation Poster Session II

PB1052 Comparison of blood group serology and whole genome sequencing identifies rare blood group alleles in Oman.

Authors:

P. Haffener[1], A. Al-Riyami[2], S. Al-Zadjali[3], G. Busby[4], M. Al-Rawahi[5], S. Al Hosni[2], A. Al Marhoobi[5], A. Al Sheriyani[6], E. M. Leffler[1]; [1]Univ. of Utah, Salt Lake City, UT, [2]Sultan Qaboos Univ. Hosp., Muscat, Oman, [3]Sultan Qaboos Comprehensive Cancer Ctr., Muscat, Oman, [4]Allelica, New York, NY, [5]Sultan Qaboos Univ., Muscat, Oman, [6]Royal Oman Police Hosp., Muscat, Oman

The earliest human genetics studies focused on blood group variation, beginning with the ABO blood group. Today, there are 44 recognized blood groups that are determined by genetic variation in 49 different genes. These blood group variants are curated in resources such as the ISBT database, which has allowed for genotype-based methods to be developed for inferring patient blood types. However, for these to be of clinical use, rare and population-specific variants must be more widely surveyed and included. Failing to do so could increase the risk for transfusion rejections in repeat-transfusion patients due to blood type incompatibility. For instance, a recent study in Oman compared serological phenotypes with blood groups inferred from targeted genotyping assays and found that few blood groups had 100% concordance. To identify population-specific and rare blood group variants responsible for these discrepancies, here we expand to whole genome sequencing in an independent set of 100 Omani donors. We performed both variant calling and serological testing across 10 blood groups: ABO, RHD, RHCE, MNS, Lutheran, Lewis, Kidd, Kell, Duffy, and P. Only two blood groups, Kidd and Kell, had 100% concordance. With the exception of MNS which had a concordance of 85%, the remaining blood groups had 95%-99% concordance. Further investigation of alleles carried by discordant samples resulted in the first reports in Oman of Fy^X, a common variant in the Duffy blood group system known to reduce expression of Fy^b, and of the Henshaw variant of the MNS blood group, the most common cause of the S-s-U^{+var} blood type in African populations. We also identified the RHD(psi) variant which is a common cause of the RH(-) phenotype in Africa and can result in fetal hemolytic disease of the newborn. Lastly, we report a rare frameshift resulting in a null phenotype of the Duffy blood group, serologically identical to the Fy^{ES} allele thought to have been selected in sub-Saharan Africa due to protection against malaria. Through this approach, we show that the comparison of whole genome sequencing to serology is a valuable strategy to identify additional variants that should be accounted for to predict blood types from genetic data. These results add to our knowledge of causes of blood group variation in Oman, which will inform future testing in blood banks and transfusion services.

Session Title: Molecular Effects of Genetic Variation Poster Session III

PB1053 Comprehensive detection of *trans*-regulatory signal in pooled single-cell CRISPR screens

Authors:

N. Babushkin, X. Liu; Univ. of Chicago, Chicago, IL

Single-cell RNA-seq screens with pooled CRISPR perturbations (Perturb-seq) have emerged as a powerful technique in the field of functional genomics. In this study, we used a unified pipeline to process several large-scale Perturb-seq datasets in K562 cells and comprehensively identified *trans*- regulatory signals across the genome. The datasets in Gasperini 2019 Cell, Morris 2023 Science, Replogle 2022 Cell, and Xie 2019 Cell Rep. targeted a total of 13,729 genes, enhancers, and cCREs, totaling 3,948,462 cells. However, we found that RNA-sequencing reads mapped to multiple homologous locations on the genome lead to around 70% of the false *trans*-signals in Perturb-seq datasets, agreeing with the findings in bulk RNA-seq datasets (Saha 2019 F1000R). To reduce false positive *trans*- signals, we first thoroughly removed reads mapped to genomic regions of low mappability. We then applied SCEPTRE (Barry 2021 Genome Bio.) to identify a set of high-quality *trans*-regulatory signals. In total, we found 3,417 significant *trans*- regulatory gene pairs in three of the datasets. These significant pairs contained a reduced percentage of cross-mappable genes across all three datasets after mappability correction. To assess whether Perturb-seq could serve as a validation strategy for *trans*-eQTLs discovered in RNA-seq datasets, we compared the significant signals to *trans*- eQTLs discovered in the DGN dataset (whole blood). We found that none of the 165 significant *trans*-eQTL signals in DGN were replicated in the Perturb-seq analysis, agreeing with recent findings (Yao 2023 bioRxiv). To examine replication between different Perturb-seq studies, we tested for replication of significant *trans*- regulatory signals. We found that up to 7% of the *trans*- regulatory signals replicate between the four Perturb-seq datasets. In conclusion, our work presents a uniform data analysis pipeline and applies a well-calibrated association test to comprehensively discover high-quality *trans*- regulation signals in Perturb-seq datasets.

Session Title: Molecular Effects of Genetic Variation Poster Session I

PB1054 Comprehensive functional characterization of *SGCG* coding variants accurately predicts pathogenicity in limb-girdle muscular dystrophy type R4/2E

Authors:

G. Haller[1], C. Li[1], J. Wilborn[2], J. Daw[1], S. Pittman[1], J. Alonso-Pérez[3], J. Díaz-Manera[4], C. Weihl[1]; [1]Washington Univ., St. Louis, MO, [2]Univ. of Utah, Salt Lake City, UT, [3]Hosp. Univ.rio Nuestra Señora de Candelaria, Tenerife, Spain, [4]Newcastle Univ., Newcastle, United Kingdom

The majority of clinically sequenced patients with limb-girdle muscular dystrophy (LGMD) carry variants of unknown significance (VUS) in myopathy-causing genes, leaving them without a genetic diagnosis. α-, β-, γ-, and δ-sarcoglycan form a transmembrane complex (SGC) and loss of function mutations in any subunit causes LGMD. We utilized deep mutational scanning to perform high-throughput screening of SGCB and SGCA membrane localization for all 6340 possible amino acid changes in *SGCB*. Variant functional scores were bimodally distributed and perfectly predicted pathogenicity of known variants. Variants with less severe functional scores more often appeared in patients with slower disease progression, implying a relationship between variant function and disease severity. Positions intolerant to variation mapped to points of predicted inter-SGC interactions, validated *in silico* structural models and enabled accurate prediction of pathogenic variants in other SGC genes. These results enable immediate clinical interpretation of newly discovered *SGC* gene variants.

Session Title: Molecular Effects of Genetic Variation Poster Session II

PB1055 † Comprehensive genomics across cardiovascular diseases using DRAGEN

Authors:

S. Behera[1], M. Rossi[2], S. Truong[2], S. Mehtalia[2], J. Han[2], F. Zhang[2], G. Metcalf[1], S. Catreux[2], R. Mehio[2], M. Ruehle[2], G. Parnaby[2], F. Sedlazeck[3]; [1]Baylor Coll. of Med., Houston, TX, [2]Illumina Inc., San Diego, CA, [3]Baylor Coll. Med., Houston, TX

Different genetic variations, such as small variants (SNVs), short tandem repeats (STRs), structural variants (SVs) and copy number variations (CNVs) impact various human health outcomes, from complex human diseases to predisposition to specific conditions. Detecting these different variant classes accurately and comprehensively on a single assay and sequencing technology is challenging. The presence of repetitive, heterogeneous regions and errors or biases within the reference genome further complicates the accurate detection of these variants.
In this study, we showcase the DRAGEN germline pipeline that allows us to comprehensively study the human genome across different variant types, at scale and speed (~30min for 30x WGS). DRAGEN employs a novel multigenome alignment method whereby alternate paths representing genetic diversity in the population are added to the reference. This approach significantly improves the mapping, especially in difficult-to-map regions of the genome. Combined with advancements in the variant caller, DRAGEN achieves superior variant calling accuracy compared to other pipelines, including those using recent pangenome graph implementations.
DRAGEN facilitates a comprehensive representation by simultaneously reporting SNVs, SVs, CNVs, STRs and targeted callers across eight highly medically relevant genes (e.g. *SMN*1&2, *GBA*, *LPA*, etc.). This provides deep insights into genetic diseases, such as the *LPA* caller being able to report phased KIV-2 CNV that more accurately reflects measured Lp(a) and thus cardiovascular risk. We assessed all novel features across HG001-07 control samples and 3,202 samples from 1kGP. We validated the improved performance by benchmarking over GIAB truth sets and comparing them with other pipelines. For the HG002 sample on small variants (<50bp), genome-wide and in challenging medically relevant regions respectively, DRAGEN (F-score: 99.85% and 98.64%) performs better than BWA-GATK (99.07% and 95.19%) and BWA-DeepVariant (99.62% and 97.29%). For SV, it achieves a higher F-score (76.9% and 82.6% for INS and DEL respectively) than Manta (34.9% and 70.08%), Lumpy (0% and 66.8%) and Delly (4.7% and 68.3%). DRAGEN performs significantly better (F-score: 94.6%) for CNVs with lengths 1-10Kbp when compared to CNVnator (50.5%).
We applied DRAGEN on ~30,000 samples with rich medical records from a diverse cardiovascular cohort of African-, Hispanic- and European-American populations. By leveraging DRAGEN's comprehensive variant representation and multigenome approach, we will uncover novel insights with association studies and disease-related gene discovery, previously unattainable.

Session Title: Molecular Effects of Genetic Variation Poster Session III

PB1056 Comprehensive statistical analysis on the impact of missense mutations toward improving variant interpretation

Authors:

S. Kwon, J. Safer, S. Iqbal; Broad Inst. of MIT and Harvard, Cambridge, MA

Accurate interpretation of the pathogenicity of missense variations in the protein-coding genes is essential in clinical genetics, requiring the understanding of variants' effects in the context of protein structure and function. We addressed a challenge in the scalable linking of cross-disciplinary genomics and protein structural data through our discovery portal, Genomics 2 Proteins (G2P;g2p.broadinstitute.org). Combining the resources from G2P portal and AlphaFold database, we aim for the most comprehensive characterization of the structural and molecular effects of missense variants for the entire protein-coding human genome. We aggregated over 4.8M missense variations from ClinVar (N=1,054,009), HGMD (N=183,113), and the genome aggregation database/gnomAD (N=3,587,776), covering 17,836 protein-coding genes. Using MANE transcript to protein isoform mapping, we mapped 91% of variations onto protein sequence/structure. Variants were further annotated with 66 protein and genomic features including physicochemical properties of amino acid, protein structural and functional features, and DNA base-pair changes. We carefully filtered the dataset based on their clinical significance and remove duplicates, resulting in a dataset of 572,226 pathogenic and 2,742,585 population variations. Through Fisher's exact test, we identified 41 pathogenic (odds ratio/OR>1 and corrected p<0.05) and 23 population (OR<1 and corrected p<0.05) variant-enriched features. Furthermore, we accounted for the rarity of population variants by analyzing subsets of rare (allele frequency/AF<0.001) and common (AF > 0.01) variants. We observed a consistent number of features enriched by pathogenic and population variants regardless of the rarity of population variants, highlighting the relevance of those features in variant interpretation. Interestingly, our results differentiated between pathogenic variant-enriched features with a consistent, high effect size (~17-fold for intramolecular disulfide bond) versus those with an increased effect size when compared with rare or common population variants (e.g., protein core residues, OR=3 to 6). Notably, the rare population variants tend to share common features with pathogenic variants, with 14 features enriched with rare variants over common variants coinciding with pathogenic variant-enriched features. Our analyses show a spectrum of the effect size of protein features in variant pathogenicity aligned to the allele frequency spectrum of population variants, highlighting the importance of considering protein features associated with both rarity and pathogenicity in variant interpretation.

Session Title: Molecular Effects of Genetic Variation Poster Session I

PB1057 Consistent effect sizes of molecular quantitative trait loci observed across diverse African populations

Authors:

D. Nachun[1], M. Degorter[1], P. Goddard[1], C. Smith[1], R. Ungar[1], N. Abell[1], B. Balliu[2], D. Gurdasani[3], S. Montgomery[1]; [1]Stanford Univ. Sch. of Med., Stanford, CA, [2]UCLA David Geffen Sch. of Med., Los Angeles, CA, [3]Queen Mary Univ. of London, Cambridge, United Kingdom

Molecular quantitative trait loci (QTLs) have been primarily mapped in subjects of European descent and studies comparing QTLs between populations have been limited to reporting whether a variant has a non-zero effect size in each population. The multivariate adaptive shrinkage model (*mash*) can estimate posterior effect sizes and standard errors for QTLs tested in multiple conditions that incorporate sharing of effects between the conditions, usually for the purpose of testing if a QTL has a non-zero effect in each condition. We have extended the *mash* model to allow us to directly test for differences in QTL effect sizes across populations by using general linear hypothesis testing to fit post-hoc contrasts from posterior effect sizes and covariance matrices for each QTL. We applied this framework to molecular QTLs generated in lymphoblastoid cell lines (LCLs) from six African populations (ESN, GWD, LWK, MKK, MSL, YRI; N = 41 - 166) sampled by the 1000 Genomes and HapMap projects to test for evidence of population differences in QTL effect sizes. We found pairwise significant ($p < 1x10^{-5}$) differences in effect sizes in only 0.07-1.3% of expression QTLs and 0.006-0.18% of splicing QTLs across the six population samples we compared, implying that population differences molecular QTL effect sizes, if they exist, are too small to detect with our current sample sizes. To test if Eurasian admixture, which has been identified in several of the populations sampled for our study, affects gene expression, we inferred global and local West African and Eurasian ancestry in our samples. We used Yoruba (YRI) as a reference for West African ancestry and six additional populations as references for Eurasian ancestry (BEB, STU, CHB, JPT, TSI, CEU) and found that mean global West African ancestry varied between populations, from 79% in MKK to 100% in ESN. We tested for associations between proportion of global and local West African ancestry and gene expression and found expression of 1,393 genes and 248 genes were significantly associated with global and local West African ancestry, respectively. These results demonstrate that while admixture may affect gene expression, differences in admixture between populations does not lead to large differences in QTL effect sizes. Future studies with larger samples sizes will be able to test if smaller population differences in QTL effect sizes exist and if sharing of effect sizes is also seen in lower frequency variants. Data from other cell types and tissues and additional diverse populations will be needed to determine if these findings generalize beyond LCLs and are applicable to other populations.

Session Title: Molecular Effects of Genetic Variation Poster Session II

PB1058 † Context-aware single-cell multiome approach identified smoking and cell-type specific lung cancer susceptibility genes

Authors:

E. Long[1], J. Yin[2], S. Ju Hye[3], Y. Li[1], A. Kane[2], J. Xia[4], Y. Han[5], J. Byun[5], T. Zhang[2], W. Zhao[2], M. Landi[2], N. Rothman[2], Q. Lan[2], Y. Chang[3], C. Amos[5], J. Shi[2], J. Gu[3], E. Kim[3], J. Choi[2]; [1]Chinese Academy of Med. Sci. and Peking Union Med. Coll., Beijing, China, [2]NCI/NIH, Rockville, MD, [3]Yonsei Univ. Coll. of Med., Seoul, Korea, Republic of, [4]Creighton Univ., Omaha, NE, [5]Baylor Coll. of Med., Houston, TX

Genome-wide association studies (GWAS) have identified over fifty genomic loci that are associated with lung cancer risk. However, the genetic mechanisms and target genes underlying these loci are largely unknown, as most variants associated with lung cancer might *cis*-regulate gene expression in a context specific manner. Here, we generated the first barcode-shared transcriptome and chromatin accessibility map of 117,911 human lung cells from sex/age-matched Korean ever- (n=8) and never-smokers (n=12) with an enrichment of epithelial cell types to represent lung cancer relevant exposures and cells of origin. This resource enabled variant-peak colocalization and direct peak-gene linkage with cell-type information, which is used to characterize 2,574 candidate causal variants (CCVs) of 51 loci from 4 different lung cancer GWAS. A total of 323 CCVs (12.5%) of 35 GWAS loci (68.6%) colocalized with an accessible peak from one or more of 23 cell types, indicating an efficient variant prioritization for most GWAS loci. Importantly, 48% of the CCV-colocalizing peaks were only detected in epithelial (28%) or immune (20%) cell types. Epithelial and immune cells also presented the highest trait relevance scores by variant-to-context mapping, highlighting their importance in lung cancer etiology. Notably, accessible peaks of rare epithelial populations implicated as lung cancer origins provided support for multiple GWAS loci. Namely, AT2-proliferating cells (0.13%) colocalized with 17 CCVs from 12 GWAS loci, and CCVs in telomerase-encoding *TERT* overlapped a peak specific to basal cells (1.8%) - self-renewing progenitor cells. Linkage between accessible peaks and gene expression identified 64 high-confidence target genes from 18 lung cancer loci. Notably, 53% of them from 14 GWAS loci were differentially expressed based on smoking status in at least one cell type. Four linked *HLA* genes in the locus at chr6p21.33 contributed to the significantly elevated cellular interactome of MHC-I pathway in smokers, highlighting potential smoking-responsive mechanisms. Moreover, our data suggested that CCVs in linkage disequilibrium could regulate one or more context-dependent target genes via multiple peaks. In the locus at chr8p12, three CCVs were linked to *NRG1* via three separate but co-accessible peaks. In the locus at chr11q23.3, immune cell-specific CCV-colocalizing peaks were correlated with *JAML*, while an adjacent epithelial cell-specific CCV-colocalizing peak was correlated with *MPZL3* expression. Overall, we provide a unique single-cell resource to characterize lung cancer GWAS loci and highlight context-dependent gene regulation underlying the risk.

Session Title: Molecular Effects of Genetic Variation Poster Session III

PB1059 Contribution of genetic ancestry to changes in the urinary metabolites of extremely preterm infants.

Authors:

T. MALAVE-MENDEZ[1], M. Guardado[2], C. Chapin[1], M. Steurer[1], R. Hernandez[3], P. Ballard[1], D. Torgerson[4]; [1]Univ. OF CALIFORNIA SAN FRANCISCO, SAN FRANCISCO, CA, [2]Univ. of California, San Francisco, San Francisco, CA, [3]UCSF, Oakland, CA, [4]UCSF, San Francisco, CA

The human biofluid metabolome is influenced by environmental and genetic factors, the latter of which may include genetic ancestry-linked variation. Genetic ancestry is an indicator of varying patterns of genetic variation that differ by continental origin; however, the contribution of genetic ancestry to the metabolome remains poorly understood. We aimed to identify metabolites that associate with continental genetic ancestry in the urine of extremely premature infants at high risk of developing bronchopulmonary dysplasia. Our study included 171 infants < 29 weeks gestational age from the Trial of Late Surfactant study (TOLSURF), including those of maternal self-reported Hispanic White, non-Hispanic White, or Black/African American race/ethnicity. Untargeted metabolomics was performed using UHPLC-MS/MS on urine from each infant collected at two timepoints between days 6-14 and 23-30 postnatal age. Ancestry proportions were inferred from >800,000 genotypes assuming a 3-population model of continental African (AFR), European (EUR), and Amerindigenous (AMR) ancestry. We then tested for an association between proportions of each genetic ancestry and quantitative measures of individual metabolites using linear regression, adjusting for maternal race/ethnicity, sex, birthweight, TPN status, BPD, and gestational age. Analyses were also performed separately within each racial/ethnic group. Among all infants combined, we identified 130, 72, and 58 metabolites that associated with proportions of AFR, EUR, and AMR ancestry at $p<0.05$, respectively, at timepoint 1, and 64, 33, and 39 metabolites, respectively, at timepoint 2. Twelve metabolites were associated with the same ancestry, and in the same direction at both timepoints. In sub-analyses stratified by race/ethnicity, 16 metabolites were associated with AFR ancestry, 5 with EUR ancestry, and 10 with AMR ancestry at $p<0.05$; 22 of these were associated in the same direction with the same genetic ancestry in >1 racial/ethnic group. Among metabolites that met significance criteria, 8 were involved in metabolic pathways, 2 were xenobiotics, 6 were unknown or partially characterize metabolites, and 6 were involved in other pathways. Our study suggests that genetic ancestry plays a role in the urinary metabolome of premature infants and highlights the importance of considering genetic ancestry in metabolomic studies involving diverse populations.

Session Title: Molecular Effects of Genetic Variation Poster Session I

PB1060 Co-occurrence of Congenital Disorder of Glycosylation Type IIm and Biglycan-Related Disorder in a 3-Year-Old Puerto Rican Male: A Case Report Emphasizing the Role of Biochemical Testing in Confirming Variants of Uncertain Significance

Authors:

E. Albino[1], F. A. Benitez-Rios[2], S. Carlo[3], A. Cornier[4], F. Velez-Bartolomei[5]; [1]Univ. of Puerto Rico - MSC, San Juan, PR, [2]Ponce Hlth.Sci. Univ., Ponce, PR, [3]Ponce Hlth.Sci. Univ., Cabo Rojo, PR, [4]San Jorge Children s & Women s Hosp., San Juan, PR, [5]Hosp. Pediatrico Univ.rio Dr. Antonio Ortiz, San Juan, Puerto Rico

Congenital disorder of glycosylation type IIm (CDG2M) (OMIM #300896) is an X-linked dominant disorder caused by mutations in the *SLC35A2* gene. It is characterized by hypotonia, infantile-onset seizures, EEG abnormalities, dysmorphic facies, developmental delay, intellectual disability, and lack of speech. Biglycan (*BGN*)-related disorder, or spondyloepimetaphyseal dysplasia (SEMDX) (OMIM #300106), is an X-linked recessive disorder caused by mutations in the *BGN* gene. Its clinical features include severe short-trunk dwarfism, brachydactyly, waddling gait, normal facies, and normal intelligence. The prevalence of these disorders is yet to be determined. Definitive diagnosis relies on genetic testing following clinical suspicion. We present the case of a 3-year-old Puerto Rican male diagnosed with CDG2M and SEMDX. Exome sequencing identified a hemizygous likely pathogenic c.351+1G>T (NM_001711.6) in *BGN*, and a hemizygous variant of uncertain significance (VUS) c.194T>C (p.Phe65Ser) (NM_005660.3) in *SLC35A2*. The patient exhibits severe generalized hypotonia, postnatal growth restriction consistent with spondyloepimetaphyseal dysplasia, and global developmental delay. He presents with dysmorphic features including temporal narrowing, a mildly elongated face, flat nasal bridge, low-set ears, and a high-arched palate. Physical examination reveals hypermobile extremities and pale, doughy skin. Management includes D-galactose supplementation, therapy, sleep study, laboratory testing, and regular follow-up visits. Our patient presents with co-occurring CDG2M and SEMDX. CDG2M is associated with mutations in the *SLC35A2* gene located on chromosome Xp11.23. The missense variant (c.194T>C (p.Phe65Ser)) in *SLC35A2* is predicted to cause protein damage, resulting in the observed phenotypic expression. SEMDX, on the other hand, is caused by mutations in the *BGN* gene found on chromosome Xq28. The splice site variant (c.351+1G>T) in *BGN* likely disrupts normal RNA splicing, affecting protein function and/or expression. This case highlights the importance of genetic testing in diagnosing and managing patients with suspected genetic disorders. Furthermore, performing biochemical tests to confirm variants of uncertain significance plays a critical role in establishing accurate diagnoses. Further research is needed to determine the prevalence of these disorders and explore potential treatment options.

Session Title: Molecular Effects of Genetic Variation Poster Session II

PB1061 Coronary artery disease-associated splicing variants in vascular smooth muscle cells

Authors:

N. Perry, M. Murali, J. Korchak, B. Jordan, N. Bhuiyan, D. Lue, K. Hudson, G. Sheynkman, M. Civelek; Univ. of Virginia, Charlottesville, VA

BACKGROUND: Coronary artery disease (CAD) is the leading cause of death in the United States despite advances in lipid-lowering and anti-hypertensive drugs. Heritability estimates for CAD vary from 40% to 70%, suggesting strong genetic contributions to disease risk. Through genome-wide association studies (GWAS), more than 300 loci have been associated with risk for CAD. Despite the success of GWAS, the mechanisms whereby the majority of CAD-associated GWAS loci contribute to disease susceptibility are unknown; however, associated SNPs are predicted to function by regulating targeted effector gene expression in the vascular wall, where smooth muscle cells (SMCs) make up the medial layer. During the initiation and progression of atherosclerosis (the underlying cause of CAD), SMCs exhibit a high level of phenotypic plasticity that leads to both atheroprotective and atherogenic contributions to disease. Understanding the genetic regulatory mechanisms of vascular smooth muscle cells (VSMCs) that underlie GWAS findings will identify susceptible mechanisms in the vessel wall for possible therapeutic targeting of CAD. METHODS: An in-depth characterization of genetic regulation of splicing in aortic SMCs of 151 heart transplant donors from various genetic ancestries was performed, followed by PacBio long-read sequencing of six aortic SMC donors to detect alternative splicing (AS) in full-length isoforms from the long-read sequence data for confirmation of splicing predictions of the 164 colocalized splicing and CAD-associated genes. RESULTS: We identified 164 genes with colocalized genetic signals between splice effect associations (splicing quantitative trait loci (sQTL)) and CAD risk associations. We prioritized sQTLs in or nearby splice acceptor and donor sites at exon skipping or cryptic splice site events within the long read data, and focused only on sQTLs that are explicitly captured in CAD GWAS loci. Six genes (*ITGA1, WIPI1, PARP12, RAB5C, MRAS, PRDM16*) contained SNPs that were both within CAD GWAS loci and highly predicted to affect SMC-specific AS events. At the *PARP12* locus, the risk allele of the CAD-associated missense variant, rs2286196, is associated with an AS event that skips exon 10 and results in the loss of the PARP catalytic domain. An active PARP catalytic domain is required for PARP12 to activate NFkB signaling, a major signaling pathway involved in atherosclerosis. SUMMARY: Our results suggest that the CAD-associated *PARP12* locus may function through AS mechanisms affecting the ability of PARP12 to regulate NFkB signaling and downstream inflammatory contributions to disease development.

Session Title: Molecular Effects of Genetic Variation Poster Session III

PB1062 COVID-eQTL: Genetic dissection of COVID-19 susceptibility

Authors:

J. Lee[1], E. Jeon[1], L. Yu[2], S. Zhao[2], M. Choi[1]; [1]Seoul Natl. Univ., Seoul, Korea, Republic of, [2]Dartmouth Coll., Hanover, NH

A world-wide pandemic caused by SARS-CoV-2 virus provided an unprecedented opportunity to investigate the genetic susceptibility of virus infection and severity determination. For example, genome-wide association analysis on a cohort of more than 100K infected individuals by COVID-19 Host Genetics Initiative has identified genetic variants associated with COVID-19 severity. However, the molecular mechanisms of how genetic variants affect disease severity and prognosis remain to be elucidated. To address this, we took advantage of the multi-omics data of COVID-19 patients with mild or severe symptoms generated by the support of Korean Disease Control and Prevention Agency and performed quantitative trait loci (QTL) analysis. The data set is composed of whole genome sequencing, blood-based single cell RNA-seq, cytokine profiling, TCR/BCR, and HLA data from 313 individuals with mild or severe COVID-19 symptoms over three time points. Our QTL discovery pipeline applied on these multi-omics data uncovered a number of signals that are specific to severity and disease progression from various cell types and serum cytokine levels. This approach identified a number of variant-gene associations that are differentially regulated by disease status and their molecular upstream signals that eventually modulate disease progression in different individuals.

Session Title: Molecular Effects of Genetic Variation Poster Session I

PB1063 CRISPR/Cas9-mediated knock-in of late-onset Alzheimer's disease-risk variant, SHARPIN G186R, lessens the NF-κB pathway and accelerates Aβ secretion

Authors:

Y. Asanomi[1], T. Kimura[1], N. Shimoda[1], D. Shigemizu[1,2], S. Niida[3], K. Ozaki[1,2]; [1]Med. Genome Ctr., Res. Inst., Natl. Ctr. for Geriatrics and Gerontology, Obu, Japan, [2]RIKEN Ctr. for Integrative Med. Sci., Yokohama, Japan, [3]Ctr. for Core Facility Admin., Res. Inst., Natl. Ctr. for Geriatrics and Gerontology, Obu, Japan

Despite the identification of many novel genetic variants associated with late-onset Alzheimer's disease (LOAD), our understanding of the direct biological relevance of these variants to the disease remains largely unclear. In our previous report (Asanomi *et al*., *Mol Med*. 2019), we identified a rare functional variant of *SHARPIN*, rs572750141 (p.Gly186Arg), that is significantly associated with LOAD in the Japanese cohort comprising 4563 LOAD patients and 16459 controls (odds ratio = 6.1). Moreover, recent large-scale meta-analyses of genome-wide association studies have identified other *SHARPIN* common missense variants associated with the risk of LOAD, indicating the important role of *SHARPIN* in LOAD pathogenesis. Although functional analyses of the G186R-type SHARPIN protein in previous studies have revealed aberrant cellular localization of this protein and attenuated activation of the NF-κB pathway, all these analyses used exogenous gene transfer. To elucidate the biological relevance of LOAD-risk variants to the disease pathogenesis, more direct ways of examining the function of mutant SHARPIN protein are desired.

In this study, we used the CRISPR/Cas9 system to perform a knock-in of the LOAD-risk variant into HEK293 cells and generated cell lines with the homozygous SHARPIN G186R mutation. Although the efficiency of the knock-in was modest (<1%), the desired knock-in cells were successfully obtained through high-throughput screening by using a PCR-Invader assay. We then used these knock-in cells to investigate the effects of the *SHARPIN* variant. TNF-α-induced activation of the NF-κB pathway, a central mediator of inflammatory and immune responses, was strongly suppressed in the G186R knock-in cells, but aberrant cellular localization of the mutant protein was not apparent. Previous findings have suggested that neuroinflammation in the central nervous system plays a significant role in the pathogenesis of LOAD. The attenuated NF-κB activity caused by the *SHARPIN* variant may increase the risk of LOAD onset by altering inflammatory and immune responses in the central nervous system. Furthermore, the amount of amyloid-β (Aβ), which was implicated in LOAD pathogenesis, secreted into the culture medium was significantly increased in the G186R knock-in cells, although no significant change in the ratio of Aβ40 to Aβ42 was observed. These findings from the *SHARPIN* knock-in cells provide the importance and new biological insight on SHARPIN functions in the pathogenesis of LOAD. Further investigation of the SHARPIN-associated pathways may elucidate the mechanism underlying the onset of LOAD.

Session Title: Molecular Effects of Genetic Variation Poster Session II

PB1064 CRISPR-mediated allele substitution of a human genetic variant rs10071329 modulates*PPARGC1B* expression and alters brown adipocyte phenotype.

Authors:

S. Kalamajski, M. Huang, D. Coral, P. Franks; Lund Univ., Malmö, Sweden

Human genetic variation in *PPARGC1B* has been associated with adiposity, but the genetic variants that affect *PPARGC1B* expression have not been experimentally validated. Guided by previous observational data, here, we used CRISPR/Cas9 D10A nickase to edit the alleles of a candidate causal genetic variant rs10071329 in human brown adipocyte cell line (hBAs). Changing the rs10071329 genotype from A/A to G/G enhanced *PPARGC1B* expression throughout the adipogenic differentiation, identifying rs10071329 as a *cis*-eQTL. The higher *PPARGC1B* expression in G/G cells coincided with higher accumulation of triglycerides, and higher expression of mitochondria-encoded genes, but without marked effect on adipogenic marker expression. Furthermore, G/G cells had improved basal- and norepinephrine-stimulated mitochondrial respiration, possibly relating to the improved mitochondrial gene expression. The G/G cells also had higher norepinephrine-stimulated glycerol release, indicating improved lipolysis. Altogether, this shows that rs10071329 is a *cis*-eQTL, with the G allele conferring enhanced *PPARGC1B* expression, with consequent improved mitochondrial function and response to norepinephrine in brown adipocytes. This genetic variant, a previously unidentified eQTL, resides downstream of the *PPARGC1B* gene and probably constitutes a part of a gene expression-regulatory enhancer element that could play a causal role in energy metabolism and adiposity. It remains to be shown if people with different rs10071329 genotypes respond differently to e.g. weight loss intervention, where *PPARGC1B* could play an important role in mitochondrial activity.

Session Title: Molecular Effects of Genetic Variation Poster Session III

PB1065 Cross-cohort eQTL analyses of 6,602 multi-ancestry TOPMed whole blood RNA-seq samples uncover regulatory relationships

Authors:

P. Orchard[1], F. Aguet[2], T. Blackwell[3], K. Ardlie[4], A. Smith[3], R. Joehanes[5,6], X. Li[7], M. Wang[8], C-T. Liu[9], A. Saferali[10], J. Wu[8], M. R. G. Taylor[11], N. Heard-Costa[12], H. Tang[8], P. J. Castaldi[13], G. Abecasis[3], J. I. Rotter[14], L. Kachuri[15], D. Levy[5,6], L. M. Raffield[16], L. J. Scott[3], S. C. J. Parker[1], NHLBI TOPMed Consortium; [1]Univ. of Michigan, Ann Arbor, MI, [2]Illumina, Inc., Foster City, CA, [3]Univ. of Michigan Sch. of Publ. Hlth., Ann Arbor, MI, [4]Broad Inst. of MIT and Harvard, Cambridge, MA, [5]Natl. Heart, Lung, and Blood Inst., NIH, Bethesda, MD, [6]Framingham Heart Study, Framingham, MA, [7]Univ. of Arizona, Tucson, AZ, [8]Stanford Univ., Stanford, CA, [9]Boston Univ. Sch. of Publ. Hlth., Boston, MA, [10]Brigham & Women's Hosp., Boston, MA, [11]Univ. of Colorado Anschutz Med. Campus, Aurora, CO, [12]Boston Univ. Chobanian & Avedisian Sch. of Med., Boston, MA, [13]Brigham and Women's Hosp., Boston, MA, [14]Lundquist Inst., Harbor-UCLA Med. Ctr., Torrance, CA, [15]Stanford Univ. Sch. of Med., Stanford, CA, [16]Univ. of North Carolina at Chapel Hill, Chapel Hill, NC

Most genetic variants associated with complex traits occur in non-coding genomic regions and are hypothesized to affect gene expression. To identify variants that regulate gene expression, we performed *cis*- and *trans*-expression quantitative trait locus (*cis*/*trans*-eQTL) analyses using whole blood RNA-seq and whole genome sequencing data from the NHLBI Trans-Omics for Precision Medicine (TOPMed) program from 6,602 samples of predominantly European (68%), African (21%) and Indigenous American (10%) ancestry. We hypothesized that many *trans* signals would overlap *cis* signals, enabling identification of candidate genes mediating *trans* effects.

At a MAF ≥ 0.01, we identified 19,422 genes with ≥ 1 *cis*-eQTL (5% FDR; *cis*-eGenes) and 71,092 total independent *cis*-eQTLs (SuSiE 95% credible sets). At MAF ≥ 0.05, we identified 1,743 *trans*-eGenes. *Trans*-eVariants were enriched for overlap with *cis*-eQTL credible sets (Fisher's exact test against MAF-matched variants; odds ratio = 8.4; $p = 3x10^{-45}$; 30% of unique *trans*-eVariants overlap *cis*-eQTL) and *cis*-eGenes for *cis*-eQTL overlapping *trans*-eQTL were 3.5-fold enriched for transcription factor genes ($p < 5x10^{-4}$). 167 variants were *trans*-eQTL for >1 gene (1,075 total genes). For example, one *cis*-eVariant for *ERN1*, which encodes endonuclease IRE1a, was a *trans*-eVariant or in high LD (>0.9) with a *trans*-eVariant for thirteen *trans*-eGenes, including known *ERN1* downstream target *XBP1* and *XBP1* target genes including *DNAJB9*. *ERN1* is a regulator of the endoplasmic reticulum (ER) stress response, and five of the twelve *trans*-eGenes in the KEGG database were in the "Protein processing in ER" pathway (31.6-fold enrichment; nominal $p=1.8x10^{-7}$). Among ER response pathways IRE1a-XBP1 is the most highly conserved, with an emerging role in regulation of inflammation and immune response.

To identify *trans*-eGenes that may share multiple signals with a potentially regulatory *cis*-eGene, we finemapped *trans*-eQTL signals within the 2Mb window centered on each *trans*-eGene's lead *trans*-eVariant. Within the 2Mb windows, 300 of the 1,743 *trans*-eGenes had >1 *trans*-eQTL (2,080 total *trans*-eQTL credible sets). We found 31 *cis*-eGenes with > 1 *cis*-eQTL signal colocalizing with > 1 *trans*-eQTL signal from at least one *trans*-eGene (145 unique *cis*-eGene - *trans*-eGene pairs). For example, *trans*-eGene *BTN3A3* showed 4 *trans*-eQTL credible sets that colocalized with 4 *cis*-eQTL credible sets for its known regulator *NLRC5*. This example provides proof-of-principle for our hypothesis.

In summary, this dataset demonstrates the utility of large eQTL studies to provide insight into regulatory pathways involving *trans*-eQTLs mediated by *cis*-eGenes.

Session Title: Molecular Effects of Genetic Variation Poster Session I

PB1066 Cross-protein transfer learning using deep mutational scanning data substantially improves proteome-wide disease variant prediction

Authors:

C. Ye, M. Jagota, C. Albors, R. Rastogi, A. Koehl, N. Ioannidis, Y. S. Song; Univ. of California, Berkeley, Berkeley, CA

Genetic variations within the human genome significantly influence individual disease susceptibility. Numerous computational strategies have been developed to predict missense variant impacts, using a variety of predictive signals. Nevertheless, the etiological impacts of a vast number of variants remain unknown. Deep Mutational Scanning (DMS) experiments, which perform saturation mutagenesis on proteins, can assess the functional impacts of a large proportion of possible mutations for each protein. However, DMS experiments are currently limited to a select few proteins and do not scale easily to the entire proteome. To address this, we introduce a robust learning framework that harnesses DMS data to construct precise computational predictors of disease variant pathogenicity. We train a cross-protein transfer (CPT) model using DMS data from only five proteins and achieve state-of-the-art performance on clinical variant interpretation for unseen proteins across the human proteome. High sensitivity is crucial for clinical applications and our model CPT-1 particularly excels in this regime, substantially improving on recently developed deep learning models (such as ESM and EVE) for protein sequences. Furthermore, for proteins not used to train REVEL, a supervised method widely used by clinicians, we show that CPT-1 compares favorably with REVEL.

Our framework combines predictive features derived from general protein sequence models, vertebrate sequence alignments, and AlphaFold2 structures; and it is adaptable to the future inclusion of other sources of information. We find that vertebrate alignments provide a strong signal for variant effect prediction that is non-redundant with recent deep learning-based models trained on massive amounts of protein sequence data. The utility of integrating vertebrate alignments across the human proteome points to exciting future directions. There are ongoing efforts to sequence a large number of vertebrate genomes; as these data become available, we anticipate that more powerful models could be applied to deeper vertebrate alignments.

Session Title: Molecular Effects of Genetic Variation Poster Session II

PB1067 De novo decipherment of genetic architecture with language models

Authors:

S. Bhate; Broad Inst. of MIT and Harvard, Cambridge, MA

We observe that the genetic architecture of single-cell gene expression exhibits consistency with phenotypic geometry and with genomic syntax. Specifically, molecularly similar cell types have similar genetic associations, and syntactically similar genomic loci have similar phenotypic associations. We show that assuming, a priori, consistency of genetic architecture with phenotypic geometry and genomic syntax enables identifying quantitative trait loci (QTLs) for phenotypes with a geometric structure de novo: from only phenotypic data and the reference genome sequence - without genetic variation, annotations or external data. We develop Phenotype Syntax Alignment (PSA), an optimization algorithm that, searches de novo through possible genetic architectures for a phenotype in a reference genome to find one consistent with phenotypic geometry and with genomic syntax, using language models. PSA identifies de novo from the geometry of single-cell gene expression, cis-, trans- and causal expression QTLs in a cell-type, tissue and gene specific manner to 10kb resolution, as well as regulators of gene expression empirically validated by CRISPR perturbation. Moreover, PSA identifies, de novo from immunofluorescence imaging, causal genetic pathways regulating the tumor microenvironment as well as QTLs for its cellular composition. Consistency with phenotypic geometry and genomic syntax is thus a general principle underlying the genetic architectures of molecular phenotypes enabling their de novo decipherment.

Session Title: Molecular Effects of Genetic Variation Poster Session III

PB1068 *De novo* variants in *PLCG1* are associated with neurological issues and other clinical features.

Authors:

M. Ma[1], Y. Zheng[2], S. Lu[3], X. Pan[3], K. Worley[4], L. Blieden[1], A. Allworth[5], D. Li-Kroeger[3], J. Rosenfeld[3], D. Dutta[3], S. Yamamoto[3], M. Wangler[6], I. Glass[7], P. Prontera[8], S. Lalani[9], H. Bellen[3]; [1]Baylor Coll. of Med., HOUSTON, TX, [2]Xiamen Univ., Xiamen, China, [3]Baylor Coll. of Med., Houston, TX, [4]Baylor Coll. Med., Houston, TX, [5]Univ. of Washington, Seattle, WA, [6]Molecular and Human Genetics, Baylor Coll. of Me, Houston, TX, [7]Univ of Washington, Seattle, WA, [8]Med. Genetics Unit, Univ. and Hosp. of Perugia, Perugia, Italy, [9]Baylor Coll. Med., Sugar Land, TX

Phospholipase C isozymes hydrolyze phosphatidylinositol 4,5-bisphosphate into inositol 1,4,5-trisphosphate and diacylglycerol, important signaling molecules involved in many cellular processes. *PLCG1* encodes the PLCγ1 isozyme that can be activated by receptor tyrosine kinases. *PLCG1* is broadly expressed, and hyperactive somatic mutations of *PLCG1* are frequently observed in multiple cancers. However, germline variants of *PLCG1* have so far not been associated with Mendelian diseases. Here, we report three independent individuals with *de novo* missense variants in *PLCG1* (p.D1019G, p.H380R, and p.D1165G), who present with hearing, vision, heart, brain structure and immune system issues as well as other phenotypes. However, they do not share all phenotypes. To assess the impact of the three variants *in vivo*, we generated and modeled the variants in the *Drosophila* orthologue, *small wing* (*sl*) (p.D1041G, p.H384R, and p.D1184G). We created a sl^{T2A} allele by inserting a splicing acceptor-T2A-GAL4-polyA cassette into the first coding intron of *sl*. The polyA tail arrests transcription and creates a strong loss-of-function mutation. The GAL4 that is produced allows us to assess the endogenous expression pattern of *sl* by using a *UAS-fluorescent reporter*. We show that *sl* is broadly expressed, including in wing discs, eye discs, and a subset of neurons and glia. Loss of *sl* is known to cause a reduced wing size, ectopic wing veins and additional photoreceptors which are all observed in sl^{T2A}. Although the mutant flies are homozygous viable we show that they have a reduced lifespan and display severe locomotion defects. Expressing wild type *UAS-sl* fully rescues the loss-of-function phenotypes of sl^{T2A} mutants. In contrast, the three *UAS-sl* variants cause lethality, suggesting that the variants are toxic and cause a gain-of-function or neomorphic effect. We also performed ectopic expression assays to test the human *PLCG1* variants in fly. Expressing the p.H380R cDNA causes mild phenotypes indistinguishable from expressing the reference cDNA. However, expressing the p.D1041G or p.D1184G cDNA results in phenotypes that differ from those observed in loss-of-function mutants but resemble the phenotypes of expressing a established hyperactive variant, again arguing that they are gain-of-function rather than dominant-negative alleles. In summary, our data indicate that the p.D1041G and p.D1184G are strong gain-of-function variants while the p.H380R is a mild gain-of-function variant. Our data also argue that these *de novo* missense variants in *PLCG1* may be causative for the features observed in the probands.

Session Title: Molecular Effects of Genetic Variation Poster Session I

PB1069 Deciphering the molecular and cellular impact of epilepsy-associated *GNAI1* variants.

Authors:

E. Fathi[1], H. Chen[2], E. Almanza Fuerte[1], L. Sadleir[3], M. Innes[4], S. Sheppard[5], H. Mefford[6]; [1]St. Jude Children's Res. Hosp., Memphis, TN, [2]St Jude Children s Res. Hosp., Memphis, TN, [3]Univ. of Otago, Wellington, New Zealand, [4]Univ. of Calgary, Alberta, AB, Canada, [5]NICHD/NIH, Rockville, MD, [6]St. Jude Children`s Res. Hosp., Memphis, TN

Developmental and epileptic encephalopathies (DEE) are a genetically heterogenous group of severe epilepsy disorders characterized by intellectual disability, developmental delay, intractable seizures, and autistic features. *De novo* mutations in the *GNAI1* gene, encoding the inhibitory G protein subunit alpha I1 ($G\alpha_{i1}$), are one genetic cause of DEE for which there is no effective treatment. When bound to GDP, $G\alpha_{i1}$ forms an inactive heterotrimer with β and γ G-protein subunits. Upon activation, GDP is exchanged for GTP, the α subunit dissociates from the βγ complex, and each (α and βγ) are now active and can act on downstream signaling pathways including cAMP, mTOR, and MAPK/ERK. Most pathogenic variants are missense variants that affect the GTP binding pocket; *de novo* truncating variants in *GNAI1* have not yet been reported. The impact of reported pathogenic variants on protein function (gain vs. loss of function) is not known but is critical for targeted therapy development. We hypothesize that pathogenic variants in *GNAI1* impair GTP binding and hydrolysis and disrupt downstream signaling pathways.To investigate the molecular and cellular function of *GNAI1* disease-causing variants, we created heterozygous CRISPR-engineered U87-MG (glioblastoma origin) cell lines with a pathogenic variant (Gly40Cys, Thr48Lys, Lys270Arg, Gln172del) or a premature truncation (Ile49Ter) as a control for LOF. Our preliminarily result showed no significant change in mRNA and protein expression of CRISPR-engineered cells (except early termination) when compared to control. We produced purified $G\alpha_{i1}$ WT and mutant proteins to measure GTPase activity using a fluorescent GTP analog, BODIPY-GTP. While $G\alpha_{i1}$ WT binds to BODIPY-GTP, the mutant proteins are incapable of binding to GTP and hydrolyzing it. These data suggest that the clinically observed $G\alpha_{i1}$ pathogenic variants in *GNAI1*-related DEE lose their ability to bind to GTP and hydrolyze it, which is critical in modulating important pathways such as mTOR and MAPK/ERK. Indeed, in heterozygous cells lines we saw significant increase in phosphorylated S6 in CRISPR-engineered cells, indicating activation of the mTOR pathway.We will validate our preliminary results using *in vitro* studies in 2D/3D neurons from patient cells. This study will uncover the molecular mechanism of the *GNAI1* pathogenic variants, which will be critical for developing treatments. Our engineered and patient patient-derived cells represent a valuable tool to characterize the effect of pathogenic variants on $G\alpha_{i1}$ function and the impact on downstream cellular pathways.

Session Title: Molecular Effects of Genetic Variation Poster Session II

PB1070 Deciphering the molecular basis of a *GRIA2* related neurodevelopmental disorder

Authors:

R. Segel[1,2], L. Berkun[1], A. Porecki[3], Y. Stern-Bach[3]; [1]Med. Genetics Inst., Shaare Zedek Med. Ctr., Jerusalem, Israel, [2]Faculty of Med., the Hebrew Univ., Jerusalem, Israel, [3]Dept. of Biochemistry & Molecular Biology, IMRIC Faculty of Med., The Hebrew Univ. of Jerusalem, Jerusalem, Israel

GRIA2 encodes GluA2, a central component of AMPA-type glutamate receptors that play a crucial role in fast excitatory neurotransmission in the brain. These receptors facilitate synapse formation and trigger activity-dependent processes underlying learning and memory. Mutations in *GRIA2* have been associated with intellectual disability (ID) and intractable seizures. In this study, we investigated a de novo missense variant (L772W) in *GRIA2* identified through exome sequencing in a girl with severe ID and microcephaly, but no seizures. The patient exhibited significant developmental delay at the age of 1 year with limited abilities to sit unsupported, was non-verbal, and had a short attention span. She began walking at the age of 3 years, but at the age of 5 she still had no words, showed autistic behavior but still had no seizures. Physical characteristics were normal, including facial features, weight, length, and head circumference. She had cortical visual impairment and mild ventricular enlargement in brain imaging. Genetic analysis included chromosomal microarray analysis and trio whole-exome sequencing, which revealed *de novo* missense variant in *GRIA2:* c.2315T>G, p.Leu772Trp. This variant was classified as a variant of uncertain significance based on ACMG criteria. Functional analysis in heterologous cells demonstrated that the mutated protein exhibited normal expression level, membrane localization, and maturation into functional receptors. However, the mutation altered receptor kinetics, resulting in faster desensitization compared to wild-type receptors. This alteration led to a shortened open channel duration, ultimately resulting in reduced overall receptor activity. The reduced channel activity caused by the variant may explain the patient's phenotype through decreased overall excitatory neurotransmission. Alternatively, reduced GluA2 activity might trigger compensatory mechanisms, leading to increased excitatory activity through calcium-permeable AMPA receptors formed by the GluA1 unit, potentially causing cell death. Further investigations are necessary to differentiate between these possibilities for better understanding of the underlying mechanisms.

Session Title: Molecular Effects of Genetic Variation Poster Session III

PB1071 Deep mutational scan of TAP2 reveals residues of functional importance for MHC-I cell-surface presentation.

Authors:

G. Nagy, M. Diabate, R. Hardy, H. Jeon, T. Banerjee, S. Dhar, K. Burgess, D. Chung, P. N. Tsichlis, J. D. Parvin; The Ohio State Univ., Columbus, OH

Defects in cell-surface presentation of MHC-I bound to peptide can lead to immune evasion of cancer cells, conferring a worse prognosis for the cancer patient. One mechanism by which MHC-I is downregulated on the cell surface is through impaired function of the Transporter Associated with Antigen Processing (TAP) complex, which contains a heterodimer of TAP1 and TAP2. Exploring how to upregulate or restore functional expression of the TAP complex in tumors with nonfunctional TAP could improve the response of the immune system to a tumor cell. Nonfunctional TAP2 may be due to mutation or post-translational modification (PTM), and functional evaluation of missense variants and PTM mimetics can illuminate the functional constraints of TAP2. Here we present a deep mutational scan of TAP2, in which an amino acid within the protein is exchanged for all twenty possible substitutions (nineteen other amino acids and one stop codon). We have targeted 96 amino acid residues for analysis, and we have functional results for 878 missense variants and 24 nonsense variants of TAP2 with clear hyper- and hypo- functional outliers identified. The functional effects of selected variants, both hypo- and hyper-active, were confirmed when tested singly for MHC-I localization. Some of these functional outliers suggest phosphorylation changes that can regulate TAP2 activity. This is one of the first deep mutational scans of a protein involved in MHC-I cell-surface presentation. Identifying PTMs that modify TAP transporter activity may reveal regulatory processes that can be targeted for therapy that would promote an increase in MHC-I cell-surface presentation and improved immune response to tumors and increased efficacy of immunotherapy.

Session Title: Molecular Effects of Genetic Variation Poster Session I

PB1072 Defining critical *MECP2* cis-regulatory elements towards identifying genetic candidates for male-biased autism

Authors:

R. Meyer-Schuman[1,2], H. Zoghbi[1,2,3,4]; [1]Dept. of Human and Molecular Genetics, Baylor Coll. of Med., Houston, TX, [2]Jan and Dan Duncan Neurological Res. Inst. at Texas Children's Hosp., Houston, TX, [3]Howard Hughes Med. Inst., Baylor Coll. of Med., Houston, TX, [4]Dept. of Pediatrics, Baylor Coll. of Med., Houston, TX

The levels of methyl-CpG binding protein 2 (MeCP2), encoded by the X-chromosome gene *MECP2*, must be precisely controlled in the developing brain. Loss-of-function mutations in *MECP2* cause the progressive neurological disease Rett syndrome in females with mosaic expression and males with hypomorphic mutations. Gain-of-function *MECP2* duplications cause an equally severe neurodevelopmental disorder primarily in males. In mice, even partial loss-of-function or partial gain-of-function *MECP2* mutations cause neurological phenotypes. Male mice with a 30% decrease or 50% increase in MeCP2 levels show autism-like phenotypes, including altered anxiety, activity, and social behavior. This sensitivity to MeCP2 dosage indicates that *MECP2* gene expression must be tightly regulated. It also suggests that mutations in *cis*- or *trans*- acting factors that modulate MeCP2 levels could be candidates for autism phenotypes in humans. Such mutations may disproportionately impact males, who have only one copy of *MECP2* and are thus more sensitive to even modest changes in MeCP2 levels.

Previous studies identified four mouse *Mecp2* cis-regulatory elements (CREs), two of which were only present in mice and two of which were conserved in humans. We hypothesize that additional CREs buffer *MECP2* expression in humans, which may have evolved since the human and mouse lineages diverged. Multiple human brain ATAC-seq datasets were analyzed to identify two regions of open chromatin within or near the *MECP2* locus. Each corresponding sequence was deleted in human induced pluripotent stem cells (iPSCs) using CRISPR-Cas9 mediated genome editing. These iPSCs were then differentiated into glutamatergic neurons and *MECP2* RNA and protein levels were assessed. Preliminary data indicates that one candidate CRE acts as a *MECP2* repressor element and the other as an enhancer. Future work will further pinpoint the critical regulatory sequences within these larger elements, as well as their associated transcription factors.

This study will ultimately provide a foundation for identifying pathogenic non-coding mutations that change *MECP2* levels, as well as mutations in transcription factors that regulate *MECP2*. Such mutations are likely to cause neurodevelopmental phenotypes in humans and may contribute to some of the observed sex bias in neurological conditions like autism spectrum disorder.

Session Title: Molecular Effects of Genetic Variation Poster Session II

PB1073 Defining the spectrum of medically actionable variation in the HostSeq Biobank

Authors:

E. Frangione[1,2], S. Casalino[1,2], N. Aujla[1,2], D. Di Iorio[1,2], J. De Los Rios[3], R. Mahajan[1,2,3], C. Fung[1,2], G. Morgan[1,2,3], M. Scott[1,2,3], D. Wolday[1,2], J. Young[1,2], B. Dickson[1], S. Arnoldo[3,4], E. Bearss[1], A. Binnie[3,4], B. Borgundvaag[1,3], H. Chertkow[5], M. Clausen[6], M. Dagher[7], L. Devine[1,3], S. M. Friedman[1,8], A-C. Gringas[1,2,3], L. Goneau[9], D. Kaushik[4], Z. Khan[10], E. Lapadula[1,2], G. MacDonald[1,3], T. Mazzulli[1,3], A. McGeer[1,2,3], S. McLeod[1,3], C. Mighton[1,2,3,6], T. J. Pugh[8,11], D. Richardson[4], J. Simpson[8,11], S. Stern[10], L. Strug[12], A. Taher[8,3,10], I. Wong[10], N. Zarei[10], S. Haller[10], S. Thakur[1], E. Greenfield[1,3], Y. Bombard[3,6], A. Noor[1,3], H. Faghfoury[8], J. Taher[1,3], J. Lerner-Ellis[1,2,3]; [1]Mount Sinai Hosp., Sinai Hlth., Toronto, ON, Canada, [2]Lunenfeld-Tanenbaum Res. Inst., Sinai Hlth., Toronto, ON, Canada, [3]Univ. of Toronto, Toronto, ON, Canada, [4]William Osler Hlth.System, Brampton, ON, Canada, [5]Baycrest Hlth.Sci., Toronto, ON, Canada, [6]Unity Hlth.System, Toronto, ON, Canada, [7]Women's Hlth.Coll., Toronto, ON, Canada, [8]Univ. Hlth.Network, Toronto, ON, Canada, [9]Dynacare Med. Lab., Toronto, ON, Canada, [10]Mackenzie Hlth., Richmond Hill, ON, Canada, [11]Ontario Inst. for Cancer Res., Toronto, ON, Canada, [12]The Hosp. for Sick Children, Toronto, ON, Canada

Background: DNA biobanks have the potential to identify medically actionable variation that can help explain disease risk, susceptibility or underlying conditions in research study participants. In the GENCOV/HostSeq study, participants were able to receive their genetic results back during the COVID-19 pandemic. Databases such as ClinVar provide clinical classifications to help inform variant interpretations. However, evidence from ClinVar relies on aggregated data across multiple submitters, often with conflicting interpretations. ClinVar is also replete with false positive and negative classifications and captures only a limited proportion of human variation. Here we summarize medically actionable pathogenic variation in HostSeq participants using a clinically derived variant classification pipeline to determine the relative concordance to ClinVar classifications.
Aims: 1) Provide a summary of the frequency of medically actionable genetic variants identified in HostSeq participants. 2) Evaluate the accuracy and concordance of ClinVar interpretations of pathogenicity to variant classifications assessed in-house.
Methods: Variants from the first 1281 individuals in the cohort were annotated with ANNOVAR to detect those found in ClinVar, and filtered to those classified as pathogenic by two or more reputable laboratories. The ACMG secondary findings v3.0 panel and a custom carrier screening panel were used. Variants were analyzed in-house following ACMG criteria and compared to ClinVar. The frequency of pathogenic variation was analyzed by ancestry.
Results: In total, 1006 variants from 1281 individuals were identified with a pathogenic ClinVar classification after filtration. Of these, 373 variants were observed exactly once. There were 42 variants detected from the ACMG panel (0 homozygous) and 471 variants from the carrier screening panel (3 homozygous). There were 689 individuals (54%) that had at least one pathogenic variant, with an average of ~1.5 per participant. Two individuals were compound heterozygous for a pathogenic variant. To date, our in-house pipeline confirmed 1523 pathogenic variants from 392 individuals, 1321 of which were identified in ClinVar (87%). However, only 500 of these variants had a concordant pathogenic classification in ClinVar (38%).
Conclusion: This study summarizes pathogenic variation using a clinically curated pipeline as compared to ClinVar in an ostensibly healthy population of Ontario adults. Our in-house pipeline was more accurate for the detection and classification of pathogenic variation. Accurate interpretations are crucial for diagnosis, clinical correlation and management.

Session Title: Molecular Effects of Genetic Variation Poster Session III

PB1074 Determining pathogenicity of Variants of Uncertain Significance in genes the Sonic Hedgehog Pathway via high-throughput experimental approaches.

Authors:

D. Baldridge, J. Shepherdson, B. Cohen; Washington Univ. in St. Louis, Saint Louis, MO

One major limitation for the molecular diagnosis of individuals with suspected monogenic disorders is the abundance of Variants of Uncertain Significance (VUS) due to insufficient evidence for determining if variants are pathogenic or benign. High-throughput functional assessment of variant effects offers a scalable solution to this critical problem in genomic medicine. We demonstrate the value of these studies by deploying recently developed experimental approaches involving the simultaneous assessment of hundreds to thousands of variants in a protein of interest, using reliable, carefully calibrated, cell-based assays. We have successfully completed a parallelized mutational scan for nearly 200 clinically observed variants in the gene, *GLI2*, part of the sonic hedgehog signaling pathway. Pathogenic variants in the transcription factor *GLI2* cause Culler-Jones syndrome which involves endocrine and skeletal patterning abnormalities. Using cells engineered with a GFP reporter and transduced with lentivirus expressing *GLI2* variants, we conducted a SortSeq experiment, coupling fluorescence-activated cell sorting (FACS) to sequencing, to demonstrate the functional effects of these variants. Our results are highly reproducible and show near perfect discrimination of known pathogenic and benign variants in *GLI2*. We use this system to assay all 170 clinically relevant variants in *GLI2* that were submitted to ClinVar, including about 100 missense VUS, many of which may be reclassified as likely pathogenic or benign based on these results. We also identify potentially mis-classified variants and emphasize limitations of the current “gold standard” ACMG variant classification approach. Further, we provide evidence to support the expansion of this approach to include at least three other clinically relevant genes in the sonic hedgehog pathway, demonstrating the feasibility for a pathway-based approach for scaling up mutational scanning to study many genes in the genome.

Session Title: Molecular Effects of Genetic Variation Poster Session I

PB1075 Development of the first two Charcot-Marie-Tooth Disease Type 4C (*SH3TC2*) models using CRISPR-Cas9 Base Editing Strategies on human induced pluripotent stem cells (hiPSCs).

Authors:

C. Loret[1], A. Pauset[2,3], P-A. Faye[1,4], V. Prouzet-Mauléon[2,3], A. Nizou[1], F. Miressi[1], F. Sturtz[1,4], F. Favreau[1,4], B. Turcq[2,3], A-S. Lia[1,4,5]; [1]NeurIT UR 20218, GEIST Inst., Univ. of Limoges, Limoges, France, [2]CRISP'edit, TBMCore UAR CNRS 3427, US Inserm 005, Univ. of Bordeaux, Bordeaux, France, [3]Modeling transformation and resistance in leukemia, BRIC Inserm U1312, Univ. of Bordeaux, Bordeaux, France, [4]Dept. of Biochemistry and Molecular Genetics, CHU Limoges, Limoges, France, [5]Dept. of Bioinformatics, CHU Limoges, Limoges, France

The Charcot-Marie-Tooth disease (CMT) is the most prevalent inherited peripheral neuropathy. More than 100 altered genes have been identified as responsible for this disease. Most of them, such as *SH3TC2*, the most frequently mutated gene for the demyelinated autosomal recessive CMT form (AR-CMT), are expressed predominantly in neuronal cells. To study the pathophysiology of the disease and to test new therapeutic approaches, it is crucial to develop appropriate cellular models. Human induced pluripotent stem cells (hiPSCs) offer a powerful tool to study neuropathological disorders when the affected cells are inaccessible in patients.
However, the hiPSCs reprogramming is a time-intensive process, and clones generated for various alterations possess different genetic background. Furthermore, to our knowledge, there is currently no hiPSCs model available for CMT4C corresponding to *SH3TC2* variations. Our goal was to generate the first two *SH3TC2* models starting from one of our healthy individuals' iPSCs using the CRISPR-Cas9 technology: a model with the variation c.2860C>T, resulting in p.Arg954* and another one with the variation c.211C>T, resulting in p.Gln71*.
We selected a CRISPR-Cas9 Cytosine Base Editing (CBE) approach, facilitating the specific conversion of a base pair (C>T) in the targeted DNA sequence. As hiPSCs are time-consuming and expensive to grow and to engineer, it is quite challenging to test different pairs of sgRNA-CBE for each desired alteration. Herein, we decided to optimize the generation of variations in *SH3TC2* using CBE, on cultivable and transfectable HEK-293T cells to quickly and easily evaluate different gRNA-CBE possibilities.
Subsequently, we applied the selected CRISPR-Cas9 Base Editing strategies to hiPSCs, derived from a healthy individual, we achieved 90.5% and 93% On-Target activity for c.211C>T, p.Gln71* and c.2860C>T, p.Arg954* respectively, ensuring correct editing in at least one allele.
In conclusion, we present here the creation of the first two *SH3TC2* hiPSCs models using CRISPR-Cas9 Base Editing. These hiPSCs models will be then differentiated into neuronal cells in order to investigate the pathophysiology of CMT4C, they would also be interesting models to test therapeutic strategies.

Session Title: Molecular Effects of Genetic Variation Poster Session II

PB1076 DGKk loss of function contributes to Fragile X syndrome and other intellectual disabilities.

Authors:

O. Cakil[1], B. Zambo[1], N. Drouot[1], F. Tran Mau-Them[2,3], P. Edery[4], A. Accogli[5,6], V. Salpietro[7,8], M. Choi[9], S. Jang[9], A. Piton[1,10,11], H. Moine[1]; [1]Inst. de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), CNRS UMR7104, INSERM U1258, Université de Strasbourg, Illkirch Graffenstaden, France, [2]Unité Fonctionelle Innovation en Diagnostic Génomique des maladies rares, INSERM UMR1231 GAD, CHU Dijon Bourgogne, Dijon, France, [3]INSERM UMR1231 GAD, F-21000, Dijon, France, [4]Service de Génétique, CHU, Lyon, France, [5]Dept. of Med. Genetics of Specialized Med., McGill Univ., Montreal, QC, Canada, [6]Dept. of Human Genetics, McGill Univ., Montreal, QC, Canada, [7]Dept. of Pediatrics, Univ. of L'Aquila, L'Alquila, Italy, [8]Dept. of Neuromuscular Disease, UCL Inst. of Neurology, London, United Kingdom, [9]Dept. of BioMed. Sci., Seoul Natl. Univ. Sch. of Med., Seoul, Korea, Republic of, [10]Unité de Génétique Moléculaire, IGMA, Hôpitaux Univ.ire de Strasbourg, Strasbourg, France, [11]Inst. Univ.ire de France, Paris, France

Fragile X syndrome (FXS) is a neonatal- to child-onset neurodevelopmental disorder characterized by intellectual disability, abnormal behaviors and characteristic physical features. FXS results from the loss of expression or function of the *FMR1* gene encoding the RNA binding protein FMRP with an impact on the post-transcriptional control of hundreds of neuronal mRNA targets. Among its numerous mRNA targets, FMRP most strongly interacts with diacylglycerol kinase kappa (DGKk) mRNA in cortical neurons to promote DGKk protein synthesis. DGKk converts diacylglycerol (DAG) to phosphatidic acid (PA), two signaling lipids regulating cytoskeletal rearrangement and protein synthesis downstream of G-protein coupled receptors. In FXS condition, the absence of FMRP leads to the defect of expression of DGKk associated with an imbalance of DAG and PA in the brain of FXS patients and the *Fmr1*-KO mouse model. Re-expression of DGKk with adeno-associated viruses (AAVs) in the *Fmr1*-KO mouse corrects its behavioral phenotypes. To determine the role of DGKk in FXS condition, here we analyzed the phenotypes of *Dgkk*-KO mouse model and show that it shares behavioral (hyperactivity, stereotypies) and physiological (overgrowth) similarities with the *Fmr1*-KO model suggesting that the absence of DGKk is sufficient to recapitulate core FXS phenotypes. Analyses of *Dgkk*-KO neurons revealed overactivation of the DAG signaling in association with abnormal dendritic spine morphology, deregulated synaptic activity and increased rate of neuronal protein synthesis, hallmarks of FXS. To date, the *DGKk* gene has never been directly associated with a pathology, potentially due to its misannotation until recently as a non-coding gene. Analyses of patient exomes (and NGS) identified *DGKk* missense variants in unrelated patients with intellectual disability or cardiomyopathy. Functional analysis of these *DGKk* variants in non-neuronal and neuronal cells showed that a number of them respectively have membrane or dendritic mislocalization, suggesting an impairment of *DGKk* function correlates with patient condition. Altogether our data support the involvement of *DGKk* in intellectual disability conditions, including FXS.

Session Title: Molecular Effects of Genetic Variation Poster Session III

PB1077 Discovery and functional validation of genes associated with muscular fitness and cardiometabolic traits.

Authors:

T. Schnurr[1], C. Jin[1], E. Bielczyk-Maczynska[1], M. Slocumb[1], P. Saliba-Gustafsson[1], M. Gloudemans[1], T. Hansen[2], E. Ashley[1], I. Carcamo-Orive[1], J. W. Knowles[1]; [1]Stanford Univ., Stanford, CA, [2]Univ. of Copenhagen, Copenhagen, Denmark

Background: Handgrip strength is a proxy for muscular fitness and is associated with insulin resistance, glucose metabolism, and metabolic disease. Genome-wide association studies have identified more than 100 loci associated with handgrip strength. Genetic correlations with handgrip strength are enriched for cardiometabolic traits, and Mendelian randomization suggests a causal effect of handgrip strength on cardiometabolic traits. The mechanisms linking muscular fitness to cardiometabolic health outcomes are incompletely understood.
Aim: To apply colocalization analysis to identify candidate causal genes for muscular fitness and explore their overlap with cardiometabolic traits by integration of expression and splicing quantitative trait loci (QTLs) for handgrip strength and cardiometabolic traits in skeletal muscle tissue, and to functionally validate these candidate genes in an immortalized human skeletal muscle cell line (HMCL-7304).
Results: We identified three muscle-specific candidate genes by integrating our custom LD-adjusted colocalization analysis for handgrip strength and nine cardiometabolic traits combined with expression and splicing QTLs from metabolically relevant tissues from GTEx v8. For functional analysis, we successfully established stable HMCL-7304 cell lines for gene knockdown (using CRISPRi) and cDNA overexpression of candidate genes. Preliminary results from RNA-seq highlight shared pathways across the candidate genes including ATP processes, cell adhesion and calcium ion binding.
Conclusion and next steps: We identified candidate genes for handgrip strength with relevance for cardiometabolic traits. Our preliminary data from functional analysis in an immortalized human skeletal muscle cell line suggests that shared pathways include ATP processes, cell adhesion and calcium ion binding. We are conducting functional assays with focus on metabolic readouts, mitochondrial function, and contractility.

Session Title: Molecular Effects of Genetic Variation Poster Session I

PB1078 Discovery and validation of novel insulin resistance loci implicated in adipocyte function.

Authors:

P. Toh[1], M. García-Ureña[1], T. Kilpeläinen[2]; [1]Novo Nordisk Fndn. Ctr. for Basic Metabolic Res., Copenhagen, Denmark, [2]Univ. of Copenhagen, Copenhagen, Denmark

Insulin resistance is a pathological condition where cells are unable to respond to insulin, resulting in elevated blood glucose levels. It presents an increased risk of developing various cardiometabolic diseases such as type 2 diabetes, a significantly prevalent disease as the global population gradually becomes more obese. Previous studies have identified 53 insulin resistance loci by triangulating genome-wide association studies (GWAS) results for three hallmark traits of insulin resistance: higher fasting insulin, higher triglycerides, and lower HDL cholesterol. Many of the identified loci were implicated in dysfunctional peripheral adipose storage capacity and adipocyte function. We aim to discover further insulin resistance loci by leveraging the results from the latest and largest GWAS and validate the function of the prioritised loci in SGBS preadipocytes. We identified 210 independent insulin resistance loci with joint genome-wide significant association ($P<5x10^{-8}$) with increased fasting insulin adjusted for BMI, increased triglycerides, and decreased HDL cholesterol. We prioritised novel insulin resistance genes in the loci based on differential expression in the adipose tissue (eQTL) and expression in adipocytes or adipose stem and progenitor cells. Five genes (*TNFAIP8*, *ZNF703*, *KCNU1*, *MT1HL1* and *SPON1*) were selected for further validation *in vitro* using SGBS preadipocytes. We observed that *MT1HL1* and *ZNF703* expression decreases during differentiation, while the other three genes show opposing effects. We are currently performing knockdown of these target genes using siRNA. We will track changes during differentiation in the absence or presence of insulin and follow up with functional analyses for insulin sensitivity in the mature adipocytes to measure glucose uptake, triglyceride accumulation and intracellular glycerol levels. Our studies will provide insights into these novel insulin resistance loci and their associated genes, to further understand the underlying mechanisms that lead to the development of insulin resistance.

Session Title: Molecular Effects of Genetic Variation Poster Session II

PB1079 Discovery of target genes at lipid and metabolic disease loci in liver through a targeted CRISPRi screen.

Authors:

G. Pandey[1], S. Vadlamudi[2], A. Moxley[3], K. Currin[3], J. Nicholas[4], J. McAfee[1], K. Broadaway[2], K. Mohlke[5]; [1]Univ. of North Carolina, Chapel Hill, Chapel hill, NC, [2]Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, [3]UNC Chapel Hill, Chapel Hill, NC, [4]Univ. of North Carolina, Chapel Hill, Chapel Hill, NC, [5]Univ North Carolina, Chapel Hill, NC

Genome-wide association studies (GWAS) have identified hundreds of risk loci for liver disease and lipid-related metabolic traits, although most target genes and mechanisms remain unknown. In a previous study, we generated a catalog of metabolic GWAS variants that may alter liver regulatory elements (peaks) by colocalizing GWAS signals with chromatin accessibility quantitative trait loci (caQTL). We predicted target genes of these peaks using colocalization with liver tissue eQTL, proximity to transcription start sites, correlation of distal caQTL peaks with promoter peaks, and/or chromatin conformation. In the current study, we sought to validate target genes at these loci using functional assays. We focused on loci for which a proxy GWAS variant overlapped a liver caQTL peak linked to a gene using ≥2 methods. We did transcriptional reporter assays in HepG2 cells to test the regulatory effects of variants. We observed significant allelic differences in transcriptional activity for variants at 4 GWAS signals for liver enzymes or lipid traits: rs13395911 near *EFHD1*, rs11644920 near *LITAF*, rs34003091 near *ZNF329*, and rs9556404 near *GPR180*. For the two non-promoter regulatory elements (rs13395911 and rs11644920) linked to *EFHD1 and LITAF*, we used CRISPRi with an inducible dCas9-KRAB system in HepG2 cells to assess effects of enhancer knockdown on expression of the predicted target and nearby genes. We introduced ≥6 gRNAs for each element into cells via lentiviruses and measured gene expression 72 hours after transduction. Compared to cells infected with non-targeted control gRNAs, cells infected with gRNAs targeting enhancers surrounding rs13395911 and rs11644920 showed significantly (P<.05) lower expression levels of *EFHD1* (23.7% decrease) and *LITAF* (24.7% decrease). At both loci, directions of effect matched the caQTL and eQTL associations. Near *LITAF*, expression level was also reduced for adjacent genes in the same topological domain, *SNN* and *TXDC11*. In a second round of analyses, we identified 476 GWAS signals that colocalized with liver caQTL, but that doesnot have an eQTL-target gene. Among these, we selected 25 caQTL elements containing a GWAS proxy variant for which the peak is detectable and predicted genes are expressed in HepG2; CRISPRi knockdown of these elements to identify target genes is underway, and further studies will investigate effects of enhancer knockdown on cellular processes. In summary, we used a scalable framework to validate predicted regulatory effects of liver lipid GWAS variants and identified their target genes, which will aid identification of cell mechanisms and elucidate the causes of liver disease.

Session Title: Molecular Effects of Genetic Variation Poster Session III

PB1080 DisCO-VG: Disease-specific optimization method for prioritizing variant-target gene pairs.

Authors:

E. Chou, A. Guerra, S. Li, T. Qin, K. Wang, Z. Zhang, J. T. Elder, L. C. Tsoi, M. A. Sartor; Univ. of Michigan, Ann Arbor, MI

Prioritizing GWAS-identified variants and linking them to their target genes are critical steps in predicting regulatory noncoding variants in disease and prioritizing genes for drug targets. Methods such as the combined SNP-to-gene linking strategy (cS2G) and Single-Cell Enhancer Target gene mapping (SCENT) have shown excellent results pinpointing causal variants and disease genes. However, these methods have limitations. For one, SCENT prioritizes singular enhancer-gene pairs, while it is known that enhancers containing variants can regulate several genes. Furthermore, cS2G does not necessarily prioritize SNP-gene-disease triplets with cell type or tissue specificity. Here, we apply a statistical learning approach, DisCO-VG (Disease-specific CTMC Optimization for Variant-Gene pairs), for prioritizing cell type-specific, disease-associated noncoding variant-target gene pairs. DisCO-VG leverages CTMC (Coupled Tensor-Matrix Completion), for which excellent performance has been observed in drug-target interaction and chemical exposure-target gene predictions. By approaching the problem of linking variants to target genes as a matrix completion problem, where variants make up the rows and genes make up the columns, we apply biological knowledge of a variant affecting multiple genes by design. Utilizing significant GWAS associations, eQTLs, predicted enhancer-gene links, and distance between variant and gene, the variant-gene pairs will have entries in this matrix ranging from 0 to 1, with a higher score representing a higher probability of targeting. DisCO-VG also utilizes supporting information in the form of tensors describing gene-gene and variant-variant relationships, respectively. These coupled tensors can incorporate multiple sources of evidence. The input into these tensors include GTEx expression correlation across cell types and GO term similarities for gene-gene similarities; for variant-variant similarities, they include belonging to similar transcription factor binding motifs and having similar ATAC-Seq profiles across multiple experimental conditions. Our initial application is studying psoriasis, where we are using data derived from peripheral blood mononuclear cells of both healthy and psoriatic samples to benchmark DisCO-VG against existing methods. Ultimately, we output a list of prioritized cell type- and disease-specific variant-gene pairs that can be experimentally validated and become the basis for potential drug targets.

Session Title: Molecular Effects of Genetic Variation Poster Session I

PB1081 Dissect the genetic architecture of congenital vertebral malformation in the context of the developing spine

Authors:

N. Wu[1], S. Zhao[2], L. Zhao[1], G. Li[3], Z. Zhao[1], K. Xu[1], J. Cai[1], Q. Li[1], Z. Wu[4], J. Zhang[1]; [1]Peking Union Med. Coll. Hosp., Beijing, China, [2]Baylor Coll. of Med., Houston, TX, [3]Peking Union Med. Coll. Hosp., beijing, China, [4]Peking Union Med. Coll. Hosp., Chinese Academy of Med. Sci., Beijing, China

Congenital vertebral malformation (CVM), affecting 0.13-0.50 per 1000 live births, has an immense locus heterogeneity and complex genetic architecture. In this study, we analyzed exome/genome sequences from 873 probands with CVM and 3794 control individuals. Clinical interpretation identified Mendelian etiologies in 12.0% of the probands and revealed a muscle-related disease mechanism. Gene-based burden test of ultra-rare variants identified risk genes with large effect sizes (ITPR2, TBX6, TPO, H6PD, and SEC24B). Exome-wide association study (EWAS) uncovered the polygenic architecture. To dissect the complex genetic architecture of CVM, we estimated the proportion of heritability explained by burden and EWAS signals. The five burden test signals and the three independent EWAS signals accounted for 4.2% and 11.5% of the phenotypic variance in CVM patients without a molecular diagnosis, respectively (3.7% and 10.1% in all CVM patients). Taking together the Mendelian etiology and burden/EWAS heritability, we were able to explain around 25.8% of the overall heritability of CVM. To further investigate the biological relevance of the genetic association signals, we performed single-cell RNA-seq on human embryonic spines. The burden test and EWAS signals were enriched in the notochord at early developmental stages and myoblast/myocytes at late stages, suggesting their critical role in the developing spine. The combined analysis of genetic data and embryonic single-cell transcriptome data provides insights into the developmental biology and pathology of the human spine.

Session Title: Molecular Effects of Genetic Variation Poster Session II

PB1082 Dissecting the genetic basis of cell-type specific immune activation across indigenous Cameroonians.

Authors:

D. Kelly[1], N. Xie[1], Y. Feng[1], W. JIA[1], E. Basner-Tschakarjan[1], W. Beggs[1], C. N. Folkunang[2], A. K. Njamnshi[2], C. Ye[3], E. Luning Prak[1], S. A. Tishkoff[1]; [1]Univ. of Pennsylvania, Philadelphia, PA, [2]Faculty of Med. and BioMed. Sci., Univ. of Yaoundé I, Yaoundé, Cameroon, [3]UCSF, San Francisco, CA

The intensity and nature of the immune response to infection varies globally across individuals, affecting susceptibility to both communicable and non-communicable diseases. This variation is shaped by environmental factors, such as pathogen exposure and treatment accessibility, as well as genetic factors that have, in turn, been shaped by evolution. A valuable model for studying inter-individual variation in immune activation is the *in vitro* stimulation of immune cells; when *in vitro* stimulation is combined with state-of-the-art single-cell transcriptomic techniques, expression changes following immune activation can be studied in an individual, stimulation, and cell type-specific manner. As with the majority of genomic studies, however, most *in vitro* stimulation and transcriptomic studies have focused on individuals of Eurasian ancestry or westernized settings. To broaden our understanding of how human immune variation is shaped by environmental and genetic factors, we have performed *in vitro* stimulation of peripheral blood mononuclear cells (PBMCs) using LPS and IFN-β in a Cameroonian cohort comprised of 58 Tikari agriculturalists, 58 Fulani pastoralists, and 44 Central-African Rainforest Hunter Gatherers (CRHGs). After performing single-cell mRNA-sequencing of approximately 1.5 million stimulated and unstimulated PBMCs we find extensive population and cell type variation in gene expression across control and stimulation conditions, as well as variation in the intensity of the stimulation response. Combining these data with genetic data from the same individuals, we map quantitative trait loci (QTLs) regulating expression across cell-types and conditions. Finally, integration of QTL data with scans of natural selection allows us to understand how evolution has shaped genetic variation controlling the immune response in these populations.
This work has been funded by the Chan Zuckerberg Initiative, NIH grant R35 GM134957-01, and the American Diabetes Association Pathway to Stop Diabetes grant #1-19-VSN-02.

Session Title: Molecular Effects of Genetic Variation Poster Session III

PB1083 Dynamic effects of disease-associated variants on chromatin accessibility across human immune cells

Authors:

Z. Mu; Univ. of Chicago, Chicago, IL

A small fraction (~25%) of autoimmunity-associated variants from genome-wide association studies (GWAS) is explained by steady-state expression quantitative trait loci (eQTL) discovered in bulk RNA-seq studies (Mu et al., Geno. Biol., 2021). Epigenetic variation may offer novel insights in understanding the remaining loci. Here we created a large, unified single-cell chromatin accessibility (scATAC) map in peripheral blood of 56 rheumatoid arthritis (RA) and healthy donors, comprising ~220,000 cells.

We applied a novel topic modeling method, fastTopics, to single-cell count data and identified cell trajectories associated with RA. We observed an effector T cell trajectory leading to a rare CD8+ T cell population in RA patients that is masked in standard cluster analyses. Notably, this trajectory is enriched in RA heritability (OR=28.2, p=1.3E-8) and tracks the activation of known RA genes like IL2, CXCL13 and LILRB1. Next, we identified 25,107 cis-regulatory elements (cREs) with significant chromatin accessibility QTL (caQTL, 10% FDR) with RASQUAL in whole blood and seven immune cell types, quadrupling the hits compared with previous study. RASQUAL uses allelic-imbalance to increase power, but its output is incompatible with most downstream analysis. To overcome this, we used a single-cell Poisson mixed-effects model to estimate standard summary statistics for significant caQTLs. We found that caQTLs show higher levels of sharing across cell types than eQTLs do; consequently, shared caQTLs often colocalize with eQTLs in just a few specific contexts, raising the possibility that many caQTLs are nonfunctional in many contexts. Last, we colocalized caQTLs with 47 GWAS of autoimmune diseases and blood phenotypes. Remarkably, caQTL explained ~50% more GWAS loci than eQTLs. When GWAS loci colocalize with both caQTL and eQTL in the same contexts, we are more likely to nominate putative causal genes (OR=1.6, p=2.7E-7). However, 57% of colocalized GWAS loci only have caQTL colocalization, and we found that they may represent cases where the causal effect is mediated through nearby cREs but in an as yet uncharacterized context.

To conclude, we show that caQTLs greatly complement eQTL in explaining disease GWAS, but the large number of caQTL-GWAS colocalization could be due to the widespread sharing of caQTL across contexts. Thus, even in the presence of clear colocalization signals, pinpointing the causal context for these GWAS loci remains difficult. Our findings indicate that we must broadly study disease-relevant contexts and conduct experimental validation on the identified cREs in order to truly understand the mechanism of disease SNPs.

Session Title: Molecular Effects of Genetic Variation Poster Session I

PB1084 Effects of pathogenic variation in *ADAMTSL4* on vascularsmooth muscle cell function relevant to spontaneous coronary artery dissection.

Authors:

C. Pinnock, K. Hunker, S. Ganesh; Univ. of Michigan, Ann Arbor, MI

Spontaneous coronary artery dissection (SCAD) is the cause of 1.7-4% of acute coronary syndromes (ACS), with a strongly sex-biased prevalence, with more frequency occurrence in female individuals. In a whole exome sequencing study of SCAD, we recently reported exonic variants in *ADAMTSL4* that were predicted to be deleterious by in silico analysis. We and others have previously reported a robustly associated common variant, rs12740679 at chromosome 1q21.2, identified by genome-wide associations study of SCAD that demonstrated a colocalized expression quantitative trait locus effect on *ADAMTSL4*, with reduced gene expression associated with the SCAD risk allele. Expression of myc-tagged *ADAMTSL4* construct of the coding variant *ADAMTSL4 c.2263dupG (p.Gly758Trpfs*59)* in 293T cells demonstrated 50% reduction in the secretion of *ADAMTSL4* from the transfected cells (N = 3), as compared to the *ADAMTSL4 WT*-transfected control. Based upon this finding, which is concordant with the GWAS risk allele association with reduced *ADAMTSL4* expression, we modeled these effects by transient siRNA transfection of female human coronary artery smooth muscle cells (huCASMCs) and performed in vitro studies to understand the function of *ADAMTSL4* in the huCASMC extracellular matrix. We hypothesized that we would observe effects on fibrillogenesis and ECM maturation/stability owing to previously described interactions of *ADAMTSL4* with fibrillin 1 (*FBN1*) and lysyl oxidase (*LOX*), respectively. After achieving an 88±9.73% (N = 5) knock down of *ADAMTSL4* mRNA expression following transfection we observed a significant increase in cell migration (N = 5) and wound healing (N = 3). Cell adhesion and contraction was quantified using a traction force microscopy (TFM) assay, and a significant increase in traction force (N = 6) was observed following *ADAMTSL4* loss of function. Biomaterial properties were assessed via tensile testing engineered tissue rings (ETR) created with *ADAMTSL4* siRNA transfected female huCASMCs, showed a significant decrease in ultimate tensile strength, elastic modulus, and failure strength following candidate gene knock down (N = 3). A microfibril assay using *ADAMTSL4* siRNA transfected huCASMCs cultured in conditioned media with excess secreted *ADAMTSL4 WT* or *ADAMTSL4 c.2263dupG (p.Gly758Trpfs*59)* via immunofluorescence staining showed a decrease in *FBN1* binding and fibrillogenesis when cells were cultured in the *ADAMTSL4* variant-conditioned media. These findings support a role for *ADAMTSL4* in arterial smooth muscle and suggest that *ADAMTSL4* loss may cause abnormal cellular and extracellular effects which can lead to arterial destabilization in SCAD.

Session Title: Molecular Effects of Genetic Variation Poster Session II

PB1085 Elucidating ancestral differences in expression quantitative trait loci architecture relative to Alzheimer's Disease Status.

Authors:

M. Mews[1], N. Wheeler[1], P. Benchek[1], T. Gu[2], L. Gomez[2], C. Kosanovic[3], M. Cornejo-Olivas[4], B. Feliciano-Astacio[5], G. Beecham[2], G. Byrd[6], J. Haines[1], M. Pericak-Vance[3], A. Griswold[2], W. Bush[1]; [1]Case Western Reserve Univ., Cleveland, OH, [2]Univ. of Miami, Miami, FL, [3]Univ. of Miami Miller Sch. of Med., Miami, FL, [4]Inst. Natl. de Ciencias Neurologicas, Lima, Peru, [5]Univ. Central del Caribe, Bayamón, PR, [6]NC A&T State Univ, Greensboro, NC

Alzheimer's disease (AD) has increased prevalence in African American (AA) and admixed Hispanic (HISP) populations relative to Non-Hispanic Whites (NHW). While it is known that the underlying genetic risk for AD differs between genetic ancestries, we understand less about how these differences in the genetic architecture of AD may contribute to the observed prevalence differences. To explore how the genetic components underlying gene expression may differ among the populations and whether AD status contributes to changes in gene expression, we performed expression quantitative trait loci (eQTL) analysis from whole-blood RNAseq gene expression and genome-wide genotyping data from AA, HISP, and NHW individuals.

RNAseq data was processed using the Alzheimer's Disease Sequencing Project (ADSP) FunGen-xQTL computational protocol. Genotyping array data was imputed using the TOPMed imputation panel and filtered to include variants with an imputation accuracy prediction of r2 > 0.8 and those directly genotyped by the array. We then performed eQTL analysis stratified by population and AD status.

We identified 147,893 genome-wide significant eQTLs for NHW (N=236) individuals, 106,062 eQTLs for HISP (N=339), and 17,533 eQTLs for AA (N=224). By comparing these groups, we identified 4,204 shared significant variants amongst all populations with effect size directions 100% concordant. Within the groups, we identified AD-specific eQTLs with 507 found within NHW, 214 within HISP, and 26 within AA. From cognitively unimpaired (CU) individuals, we identified CU-specific eQTLs with 688 found within NHW, 143 within HISP, and 269 within AA. *No AD-specific or CU-specific eQTLs were shared amongst or between the three populations*. Among the significant AD-specific variants within each population, 27% of NHW variants (139/507), 37% of HISP variants (79/214), and 77% of AA variants (20/26) were not previously identified as significant eQTLs in eQTLGen, the European-descent blood eQTL analysis of 31,684 individuals (Võsa & Claringbould et al. 2021). Similarly, among the significant CU-specific variants within each population, 36% of NHW variants (247/688), 63% of HISP variants (90/143), and 45% of AA variants (121/269) were not previously identified as significant eQTLs in eQTLGen.

Our results suggest prominent differences in eQTL architecture between populations with AA samples exhibiting at least an 80% reduction in eQTLs relative to NHW and HISP. The presence of AD and CU-specific variants solely found within each population highlight the importance of personalized risk assessment informed by genetic ancestry.

Session Title: Molecular Effects of Genetic Variation Poster Session III

PB1086 Elucidating therapeutic pathways for disease through the integration of fine-mapping and genetic colocalization applied to 2,923 circulating proteins in 38,697 UK Biobank participants

Authors:

C. Robins[1], Q. Wang[2], N. Bowker[2], J. Davitte[1], A. Cortes[2], T. Richardson[2], R. Scott[2]; [1]GSK, Collegeville, PA, [2]GSK, Stevenage, United Kingdom

Trait-associated protein quantitative trait loci (pQTL) are a valuable source of evidence to support drug development. In this study, we developed an innovative approach that integrates fine-mapping and genetic colocalization to systematically evaluate the impact of 2,923 circulating proteins on 340 disease phenotypes and biomarkers. We performed genome-wide pQTL and fine-mapping analyses using genetic and plasma proteomic data from 38,697 UK Biobank (UKB) participants of European ancestry. These results were integrated with fine-mapped summary statistics from GWASs of curated UKB disease and biomarker phenotypes using genetic colocalization methods. We identified a total of 76,369 protein-trait pairs with colocalization evidence (posterior probability >80%), which suggests shared causal variants underlying both plasma protein levels and disease. Of the identified protein-trait pairs, 2.8% (2,135) have colocalization evidence at *cis* loci (within 1 Mb of the protein-coding gene), while 99% (75,662) have colocalization evidence at *trans* loci. The majority of the protein-trait pairs with *cis* evidence were identified only through conditional colocalization (60%, 1,279 pairs) and would not have been identified without conditional analysis to resolve distinct signals. This included 273 protein-trait pairs supported by multiple independent *cis*-pQTL colocalizations. For instance, multiple independent *cis*-pQTL for LPL provided evidence of colocalization with triglyceride levels, highlighting the lipoprotein lipase pathway as a therapeutic target for treating atherosclerosis. Additionally on this pathway, secondary *cis*-pQTL for ANGPTL3 colocalize with lipoprotein lipid traits. This is further consolidated by rare protein truncating variants in the coding region of *ANGPTL3* that were identified through exome sequencing in UKB. Of the protein-trait pairs with *cis*-pQTL colocalization evidence, 67% (1,428/2,135) were also supported by *trans* pQTL colocalization evidence. Many of these *trans* pQTL findings reveal known biology, including protein-protein interactions and shared disease mechanisms. For example, ANGPTL3 *trans*-pQTL located at the *APOA5* locus colocalize with lipoprotein lipid traits. APOA5 is another member of the lipoprotein lipase pathway, highlighting a coordinated system of effects which may have been overlooked without jointly considering evidence from *cis*- and *trans*-pQTL. Taken together, our approach can identify and harness independent secondary pQTL signals and allelic series to improve downstream causal inference and develop our understanding of the relationships between proteins and disease.

Session Title: Molecular Effects of Genetic Variation Poster Session I

PB1087 Engineering a synthetic humanized*RET*mouse to model multifactorial Hirschsprung disease

Authors:

R. Fine[1], A. Chakravarti[1], J. Laurent[1], J. Boeke[1], B. E. Low[2], A. Zuberi[2], M. V. Wiles[2]; [1]NYU Sch. of Med., New York, NY, [2]The Jackson Lab., Bar Harbor, ME

Hirschsprung disease (HSCR) is multifactorial but with ~50% of attributable risk from the receptor tyrosine kinase RET gene. To understand the genetic and cellular events that lead to aganglionosis in HSCR, we have built a humanized mouse model with an intact 77kb human RET genomic locus inserted into the Rosa26 safe-harbor locus using direct zygote injection and recombination with the Bxb1 serine-integrase. Our approach takes advantage of cutting-edge "Big-DNA" technology in budding yeast and E. coli to synthetically construct these loci. We have confirmed integration using nanopore Cas9-targeted sequencing (nCATS) as well as Illumina short-read sequencing. The intact locus includes all intron sequences and ~10kb upstream of the RET promoter, which enables direct modeling of known coding and regulatory human pathogenic variants in mice. We have functionally confirmed expression of the human RET gene using qRT-PCR. While this initial model will answer a number of outstanding questions, the majority (42%) of HSCR risk can be further defined from three HSCR-associated noncoding variants, rs2506030, rs7069590, and rs2435357, within RARB, GATA2 and SOX-10 binding RET enhancers. We have demonstrated topological chromatin interactions between the three enhancer variants and the RET promoter by 5C-ID sequencing in CHP212 cells. The 77kb strain harbors the susceptibility alleles at rs7069590 and rs2435357. To further increase our model's accuracy, we have additionally constructed a 179kb locus in yeast with a defined risk haplotype to include the remaining distal enhancer. We are further advancing an mSWAP-IN delivery approach in mouse embryonic stem cells to deliver the 179kb locus directly to the mouse Ret locus. We will present our approach, the progress towards the various RET humanized mouse strains harboring noncoding regulatory variants and their transgenic phenotypes at the genetic, gene expression, cell behavior and aganglionosis levels.

Session Title: Molecular Effects of Genetic Variation Poster Session II

PB1088 Enhancer RNA transcriptomics-wide association study reveals an atlas of pan-cancer susceptibility eRNAs

Authors:

W. Chen, L. Li; Shenzhen Bay Lab., Shenzhen, China

Increasing evidence has evidenced that enhancer RNAs (eRNAs) play essential roles in human disease. However, it is still a major challenge to comprehensively identify the subset eRNAs relevant to disease. Here we constructed the first atlas of common and rare genetic influences on enhancer RNA expression across 49 human tissues. We identified 11,757 eRNA quantitative trait loci (eRNA-QTLs) associated with the expression of 89.75% annotated eRNAs. eRNA-QTLs significantly contribute to cancer heritability, and 28.48% of cancer variants are co-localized by eRNA-QTLs. We performed an enhancer RNA transcriptome-wide association study (eTWAS) and identified 259 cancer susceptibility eRNAs for 23 cancer types. Their nearby target genes of these eRNAs were strongly enriched in oncogenes. More interestingly, rare functional variants associated with extreme eRNA expression can also substantially affect cancer risks. Additionally, we developed a comprehensive and flexible data portal for exploring the genetic basis of eRNAs. Overall, this study revealed the systematic landscape of genetic effects on eRNAs across human tissues and significantly expanded the diseases-associated eRNAs.

Session Title: Molecular Effects of Genetic Variation Poster Session III

PB1089 Enhancer-promoter compatibility and cellular context in disease-associated gene regulation

Authors:

A. Tovar, Y. Kyono, K. Nishino, M. Bose, A. Varshney, J. Kitzman, S. Parker; Univ. of Michigan, Ann Arbor, MI

Complex diseases arise from the combined effects of genetic, environmental, and lifestyle factors. Recent genome-wide association studies have revealed that the majority of complex disease-associated loci are found in noncoding regions. However, the framework to translate these noncoding variants to specific transcriptional processes is ill-defined, unlike the codon-amino acid system for coding variants. Massively parallel reporter assays (MPRAs) have emerged as a useful method to screen large collections of putative regulatory elements, efficiently prioritize candidate causal variants at disease loci, and learn principles of disease-associated regulatory variation. Previously, we used a modular MPRA to discover sequence features associated with cell- and sequence context-specific regulatory activity. Using a panel of type 2 diabetes (T2D)-associated fragments paired with a synthetic housekeeping promoter (SCP1) or the biologically relevant human insulin promoter (*INS*), we discovered a subset of fragments displaying higher enhancer activity with the *INS* promoter. Interestingly, these fragments were enriched for pancreas- and neuroendocrine-relevant transcription factor motifs, including HNF1 family members. Exonic mutations in *HNF1A*/*HNF1B* cause maturity onset diabetes of the young (MODY) while T2D-associated variants are found in their cognate binding motifs, suggesting genetic convergence between rare and common genetic variation. We used a follow-up MPRA library to investigate two main questions: (1) do HNF1 binding motifs contribute to *INS* promoter-biased enhancer activity? and (2) is this phenomenon present in other promoter and cellular contexts? We indeed observed several instances where deletion or mutation of HNF1 motifs disrupted the *INS* promoter-biased enhancer activity. By contrast, when we delivered the same library to the LHCN-M2 human skeletal muscle line, we observed that many of these fragments displayed low enhancer activity that was not disrupted by HNF1 motif deletion or mutation, including when cloned with the promoter for the muscle-specific gene *MYBPC2*. Additionally, we observed instances where variants displayed allelic activity only in the presence of the *MYBPC2* promoter in LHCN-M2 cells and not in other contexts. Together, our study suggests that cell-specific regulatory activity is partially influenced by enhancer-promoter compatibility, and indicates that careful attention should be paid when designing MPRA libraries to capture context-specific regulatory processes at disease-associated genetic signals.

Session Title: Molecular Effects of Genetic Variation Poster Session I

PB1090 Enhancing the activity of YKL-39: investigating mutations and enzymatic properties.

Authors:

T. Kanaizumi, K. Suzuki, M. Ohkura, K. Okawa, **F. Oyama**; Kogakuin Univ, Tokyo, Japan

YKL-39, also known as CHI3L2, is a secretory protein that lacks chitinase activity. Despite its enzymatic inactivity, YKL-39 is expressed in healthy individuals and found to be overexpressed in various diseases such as osteoarthritis (OA), rheumatoid arthritis (RA), glioblastoma, multiple sclerosis (MS), amyotrophic lateral sclerosis (ALS), Alzheimer's disease (AD), and certain cancers. This highlights its potential as a biomarker and therapeutic target. The absence of chitinase activity in YKL-39 is attributed to mutations in the conserved catalytic sequence. Previous studies have demonstrated that introducing two specific amino acid mutations can activate YKL-39 (Schimpl et al., Biochem. J., 446, 149-157, 2012). In our study, we aimed to further enhance the activity of YKL-39 by introducing three substitutions in the catalytic motif of the wild-type protein, resulting in MT-YKL-39. The comparative evaluation revealed increased activity of MT-YKL-39 compared to WT-YKL-39, with an optimal pH of 5.0, similar to chitotriosidase (CHIT1), an active human chitinase. Furthermore, we generated a chimera by fusing MT-YKL-39 with CHIT1 and conducted a chitinase activity assay, which indicated increased chitinase activity levels in some chimeras. Future investigations will focus on detailed studies of the enzymatic properties and pH dependence of chitinase activity in the chimeras. Our ongoing research aims to activate and further understand the potential of MT-YKL-39.

Session Title: Molecular Effects of Genetic Variation Poster Session II

PB1091 Enrichment of incidental/secondary genetic findings of G6PD, TTR, TTN, RYR1, and BRCA2 in a cohort of 8667 African-American pediatric subjects.

Authors:

E. Bhandari[1], A. Saeidian[1], M. March[1], L. Youssefian[2], D. Watson[1], M. Harr[1], G. Kazemioula[3], F. Aliazami[4], F. Mentch[1], X. Wang[1], A. Saeidian[5], H. Vahidnezhad[6], M. Akbarzadeh[7], L. Tsoi[8], J. Gudjonsson[8], J. Glessner[1], H. Hakonarson[1]; [1]Children's Hosp. of Philadelphia, Philadelphia, PA, [2]UCLA, Los Angeles, CA, [3]Univ. of Social Welfare and Rehabilitation Sci., Tehran, Iran, Islamic Republic of, [4]Azad Univ., Tehran Med. Sci., Dept. of Genetics, Tehran, Iran, Islamic Republic of, [5]Iran Univ. of Med. Sci., Tehran, Iran, Islamic Republic of, [6]Univ. of Pennsylvania, Philadelphia, PA, [7]Shahid Beheshti Univ. of Med. Sci., Tehran, Iran, Islamic Republic of, [8]Univ. of Michigan, Ann Arbor, MI

Purpose: We utilized clinical next-generation sequencing to identify pathogenic sequence variations not associated with a patient's primary diagnosis in a large pediatric cohort. These variations are referred to as incidental findings (IFs), secondary findings (SFs), or unsolicited findings (UFs). Our cohort was specifically enriched for Black children, which comprises 52% (8667) of the total cohort (16713) at the Center for Applied Genomics at The Children's Hospital of Philadelphia. The inclusion rate for Black individuals (adults and pediatrics) in published genome-wide studies is less than 3%; this highlights a dire need for genetics research on this population. The strong representation of Black subjects in our cohort allows us to yield information on the types and frequency of pathogenic IFs primarily affecting Black children and families. **Methods:** We searched for pathogenic (P) and likely pathogenic (LP) variants in disease-causing genes, specifically in 78 ACMG recommended and 93 non-ACMG genes that met the definition of IFs in patients of four large pediatric cohorts. The list of 93 non-ACMG genes was compiled from a literature review for genes proposed for addition to the ACMG list. Variants were filtered by P/LP classification as per ACMG/AMP guidelines, CADD vs Mutation Significance Cutoff (MSC) scores, variant type, presence in disease-causing variant databases, and IGV visualization for manual variant quality inspection. We divided the patient's P/LP variants into four disease categories: metabolic, cancer, cardiovascular, and other. We further divided our patients by self-reported ethnicity. **Results:** We report a distinctive distribution of types and frequency of IFs in Black individuals. Black subjects were 335/350 (95%) of total subjects with P/LP variants in the G6PD gene, 307/320 (96%) in TTR, 62/146 (42%) in TTN, 29/48 (60%) in RYR1, and 24/46 (52%) in BRCA2. Out of 335 Black subjects with P/LP variants in G6PD, the most common variant (327) was p.V98M. Homozygous females comprised 39/327 (12%) of this specific P/LP variant. Out of 307 Black subjects with P/LP variants in TTR, the most common variant (304) was p.V142I. The other 3 variants were p.V50M, p.V113M, and p.D119V. **Conclusions:** Our data on IFs in a large pediatric cohort enriched for Black subjects provides valuable data on P/LP variants which mainly affect Black children. The implications of identifying such in an underrepresented patient population are vast, as many previous studies on IFs has been dominated by White cohorts. This will inform medical management, reproductive planning, and additional family member genetic testing for Black children and families specifically.

Session Title: Molecular Effects of Genetic Variation Poster Session III

PB1092 eQTL analysis of canine testes identifies gene expression patterns associated with canine body size.

Authors:

R. M. Buckley[1], A. C. Harris[1], J. Hale[1], D. T. Whitaker[1], S. E. Lana[2,3], E. A. Ostrander[1]; [1]Natl. Human Genome Res. Inst., Bethesda, MD, [2]Coll. of Vet. Med. and BioMed. Sci., Colorado State Univ., Fort Collins, CO, [3]Flint Animal Cancer Ctr., Colorado State Univ., Fort Collins, CO

The role of breed-associated genetic variation in shaping morphological diversity of dogs is difficult to assess, as most of this variation is non-coding. To overcome this, we initiated the canine eQTL project. Bulk RNA-seq and low-pass WGS data was collected and analyzed for 143 canine testes samples. To address the effects of cell-type heterogeneity in bulk RNA-seq data we performed cell-type deconvolution with human single cell testes expression profiles to estimate cell-type composition of dog samples. We tested the efficacy of this approach by focusing on the well-characterized IGF1 small body size allele and its association with serum IGF1 levels. We found testes IGF1 expression was only associated with the IGF1 small body size allele when cell-type covariates were included in our models, highlighting the importance of correcting for cell-type heterogeneity in detecting eQTLs. Moreover, inclusion of cell-type covariates led to the identification of an additional 1,138 eGenes. To identify genes potentially regulating morphological traits in breeds, we performed colocalization analysis for 28 breed height and breed weight loci. Two loci, one on chromosome 7 and the other on chromosome 34, which affect both breed height and breed weight, were supportive of eQTL colocalization (posterior probability > 0.5). The chromosome 7 locus was associated with SMAD2 expression and the chromosome 34 locus was associated with IGF2BP2 expression. The expression of both genes was positively correlated with breed height and breed weight, indicating how expression changes in SMAD2 and IGF2BP2 regulate canine body size. Finally, we investigated eQTL activity within a chromosome X locus that is > 1Mb in size and is known to contribute to proportionately large breed size. IRS4 and ACSL4 were previously identified as candidates in this region as they contained missense variants and similar functional roles in other species. However, IRS4 and ACSL4 showed no significant expression change in the presence of the large breed size allele. Alternatively, AMMECR1 and CHRDL1, genes associated with short stature and bone growth in other systems, showed decreased expression with the large breed associated allele, providing as novel candidates for regulating canine body size. The canine eQTL project has identified a total of 1,838 eGenes and has helped to identify novel gene candidates for common traits in dogs. As this resource matures, it will continue to facilitate the detection of target genes from GWAS and provide a framework for comparing genetic architecture across species.

Session Title: Molecular Effects of Genetic Variation Poster Session I

PB1093 eQTLGen phase 2: Genome-wide *trans*-eQTL analysis in blood in over 35,000 individuals provides insight into the genetic architecture of molecular traits.

Authors:

R. Warmerdam[1,2], A. van der Graaf[3], eQTLGen Consortium, T. Esko[4], Z. Kutalik[3], H-J. Westra[1,2], L. Franke[1,2], U. Võsa[5]; [1]Univ. Med. Ctr. Groningen, Univ. of Groningen, Dept. of Genetics, Groningen, Netherlands, [2]Oncode Inst., Utrecht, Netherlands, [3]Univ. of Lausanne, Lausanne, Switzerland, [4]Estonian Genome Ctr., Inst. of Genomics, Univ. of Tartu, Tartu, Estonia, Tartu, Estonia, [5]Estonian Genome Ctr., Inst. of Genomics, Univ. of Tartu, Tartu, Estonia

Gene expression quantitative trait loci (eQTL) analysis has become an essential tool for understanding complex disease mechanisms. However, previous eQTL studies have been limited by sample size and have therefore mainly focused on *cis*-regulatory effects. We recently performed a *trans*-eQTL analysis in 31,684 samples (eQTLGen), but were only able to study ~10,000 genetic variants, due to computational challenges. Here, we present interim results from eQTLGen phase 2, the largest-scale genome-wide *trans*-eQTL analysis to date.

We have developed robust pipelines to perform automated, comprehensive data quality control and genotype imputation in individual cohorts. Methodology adapted from the HASE framework allows for genome-wide meta-analyses for thousands of phenotypes and samples, while limiting data transfer sizes and ensuring participant privacy. We have successfully applied the method to 12,384 available samples and are currently expanding the analysis to more than 35,000 samples.

Currently, we have been able to identify 25,796 independent *trans*-eQTLs, and we expect this to increase as we expand to 35,000 samples, providing substantially more power. This will provide us with a more complete understanding of gene expression regulation, and reveal gene-to-gene interactions through comprehensive fine-mapping, colocalization and Mendelian randomization analyses. Moreover, our findings will provide a resource for the genetics research community and offer a foundation for future functional genomics studies.

In conclusion, we highlight the importance of large-scale *trans*-eQTL analysis in understanding gene expression regulation and its implications for complex disease biology.

Session Title: Molecular Effects of Genetic Variation Poster Session II

PB1094 Examining the impact of natural genetic variation on meiosis and fertility in hybrid mice using single-cell chromatin accessibility and gene expression

Authors:

I. Agarwal, B. Myers, M. Houlard, E. Bitoun, S. Qu, A. Hinch, S. Myers; Univ. of Oxford, Oxford, United Kingdom

Meiosis involves progression through several short-lived but key transcriptional events (e.g., PRDM9 expression), and critical instances of chromatin remodeling, including PRDM9 binding to recombination hotspots, synapsis between homologous chromosomes, and silencing of unsynapsed chromatin. Many of these processes are difficult to study directly as they are stochastic properties of individual cells in very transient cell populations.
Here, we take advantage of genetic diversity in a sample of fertile B6, infertile (PWDxB6) F1 and variably-fertile PWDx(B6xCAST) F1 hybrid mice, together with simultaneous gene expression profiling by RNA-seq and mapping of open chromatin by ATAC-seq in single cells, to characterize the genetic regulation of the entire meiotic process. These mice possess ~10-fold greater genetic diversity than humans, greatly facilitating identification of allelic differences. We identify key binding events that control meiotic gene expression during specific meiotic stages, including transient allele-specific opening of chromatin and histone repositioning at Prdm9 binding sites, lagging RNA expression of Prdm9. We show how broad-scale gene-expression and chromatin patterns change in a coordinated fashion throughout meiosis: two examples are X-chromosome inactivation, and compaction of chromatin in spermatids. We identify cell stages throughout the meiotic program, including cells that have undergone the first and second meiotic division; moreover, we can identify sites of crossovers and instances of aneuploidy in single cells.
Studying hybrids with variable levels of fertility in particular allows us to examine how mutations impact regulatory binding and drive differential expression in the germline, and to investigate the molecular basis of infertility in these mice. Together, these approaches provide a powerful and generalizable toolbox to study the mechanisms of germline development, and to dissect the effects of common genetic variation on a number of molecular traits in the male germline that are directly relevant to fertility.

Session Title: Molecular Effects of Genetic Variation Poster Session III

PB1095 Exploring dual and triplets genetic variants in nucleotide excision repair-related genes: unraveling their impact on disease severity and structural consequences in patients with late-onset Xeroderma pigmentosum phenotype.

Authors:

A. BOUAZIZ[1,2,3,4,5], F. FRIKHA[6], N. KHERIJI[7], L. ROMDHANE[8,9], M. ZGHAL[10], O. MESSAOUD[1]; [1]BioMed. Genomics and Oncogenetics Lab., Inst. Pasteur of Tunis, Tunis, Tunisia, [2]Département Génétique, Faculté de Médecine de Sousse, Université de Sousse, Sousse, Tunisia, [3]Faculté de Pharmacie de Monastir, Université de Monastir, Monastir, Tunisia, [4]Natl. Sch. of Computer Sci. , Université de Manouba, Manouba, Tunisia, [5]Higher Inst. of Biotechnology of Sidi Thabet, Université of Manouba, Sidi Thabet, Tunisia, [6]Faculty of Sci. of Sfax (FSS), Univ. of Sfax, Sfax, Tunisia., Sfax, Tunisia, Sfax, Tunisia, [7]BioMed. Genomics and Oncogenetics Lab., Inst. Pasteur of Tunis, Tunis, Tunisia, Tunis, Tunisia, [8]Département des Sci. de la Vie, Faculté des Sci. de Bizerte, Université Carthage, 7021 Jarzouna, BP 21, Tunisie., Bizerte, Tunisia, [9]Lab. of BioMed. Genomics & Oncogenetics, Inst. Pasteur de Tunis, Université Tunis El Manar, 13, Place Pasteur BP 74, 1002 Tunis, Belvédère, Tunisie., Tunis, Tunisia, [10]Dept. of Dermatology, La Rabta Hosp., Tunis, Tunisia., Tunis, Tunisia

Xeroderma pigmentosum (XP) is a genetic condition primarily characterized by compromised DNA repair capabilities, resulting in a heightened vulnerability to ultraviolet (UV) radiation and a propensity towards skin cancer. Our previous work on the largest cohort worldwide enabled us to identify causative founder and private genetic variants. For a small fraction of patients, with a mild phenotype, the genetic etiology has not yet been identified. The goal of our research was to unravel the genetic basis of late-onset XP forms.

We implemented exome sequencing in three Tunisian XP patients. The Agilent SureSelect XT Human All Exon v6 panel was used for exome capture, followed by high-throughput sequencing via the NovaSeq6000 platform. The raw sequence data were meticulously processed and aligned against the human reference genome (GRCh38). Variants' pathogenicity was evaluated using several bioinformatics in-silico prediction tools as well as protein structure through advanced homology modelling and molecular dynamics. For proteins with no reference structure, domain modelling using threading and ab initio methods was applied.

Our analysis unveiled a previously unexplored genetic landscape associated with mild XP. Indeed, we have identified not only pathogenic variants in *DDB2*, *UVSSA*, *ERCC5*, and *ERCC6* genes, but, more notably, at least two homozygous variants were identified for every patient. For the first patient, we observed a variant in the *UVSSA* gene (p.620L) along with a second variant in ERCC5 (p.G1053R). The second patient exhibited two variants in the *ERCC5* gene (p.G1053R and p.G1080R), and a variant in the *ERCC6* gene (p.R382K). Lastly, the third patient presented three variants, including a variant in *DDB2* (p.A128P), a variant in *UVSSA* (p.P620L), and a variant in *ERCC5* (p.G1053R). The pathogenicity of all these variants was confirmed by at least 5 in-silico prediction tools. Furthermore, their impact on the structural stability and function of the protein was demonstrated. Indeed, structural modelling and 100ns molecular dynamics simulations revealed a destabilizing effect of the Ala128Pro substitution in the *DDB2* protein. Similarly, we observed a significant disparity in stability between the wild-type and mutated proteins.

In conclusion, this is the first study reporting double homozygous mutants in XP patients. Moreover, our integrated methodologies, including network analysis and structural characterization, contributed to a deeper understanding of the functional repercussions of the identified variants and their implications on the variable severity degree between the investigated patients.

Session Title: Molecular Effects of Genetic Variation Poster Session I

PB1096 Exploring long non-coding RNAs at skin cancer risk loci

Authors:

M. H. Law[1,2,3], I. Rivera[1], M. Seviiri[1,2], J. Beesley[1], L. T. Koufariotis[1], N. Waddell[1,4], S. MacGregor[1], M. T. Landi[5], M. M. Iles[6,7], K. M. Brown[5], J. French[1], S. Edwards[1]; [1]QIMR Berghofer Med. Res. Inst., Brisbane, Australia, [2]Sch. of BioMed. Sci., Faculty of Hlth., Queensland Univ. of Technology, Brisbane, Australia, [3]Sch. of BioMed. Sci., The Univ. of Queensland, Brisbane, Australia, [4]Faculty of Med., The Univ. of Queensland, Brisbane, Australia, [5]Natl. Cancer Inst., Bethesda, MD, [6]Leeds Inst. for Data Analytics, Univ. of Leeds, Leeds, United Kingdom, [7]NIHR Leeds BioMed. Res. Ctr., Leeds Teaching Hosp. NHS Trust, Leeds, United Kingdom

Cutaneous melanoma, basal cell carcinoma and squamous cell carcinoma are the three most frequent types of skin cancers and share common risk factors. Genome-wide association studies have identified over 100 independent or shared risk loci for these cancers. While analyses incorporating methylation and protein-coding expression data have identified putative gene targets for many cutaneous melanoma loci, basal and squamous cell carcinoma lag behind, and across the three cancers, many loci remain uncharacterised.

As with other complex traits many skin cancer risk SNPs lie in transcriptional enhancers or other regulatory elements. In addition to protein-coding genes, these SNPs may regulate the expression of long non-coding RNAs (lncRNAs). Previous research has indicated that some lncRNAs are involved in important biological pathways including those relevant to skin cancers.

To map novel skin lncRNAs we extracted RNA from melanocytes, fibroblasts and keratinocytes and performed de novo assembly. Following assembly by Trinity, the data was validated, filtered and optimised using BUSCO and TransRate. Transcripts were processed with ezLncPred using three different methods to predict coding potential (CPAT, CPC2, CNCI) parallel to using FEELnc to identify lncRNAs.

Following the exclusion of protein-coding or annotated genes we identified 16,076 novel, credible, lncRNAs. The expression of 7,589 of these lncRNAs was measurable in 106 primary melanocyte cultures and quantitative trait loci (lncQTLs) were identified for 7,400 autosomal transcripts using fastQTL. PEER factors were used to control for technical or confounding effects in the analyses.

646 of the lncRNAs had a QTL SNP significant after accounting for multiple testing. Testing for colocalisation with known skin cancer risk loci identified 12 transcripts for melanoma, seven for basal cell carcinoma and two for squamous cell carcinoma. Risk loci exhibiting lncRNA colocalisation included one on chromosome 15 containing the *OCA2* gene. There the lead SNP rs12913832, known for being the main determinant of blue vs brown eye colour, its association with skin cancer risk, and transcriptional regulation of *OCA2*, was also associated with the expression of two novel lncRNAs. We will next expand our lncRNA discovery set to include additional melanoma tumour sets, increasing our power to detect lncRNAs relevant to the development of skin cancers.

Session Title: Molecular Effects of Genetic Variation Poster Session II

PB1097 Exploring the evolutionary arms race between human PKR and poxvirus K3

Authors:

M. Chambers[1,2], T. Dever[3], M. J. Sadhu[1]; [1]Natl. Human Genome Res. Inst., Bethesda, MD, [2]Georgetown Univ., Washington, DC, [3]NICHD, Bethesda, MD

The interfaces between host and viral proteins can be dynamic spaces in which genetic variants are continually pursued, giving rise to evolutionary arms races. One such scenario is found between the mammalian innate immunity protein PKR (protein kinase R) and the poxvirus antagonist K3. Once activated, PKR phosphorylates the natural substrate eIF2α, which halts translation within the cell and prevents viral replication. K3 acts as a pseudosubstrate antagonist against PKR by directly antagonizing this halt in translation, enabling poxviruses to replicate in the cell. Exploring the impact of genetic variants in both PKR and K3 is necessary not only to highlight residues of evolutionary constraint and opportunity but also to elucidate the mechanism by which human PKR is able to subvert a rapidly evolving viral antagonist. To systematically explore this dynamic interface, we have generated a combinatorial library of PKR and K3 missense variants to be co-expressed and characterized in a high-throughput yeast selection assay. This assay allows us to characterize hundreds of thousands of unique PKR-K3L genetic combinations in a single pooled experiment. Our results highlight specific missense variants available to PKR that subvert the K3 antagonist. We find that improved PKR variants are readily available at sites under positive selection, with limited opportunity at sites interfacing with K3 and eIF2α. Additionally, we find few functional variants that improve K3 antagonism, suggesting the antagonist is optimized. We reason that this approach can be leveraged to explore the evolutionary plasticity of many other host-virus interfaces.

Session Title: Molecular Effects of Genetic Variation Poster Session III

PB1098 Exploring the regulatory potential of the human genome.

Authors:

J. Srivastava, I. Ovcharenko; NIH, Bethesda, MD

The robust regulatory landscape of the human genome results from a multi-dimensional control of gene expression involving concerted action of chromatin accessibility and organisation, transcription factor binding, alternative splicing, control by regulatory RNAs, etc. Mutations affecting these processes result in a phenotypic outcome that is dependent on the robustness of the system. Enhancers are major drivers of gene regulation and mutations in these regions can result in insignificant to lethal phenotypes. This begets fundamental questions about robustness of enhancer-based regulation. Does the level of deleteriousness correspond to enhancer categories, i.e., do mutations leading to less severe phenotype occur in developmental and pioneer factor binding enhancers (mutation resistant enhancers)? Or do they reflect shadow enhancers that have evolved to offset deleterious effects from loss-of-function mutations? Second, what are the biological processes that benefit from such robustness, and those vulnerable to mutations due to lack thereof? We use a forward genetics approach to address these questions. We generate synthetic genomes by introducing mutations at random 2% sites (quantified as the percent of single nucleotide variations between the human and chimpanzee genome) in the whole genome sequence by retaining nucleotide substitution rates between the two organisms. We then use deep learning models in various cellular contexts to identify enhancers from wild type and mutant whole genome sequences. Our preliminary results show that stem cell derived, embryonic and blood cell lines gain enhancers while differentiated cell lines representing muscle and epithelial cells lose enhancers in the event of random mutations. In addition to addressing the questions above, we are currently evaluating the number of mutations that would result in complete loss of enhancer activity among wild type enhancers and map these regions to biological pathways to assess their disease vulnerability. By comparing our data with documented genome variants, we can assess the critical sites of regulatory regions and delineate their significance in biological function.

Session Title: Molecular Effects of Genetic Variation Poster Session I

PB1099 Expression CROP-sequencing identifies causal variants in *cis*-eQTL for genes associated with inflammatory bowel disease.

Authors:

E. Greenwood[1], M. Cao[2], C. Lee[3], B. Moyo[2], G. Bao[2], G. Gibson[1]; [1]Georgia Tech, Atlanta, GA, [2]Rice Univ., Houston, TX, [3]Univ. Coll. Cork, Cork, Ireland

This study fine-maps immune expression quantitative trait loci (eQTL) to the nucleotide level in a myeloid cell line, revealing considerably greater regulatory complexity than is generally appreciated. Although statistical fine-mapping of eQTL has been widely used, extensive linkage disequilibrium limits resolution to a credible set of probable variants. Credible sets often range from tens to hundreds of variants, and it is unclear whether one or multiple of these are responsible for the eQTL signal. Here, using expression CROP-sequencing (eCROP-seq), we combine the power of CRISPR-Cas9 and single cell RNA-sequencing (scRNA-seq) to identify causal variants in *cis*-eQTL for genes implicated in inflammatory bowel disease (IBD). Each variant in an eQTL is assigned a unique gRNA that will target that region for a double stranded break. Following scRNA-seq, cells that received this perturbation are detected via the readable gRNA transcript, and gene expression of the IBD target gene in these cells are compared to all other cells in the pool that did not receive this gRNA. We select 61 IBD associated genes that are adequately expressed in the immortalized myeloid cell line, HL-60, and target ~4,400 variants in *cis*-eQTL for these genes. In 79% of genes, we identify one or more causal variants that significantly alter target gene expression at a Bonferroni-adjusted *p*-value, corrected for the number of comparisons at each locus. Our findings suggest that multiple causal variants explain the full eQTL effect for the majority of target genes and that different causal variants explain each peak if a locus has more than one eQTL signal. Additionally, we describe four classes of eQTL structure and show that the distribution of causal variants within the peak differs among the classes. Future work will include repeating eCROP-seq analysis for all IBD associated genes expressed in the immortalized T lymphocyte cell line, Jurkat. Preliminary results suggest causal variant cell-type specificity when comparing significant variants in Jurkat cells to those in HL-60 cells for the same eQTL. Our results imply that GWAS signals should often be interpreted as haplotype effects rather than single variant signals and have implications for understanding regulatory architecture and evolution.

Session Title: Molecular Effects of Genetic Variation Poster Session II

PB1100 Extensive co-regulation of neighboring genes limits the utility of eQTLs in target gene prioritization.

Authors:

R. Tambets[1], A. Kolde[1], P. Kolberg[1], M. I. Love[2], K. Alasoo[1]; [1]Univ. of Tartu, Tartu, Estonia, [2]UNC-Chapel Hill, Chapel Hill, NC

Identifying causal genes underlying genome-wide association studies (GWAS) is a fundamental problem in human genetics which is mainly rooted in the pleiotropic nature of gene expression - a single gene can be regulated by several enhancers and a single enhancer can regulate several genes. The accompanying lack of ground truth makes it difficult to benchmark models for determining causality on real-world data. To generate such a ground truth dataset, we focused on the large collection of fine mapped gene expression and protein quantitative trait locus (QTL) data from the eQTL Catalogue and the INTERVAL study. Assuming that the most likely causal gene underlying each cis-pQTL association is the gene coding for the protein allowed us to identify 872 high-confidence gene-protein pairs.

We compared three Bayesian colocalization methods (coloc.susie, coloc.abf and CLPP) and two Mendelian randomization (MR) approaches (inverse-variance weighted MR and MRLocus). We found that assigning fine-mapped pQTL variants to their closest protein coding genes outperformed all colocalization methods both in terms of precision (67% vs 72%) and recall (55% vs 77%). Furthermore, the colocalization method with the highest recall (coloc.susie - 55%) also had the lowest precision (38%). Importantly, restricting the analysis to gene-protein pairs with at least two independent colocalizing signals increased precision to 71%, but at the cost of significantly lower recall (7%).

We next used MR to further assess if multiple colocalizing variants had consistent effect size direction and magnitude on gene and protein levels. Unexpectedly, the standard inverse variance weighted MR had a very high false positive rate. In multiple cases, variants with opposite effects on gene and protein levels still yield highly significant non-zero causal effect estimates. Many of these false positives were removed by the Bayesian MRLocus approach, further increasing the precision to 75%. The remaining false positives seemed to be primarily driven by strong co-regulation between neighboring genes.

Our results highlight that linking disease-associated variants to their causal target genes remains challenging using eQTL evidence alone. Prioritizing target genes at truly novel loci likely requires triangulation of evidence from multiple sources, including rare protein altering variants and splicing QTLs that are likely to be less pleiotropic. Furthermore, MR methods need to be carefully adapted for use with eQTL signals, where high uncertainty in instrument effect size estimates must be accounted for.

Session Title: Molecular Effects of Genetic Variation Poster Session III

PB1101 Factors Influencing the Portability of Gene Expression Imputation across genetic ancestries in TWAS

Authors:

Y-H. Chang[1], A. Bhattacharya[1], Y. Ding[2], B. Pasaniuc[2]; [1]Univ. of California, Los Angeles, Los Angeles, CA, [2]UCLA, Los Angeles, CA

Transcriptome-wide association studies (TWAS) investigate genetic associations potentially mediated by gene expression, increasing power compared to genome-wide association. TWAS relies on imputation of gene expression based on models trained in reference expression quantitative trait locus (eQTL) datasets. Previous studies have identified that models trained in majority European-ancestry cohorts are unable to predict gene expression in external datasets of individuals of different ancestry. Here, through extensive simulations based on data from the 1000 Genomes Project (1000G), we examine potential factors that drive this lack of portability of expression models.

Across 1,000 simulations at 22 gene loci (one per chromosome), we simulate a training eQTL dataset of European samples and test sets of both European and African populations based on the linkage disequilibrium (LD) patterns in 1000G and across a variety of genetic architectures. We vary the heritability and number of causal eQTLs. We build a sum of single effects (SuSiE) linear regression model in the training dataset and predict expression in the test datasets. To assess the predictive performance of these models, we calculated the percent variance explained (PVE) in the test sets.

Our findings indicate that, in general, as the proportion of causal eQTLs increases from 1% to 10% and the heritability is 0.05, PVE decreases by 62% and 66% in the European and African test sets, respectively. We also examine differences in PVE across three factors in genetic architecture. First, we examined the impact of varying allele frequencies of the causal eQTLs and found no significant differences, except in cases where the causal proportion was extremely high or low. Second, we explored the effect of changing the ratio of heritability between African and European populations and observed a decrease of 0.028 in the PVE difference as the ratio changed from 0.1 to 2. Third, we are investigating how differences in LD across the datasets are associated with differences in performance.

In conclusion, our findings emphasize the importance of explicitly considering differences in heritability, causal eQTL proportion, and LD when aiming to improve the portability of gene expression imputation across ancestry.

Session Title: Molecular Effects of Genetic Variation Poster Session I

PB1102 Female neonate with bilateral microphthalmia, vermis hypoplasia, and foot deformities, new case report, case series, and phenotypic delineation ofAlkuraya-Kucinskas syndrome.

Authors:

H. Alakeel, F. AlMutairi; Natl. Guard For Hlth.Affairs, Riyadh, Saudi Arabia

We hereby report a 1 day old female neonate, with bilateral microphthalmia, ankyloblepharon, absent vermis, clenched fest, anterior placement of anus, sacral dimple with hair tuft, ASD II, and talipes equinovarus, her WES has confirmed Alkuraya-Kucinskas syndrome; Homozygous likely pathogenic mutation in KIAA1109 gene; c.4672-1G>T. ALKKUCS, OMIM # 617822, is an autosomal recessive severe neurodevelopmental disorder characterized by arthrogryposis, brain abnormalities associated with cerebral parenchymal underdevelopment, and global developmental delay. Most affected individuals die in utero or soon after birth. Additional abnormalities may include hypotonia, dysmorphic facial features, and involvement of other organ systems, such as cardiac or renal. The few patients who survive have variable intellectual disability and may have seizures. To our current literature review we have found reported 15 patients, we are hereby analyzing their phenotypic presentation, expanding it’s spectrum, and comparing their molecular finding in order to have a better understanding of Alkuraya-Kucinskas syndrome, it’s genotypic-phenotypic correlation, and estimate it’s prognosis based on current diagnosed patients.

Session Title: Molecular Effects of Genetic Variation Poster Session II

PB1103 Fewer than 2% of all possible *SPINK1* coding single-nucleotide variants affect splicing

Authors:

J. Chen[1], J-H. Lin[2,3], X-Y. Tang[4,2], W-B. Zou[2,3], S. Schutz[1], E. Masson[1,5], D. N. Cooper[6], C. Ferec[1], Z. Liao[2,3], H. Wu[2,3]; [1]Univ Brest, Inserm, EFS, UMR 1078, GGB, Brest, France, [2]Dept. of Gastroenterology, Changhai Hosp., Naval Military Med. Univ., Shanghai, China, [3]Shanghai Inst. of Pancreatic Diseases, Shanghai, China, [4]Dept. of Prevention and Hlth.Care, Eastern Hepatobiliary Surgery Hosp., Naval Military Med. Univ., Shanghai, China, [5]Service de Génétique Médicale et de Biologie de la Reproduction, CHRU Brest, Brest, France, [6]Inst. of Med. Genetics, Sch. of Med., Cardiff Univ., Cardiff, United Kingdom

INTRODUCTION: Single-nucleotide variants (SNVs) occurring within the coding sequence of any gene have the potential to affect splicing. In this regard, it has been reported that up to 10% of known disease-associated missense variants alter pre-mRNA splicing. Moreover, the number of synonymous SNVs altering splicing is also steadily increasing. However, to date, a precise estimation of the prevalence and extent of the impact on splicing of all possible SNVs within the coding sequence of any gene is lacking. Herein, we sought to provide such an estimation for the four-exon *SPINK1* gene, taking advantage of the availability of a highly accurate full-length gene splicing assay (FLGSA) on the one hand and the availability of the highly reliable artificial intelligence-based splicing prediction tool, SpliceAI, on the other.
METHODS: All 720 possible SNVs within the 240 nucleotides of the entire *SPINK1* coding sequence were subjected to SpliceAI prediction. The accuracy of SpliceAI prediction was first evaluated in the context of 27 known *SPINK1* coding SNVs (24 missense and 3 synonymous) that were previously analyzed by means of FLGSA. Based upon SpliceAI-predicted delta scores, 35 possible SNVs in 16 coding nucleotide positions (with delta scores ranging from 0 to 0.93) were selected for FLGSA characterization.
RESULTS: In total, 62 *SPINK1* coding SNVs were functionally analyzed, accounting for 8.6% of all possible coding SNVs. They affected 42 (17.5%) of the 240 coding nucleotides. Notably, all three possible SNVs in 9 nucleotide positions, were functionally characterized. Of the 62 FLGSA-analyzed SNVs, 12 were found to affect splicing. Of these 12 splice-altering variants, 9 generated both wild-type and aberrant transcripts whereas the remaining 3 generated only aberrant transcripts. The splice-altering SNVs only occurred in exons 1 and 2, and predominantly affected the first and/or last coding nucleotide of the respective exon. Importantly, SpliceAI prediction and experimentally obtained findings were found to concur in all cases of the 62 analyzed SNVs. This excellent correlation, taken together with the fact that we have included all variants with a significant delta score for functional characterization, suggest that none of the not functionally analyzed SNVs should have any appreciable impact on splicing. Therefore, only 1.67% (n = 12) of the 720 possible coding *SPINK1* SNVs would be considered to affect splicing.
CONCLUSION: Combining SpliceAI-predictions with FLGSA-derived data from a large number of known and prospectively generated SNVs, we confidently conclude that fewer than 2% of all possible *SPINK1* coding SNVs affect splicing.

Session Title: Molecular Effects of Genetic Variation Poster Session III

PB1104 Fine-mapping of functional traits using fSuSiE reveals local genetic influences of epigenetic and transcriptomic regulation in human brains.

Authors:

H. Sun[1], W. Denault[2], P. Carbonetto[2], The FunGen-xQTL Consortium, M. Stephens[2], G. Wang[1]; [1]Columbia Univ., New York, NY, [2]Univ. of Chicago, Chicago, IL

Many genetic variants identified in genome-wide association studies are located in epigenetic regulatory elements such as histone modification and DNA methylation (DNAm). These epigenetic processes have been implicated in complex disease etiology through their effects on gene activities. Therefore, we would like to understand the underlying regulatory genetic mechanisms by interrogating available genomic, epigenomic, and transcriptomic data sets. Here we describe the results of our efforts to fine-map epigenomic QTLs—that is, identify putative causal epigenomic quantitative trait loci (QTLs)—from H3K9ac (haQTL) and DNAm (meQTL) data across 877 topologically associated domains (TADs) in the human prefrontal cortex from the ROSMAP project, and integrate them with 15,339 fine-mapped expression QTLs (eQTLs) identified in the same brain subjects. The fine-mapping of this heterogeneous data is based on a newly developed method, fSuSiE ("Functional Sum of Single Effects"), which can account for both spatial correlations of hundreds of epigenomics marks and linkage disequilibrium between genetic variants. fSuSiE identified numerous credible sets (CSs) that impact 29,853 transcription starting sites through complex regulatory effects, hence allowing us to gain insight into the genetic regulation of such epigenetic marks. In particular, fSuSiE identified: 10,642 fine-mapped CSs for DNAm, 1070 of them intersecting eQTL CSs, and 2,562 CSs for haQTL, 605 of them intersecting eQTL CSs. We identified 310 CSs affecting DNAm and H3K9AC (shared CSs). Overall, we found that the CSs affect the transcription starting site (TSS) in 23,238 genes through DNAm and over 12,804 through H3K9ac. Interestingly, many of the CSs affect multiple TSSs, and we found that 53% of mQTL CSs and 40% of haQTL CSs are located more than 500 kb from their epigenomic targets on the most distal TSS. Of the CSs shared between the two epigenomic marks, 77 associate with the same TSS through both marks, whereas 50 affect TSS solely through DNAm and 44 only through H3K9ac. Fine-mapped effect sizes across epigenetic marks in the shared CSs show a strong correlation (-0.6). We identify several examples where fine-mapped epigenomic QTL are likely to impact gene expression through complicated multi-omics regulations, such as those observed in genes *PSCA* and *TMEM131L*, where multiple independent QTL are involved. Our work offers promising future research directions on *cis*-regulatory networks underlying complex disease etiology.

Session Title: Molecular Effects of Genetic Variation Poster Session I

PB1105 Five novel mutations in ribosomal proteins in 11 Diamond Blackfan Anemia (DBA) patients: An Indian outlook

Authors:

P. Kedar, R. Dongerdiye; Natl. Inst. of Immunohematology (ICMR), Mumbai, India

Diamond Blackfan Anemia (DBA) is an inherited bone marrow (BM) failure syndrome characterized by a paucity of erythroid differentiation, congenital anomalies, and a predisposition for malignancies. DBA is mainly caused by the mutations in ribosomal protein(RP) genes, hence classified as ribosomopathy. More than 20 RP genes are known to be associated with DBA. The most frequent mutations are observed in RPS19, RPL5. However non-RP genes are also known to be associated with DBA. Approximately 50% of the DBA patients have physical anomalies like craniofacial defects, renal anomalies, limb anomalies, and cardiac defects. Patients usually have normochromic and macrocytic, severe anemia with reticulocytopenia. In the study, we report 11 cases of DBA identified by next generation sequencing (NGS). We have identified five novel mutations in RPL5, RPL27, RPL35A and RPS17 gene. In 10 of 11 patients, bone marrow examination suggested pure red cell aplasia which is the most prominent feature of DBA. Short stature is observed in two patients bearing mutation in RPL5 and RPL26 genes. NGS have revealed seven missense mutations in different RP gene- RPL27 (p.Ile25Thr), RPL35A (p.Ile101Ser), RPL5 (p. Asp234Gly, p.Tyr253Cys), RPS19 (p.Met1X), RPS24 (p.Val192Met), RPS26 (p.Met115Thr); two large base deletions in RPL35A, RPS17; one frameshift mutation in RPL5 (p. Asp59Tyrfs*53) and one splice site mutation in RPS26 (c.3+1G>T). In addition, increased levels of erythrocyte adenosine deaminase (e-ADA) were detected in all cases. The functional consequences of the variants were assessed and its effect on the protein structure were investigated using PyMOL. Genotype-phenotype correlation were established, although clinical heterogeneity makes it difficult to analyse. In India, DBA patients are managed by steroids, blood transfusion support, or bone marrow transplantation. Conclusively, the genome sequencing has resulted in significant advancement in understanding rare diseases like DBA.

Session Title: Molecular Effects of Genetic Variation Poster Session II

PB1106 From benign to pathogenic and back: the effect of context on genetic variants.

Authors:

T. Ciesielski[1], G. Sirugo[2], S. Iyengar[3], S. Williams[4]; [1]Case Western Reserve Univ. Sch. of Med., Cleveland, OH, [2]Univ. of Pennsylvania, Philadelphia, PA, [3]Case Western Reserve Univ., Cleveland, OH, [4]Case Western Reserve Univ, Cleveland, OH

Here we address a fundamental problem that hampers the clinical utility of genetic research: the characterization of variant effects on a single continuum from benign to pathogenic. While useful in some ways, this approach oversimplifies what we know about genetic variation and it ignores the fact that context often determines variant impact. Overall, we argue that we need to better probe the determinants of penetrance and implications of pleiotropy. For example, genetic and environmental factors often alter variant effects and thus we should expect lower variant penetrance in large studies with diverse ancestry groups and environments. When over 5000 ClinVar pathogenic and loss-of-function variants were assessed in two large biobanks, UK Biobank and BioMe, the mean penetrance was only 7%. This indicates that the participants in most of the family-based, clinical, and case-control studies that first identified these variants were more homogenous and enriched for etiologic cofactors. This effectively produced a selection bias that overestimated the average effect in the average context, but this language of bias distracts us the primary point: the average effect is not a very useful measure. We want to know what the variant means in a specific population or patient. When pleiotropy is also considered, it becomes apparent that context can change not only the likelihood of one outcome but it may completely shift the outcome of primary concern. For example, the variant that causes hemoglobin S can increase the risk of death from sickling, lower the risk of death from malaria, or increase the risk of kidney disease. The phenotypic spectrum is strongly modified by infections, the second allele, and co-inherited variants at other loci (e.g., alpha or beta thalassemia). Overall, annotation on a single contextual continuum from benign to pathogenic attempts to shoehorn a complex phenomenon into an overly simplistic framework. There is no panacea, but we propose two recommendations. First, we need to routinely evaluate contexts such as sex and genetic ancestry by conducting stratified analyses and considering female-to-male allele proportion ratios. Here we demonstrate the value of sex ratios by revealing how putatively dangerous high ratio variants can evade detection in sex-agnostic assessments. Second, we need to consistently document what we know about effect modifiers in our annotation databases. These are not the only possible approaches, but they represent the first steps on a path that can create useful annotations for specific patients and clinical scenarios. We cannot achieve accuracy until we acknowledge that there is more than one target.

Session Title: Molecular Effects of Genetic Variation Poster Session III

PB1107 From Uncertainty to Clarity: Decoding the Molecular Consequences of Two *KPTN* Gene Variants for the Diagnosis of a Rare Neurodevelopmental Disorder

Authors:

C. Washington[1], D. B. Everman[2], K. Howard[3], K. M. Butler[3], R. J. Louie[4]; [1]Greenwood Genetic Ctr., Greenville, SC, [2]Evermed Genetics and Genomics Consulting, Greenville, SC, [3]Greenwood Genetic Ctr., Greenwood, SC, [4]Greenwood Genetic Ctr., Portland, OR

This work adds to the existing evidence for *KPTN*-related disorder (MIM: 615637) by expanding its clinical and mutational spectrum. The research and evaluation of our now 12-year-old female patient offer insights into two novel biparentally-inherited sequence variants in the *KPTN* gene, NM_007059.2: c.863G>C (p.Arg288Pro) and NM_007059.2: c.863G>A (p.Arg288Gln) both located at the exon-intron boundary of exon 9. Through targeted cDNA analysis of the proband and her parents, our team observed two aberrant splicing effects: intron retention and exon skipping for the *KPTN* RNA, leading us to reclassify these uncertain variants as likely pathogenic for our patient. The re-classification of these variants provided a diagnosis for the family and informed genetic counseling for future reproductive risk. Clinically, our patient has typical features of *KPTN*-related disorder, including macrocephaly, intellectual disability, and behavior problems. Yet, our patient does not have seizures and, to our knowledge, is the second reported individual with co-morbid precocious puberty, suggesting phenotype variability. Two key points are raised in this case presentation. First, it supports whole exome sequencing as a first-tier test for individuals presenting with neurodevelopmental delay, as this approach would have decreased the time to diagnosis by seven years for our patient. Second, additional molecular analysis of suspicious sequence variants by a laboratory with expertise in RNA diagnostics is suggested. Without this additional evaluation, the diagnosis of *KPTN*-related disorder would have remained speculative. Overall, this study provides additional clinical and molecular descriptions of an individual with a *KPTN*-related disorder, supporting the need for larger cohort studies and genotype-phenotype correlations in the future.

Session Title: Molecular Effects of Genetic Variation Poster Session I

PB1108 Functional analysis of *LDLR* variants using flow cytometry.

Authors:

K. Konecna[1,2], L. Tichy[1], T. Loja[3], T. Prerovska[2], L. Fajkusova[1]; [1]Ctr. of Molecular Biology and Genetics, Univ. Hosp. Brno, Brno, Czech Republic, [2]Natl. Ctr. for Biomolecular Res., Faculty of Sci., Masaryk Univ., Brno, Czech Republic, [3]Central European Inst. of Technology, Masaryk Univ., Brno, Czech Republic

Familial hypercholesterolemia (FH) is one of the most common autosomal dominant disorders. It is characterized by elevated low-density lipoprotein (LDL) cholesterol and contributes to higher risk of cardiovascular diseases. The most common cause of FH are genetic variants in the low-density lipoprotein receptor gene (*LDLR*). As of May 2023, there are over 3000 known variants in the *LDLR* gene, however, the clinical significance of many variants is uncertain. The purpose of this study is to functionally characterize 16 variants in the *LDLR* gene found in patients with phenotypic FH, to help determine their pathogenicity.

We performed *in vitro* functional assays based on transient expression of wild-type and mutant LDL receptors in LDLR-deficient Chinese hamster ovary (CHO) cells. To analyze the relative amount of LDL receptors expressed on the cell surface (compared to a control benign variant Ala391Thr), we labeled live cells with fluorescently tagged anti-LDLR antibody and analyzed fluorescence intensity by flow cytometry. We also analyzed the expression and maturation of LDLR protein in transfected CHO cells by Western blot.

The functional study indicated that out of 16 studied variants, 3 exhibited normal cell-surface expression of LDLR, while 13 variants exhibited lowered cell-surface expression of LDLR, potentially giving rise to FH. To further determine the effect of the studied variants on the whole LDLR cycle, we will also analyze the ability of mutant LDL receptors to internalize labeled LDL particles by flow cytometry. This information will be useful for identifying disease-causing variants in patients with phenotypic FH, allowing for cascade testing of potentially affected relatives.

Session Title: Molecular Effects of Genetic Variation Poster Session II

PB1109 Functional analysis of*MFSD3*associated with dementia with Lewy bodies.

Authors:

T. Kimura[1], M. Suganuma[1], K. Sawamura[1], Y. Asanomi[2], T. Hosoyama[1], N. Shimoda[1], N. Ogiso[1], S. Niida[1], K. Ozaki[3], D. Shigemizu[4]; [1]Natl. Ctr. for Geriatrics and Gerontology, Obu City, Japan, [2]Natl. Ctr. for Geriatrics and Gerontology, Obu City, Aichi Prefecture, Japan, [3]The Natl. Ctr. for Geriatrics and Gerontology, Obu, Aichi, Japan, [4]Med. Genome Ctr., Natl. Ctr. for Geriatrics and Gerontology, Obu, Aichi, Japan

With an increasing life expectancy, the incidence and prevalence of dementia increases, and a better understanding of the etiology and pathogenesis of dementia becomes ever more urgent. Dementia with Lewy bodies (DLB) is the second most common form of degenerative dementia in elderly people. The clinical feature of DLB includes cognitive impairment, visual hallucinations, parkinsonism, fluctuating attention, decreased brain acetylcholine levels, and apathy. Three genes, *SNCA* (α-synuclein), *APOE* (apolipoprotein E), and *GBA1*(glucosylceramidase beta 1), have been convincingly demonstrated to be associated with DLB. Our previous studies reported that a variant in *MFSD3* (rs143475431, c.888T>A:p.C296*) is associated with DLB in Japanese population by whole-genome sequencing and association studies. However, the functional understanding of the variant underlying DLB remains to be elucidated. Here, we performed function analyses of the *MFSD3* variant by using zebrafish, mice, and human neural stem cells. We generated the homologous *MFSD3*knockout zebrafish and mice using the CRISPR/Cas9 system. We found that the knockout zebrafish showed a significant increase of butyrylcholinesterase in the brain tissues and the knockout mice had significantly reduced levels of curiosity and activity in the novel environment compared with wild-type mice. These results are reminiscent of symptoms such as decreased brain acetylcholine levels and apathy in DLB patients. In addition, a decrease of the cell growth rate was observed in the *MFSD3* variant homozygous human neural stem cells. Totally, the *MFSD3* variant can play an important role in the pathogenesis of DLB. Further functional verification will contribute to the elucidation of the DLB pathogenesis mechanism.

Session Title: Molecular Effects of Genetic Variation Poster Session III

PB1110 Functional characterizationof eQTLs and asthma risk loci with scATAC-seq

Authors:

j. wei[1], J. Resztak[1], A. Alazizi[1], H. Mair-Meijers[1], R. Slatcher[2], S. Zilioli[3], X. Wen[4], F. Luca[1,5], R. Pique-Regi[1]; [1]Ctr. for Molecular Med. and Genetics, Wayne State Univ., Detroit, MI, [2]Dept. of Psychology, Univ. of Georgia, Athens, GA, [3]Dept. of Psychology, Wayne State Univ., Detroit, MI, [4]Dept. of Biostatistics, Univ. of Michigan, Ann Arbor, MI, [5]Dept. of Biology, Univ. of Rome "Tor Vergata", Rome, Italy

Cis-regulatory elements (CREs) control gene transcription dynamics in response to the environment. In asthma, multiple immune cell types play an important role in the inflammatory process. Genetic variants in the CREs can also affect gene expression response dynamics and contribute to asthma risk. However, the regulatory mechanisms underlying genetic control of transcriptional dynamics and their contribution to disease remain to be elucidated. Single cell ATAC-seq enables characterization of the genome-wide CREs at the single cell level and provides a unique opportunity to study gene expression regulatory mechanisms across cell-types and environmental contexts, and to annotate regulatory variants. We activated peripheral blood mononuclear cells (PBMC) from 96 children with asthma with phytohemagglutinin (PHA) or lipopolysaccharide (LPS), and treated with anti-inflammatory glucocorticoids (dexamethasone, DEX). We previously studied the transcriptional response with scRNA-seq; we now performed scATAC-seq in 16 individuals for 98,818 cells. Across four major cell-type clusters (B-cell, Monocyte, NK-cell and T-cell) and treatments, we identified 62,424 differentially accessible regions (DARs) (FDR<10%, $|\log_2 FC|>0.5$). Both LPS and PHA preferentially activated monocytes . We analyzed chromatin accessibility genome-wide to measure motif activity and identified 354 response factors with significant motif activity changes. We observed strong positive linear dependence between motif response and their target gene expression changes, but negative in variability changes. This result suggests that an increase of TF binding tightens the variability of gene expression around the mean. We then annotated genetic variants in peaks and in response motifs and performed computational fine-mapping of eQTL signals from a pediatric asthma cohort. We found that eQTLs were highly enriched in peaks with response motifs. Integrating the fine-mapped eQTLs with asthma GWAS, we refined a credible set of 70 risk genes (FDR<10%) supported by both TWAS and colocalization analysis, 33 genes of which had eQTLs in response motifs. In conclusion, scATAC-seq enhances the understanding of molecular mechanisms of individual variants contributing to asthma risk mediated by gene expression.

Session Title: Molecular Effects of Genetic Variation Poster Session I

PB1111 † Functional dissection of metabolic trait-associated variants in steady state and stimulated human skeletal muscle cells

Authors:

K. Nishino, A. Tovar, J. Kitzman, S. Parker; Univ. of Michigan, Ann Arbor, MI

Type 2 diabetes (T2D) is a common metabolic disorder characterized by dysregulation of glucose metabolism. Genome-wide association studies have defined hundreds of signals associated with T2D and related metabolic traits, most of which are found in noncoding regions. Given their central role in insulin production and glucose homeostasis, much focus has been devoted to investigating T2D-associated genetic variation in pancreatic islets; however, metabolic disease pathogenesis and risk is distributed across other important metabolic tissues, including the liver, adipose, and skeletal muscle. Here, we used a massively parallel reporter assay (MPRA) to characterize the regulatory activity of T2D-associated variants in human skeletal muscle cells at basal and multiple stimulatory states. We constructed a library of 1,255 oligos spanning 333 metabolic trait-associated variants, half of which were previously characterized using luciferase reporters or in MPRA libraries in any metabolic tissue, and therefore serve as positive controls. We delivered this library to LHCN-M2 human skeletal muscle myocytes in one of four conditions: (1) undifferentiated or differentiated with (2) basal media, (3) media supplemented with the AMP analog 5-aminoimidazole-4-carboxamide ribonucleotide (AICAR) to mimic physiological effects of exercise or (4) media containing sodium palmitate which is known to induce insulin resistance. Across all oligos, we identified 293 (23.3%) with significant enhancer activity in undifferentiated cells, 128 (10.2%) in basal differentiated cells, 199 (15.9%) in AICAR-stimulated cells, and 276 (22%) in palmitate-stimulated cells (FDR < 0.05). Across all conditions, we identified a shared set of 95 (28.5%) variants with significant allelic activity (FDR < 0.05). Further, we identified several instances of condition-specific allelic bias: 25, 5, 6, and 10 variants in undifferentiated, basal differentiated, AICAR-stimulated, and palmitate-stimulated, respectively. Several of these variants have been previously characterized in skeletal muscle cells or other metabolic tissues; however, none have been studied in these stimulated states. Together, this work identified a set of variants with regulatory activity in skeletal muscle cells and defined conditions in which they are active. We plan to extend these analyses by incorporating existing functional genomic data (e.g., ATAC-seq, ChIP-seq, chromatin state annotations) to explore regulatory features of these variants and better understand their role in regulating disease-relevant gene expression.

Session Title: Molecular Effects of Genetic Variation Poster Session II

PB1112 Functional evaluation of melanosomal pH regulation and melanin production for *OCA2* variants found in patients with oculocutaneous albinism type 2

Authors:

M. Gillis[1], L. Lundh[1], S. Wincovitch[1], H. Dorward[1], F. X. Donovan[1], NISC Comparative Sequencing Program, V. Gotea[1], L. Elnitski[1], W. S. Oetting[2], W. J. Pavan[1], W. A. Gahl[1], S. K. Loftus[1], D. R. Adams[1]; [1]NIH, Bethesda, MD, [2]Univ. of Minnesota, Minneapolis, MN

Oculocutaneous albinism (OCA) is an autosomal recessive disorder characterized by a reduction of pigment in the hair, skin, and eyes, foveal hypoplasia, optic nerve misrouting, and increased risk of skin cancer. OCA is phenotypically and genetically heterogeneous, with variants identified in *TYR*, *OCA2*, *TYRP1*, *SLC24A5*, *SLC45A2*, *LRMDA*, and *DCT*. *OCA2*, the gene responsible for OCA type 2, encodes a chloride channel on melanosomal membranes critical for maintaining a favorable pH for melanin production. 352 OCA probands were screened using a high-depth, short-read, custom capture sequencing approach, targeting 37 OCA- and pigmentation-related genes. A total of 85 *OCA2* rare variants were identified in 146 probands, including 62 single-nucleotide and small indels, 10 predicted splice, and 13 structural variants. Applying the American College of Medical Genetics framework of Richards et al. (2015) and incorporating *OCA2* specific criteria, we assessed the pathogenicity of *OCA2* variants. This led to 54 (64%) *OCA2* variants being defined as pathogenic or likely pathogenic and 31 (36%) as being of uncertain significance. To assess the pathogenicity of the remaining variants of uncertain significance, we have developed a cell-based functional assay for *OCA2* alleles. Through transfection of human *OCA2* constructs into melan-p5 (murine pink-eyed dilution (*Oca2*$^{p\text{-}J}$/*Oca2*$^{p\text{-}J}$) immortalized mouse melanocytes, we can assess the degree of pigmentation and rescue of melanosomal pH. Variants that demonstrate near wild-type pigmentation and neutralization of pH in melanosomes would be classified as benign, whereas those with minimal pigmentation and highly acidic melanosomes would be pathogenic. This analysis provides a more complete understanding of *OCA2* gene function and thus allows for phenotype-genotype analysis to be performed.

Session Title: Molecular Effects of Genetic Variation Poster Session III

PB1113 Functional genomic analysis of rs73352950: A rare, major depressive disorder (MDD) risk variant in Mexican Americans.

Authors:

S. Kumar[1], J. Curran[2], J. Granados[1], M. Aceves[1], A. Leandro[2], J. Peralta[2], S. Williams-Blangero[2,1], D. Glahn[3,4], J. Blangero[2]; [1]UTRGV, Sch. of Med., McAllen, TX, [2]UTRGV, Sch. of Med., Brownsville, TX, [3]Boston Children s Hosp. & Harvard, Boston, MA, [4]IOL & Yale, Hartford, CT

Major depressive disorder (MDD) is a common and potentially life-threatening disease. Recent genome-wide association studies (GWAS) focusing mainly on common genetic variation and overwhelmingly in individuals of European ancestry have identified several MDD risk loci but there have been no unequivocal (or experimentally validated) causal gene discoveries. Furthermore, while the heritability estimates for MDD and recurrent MDD among Hispanics are similar to those reported in individuals of European ancestry, GWAS in Hispanics reported minimal evidence of overlap with loci previously identified in non-Hispanic whites. GWA analysis in our Genetics of Brain Structure Study (GOBS) of Mexican American families identified a rare non-synonymous variant rs73352950 in the γ-adducin (*ADD3*) gene, which shows a highly significant association (Odds Ratio = 3.88, $p = 3.1x10^{-9}$) with MDD. Follow-up experiments in induced pluripotent stem cell (iPSC) derived neural epithelial/stem cells (NSCs) from 16 individuals who were heterozygous carriers for the identified variant and 13 controls suggest that the variant affects the adherens junctions (AJs) and the overall apical-basal polarity of the NSCs. In monolayer culture, NSCs derived from the variant carriers showed diminished adherens junction (AJ) reassembly indicated by fewer rosette-like structures and defused co-localization of the AJ proteins N-cadherin and γ-adducin in the rosettes. In a 3D neurosphere assay, NSCs of the variant carriers showed significantly higher cell migration ($p = 0.0031$) as a result of higher differentiation and significantly diminished self-renewal capacity. During development and in adult neurogenesis, apical/basal polarity and NOTCH signaling regulates NSC self-renewal and differentiation. Genome-wide gene expression analysis of the same carrier and control NSC samples identified 104 significantly ($p \leq 0.05$ and FC-abs ≥ 1.5) differentially expressed genes, which were significantly enriched in neuron differentiation ($p = 2.9x10^{-7}$), heparan sulfate proteoglycan (HSPG) metabolic process ($p = 3.7x10^{-5}$), visual perception ($p = 1.8x10^{-4}$), and extracellular structure organization ($p = 2.6x10^{-4}$) gene ontology (GO) processes. Interestingly, the NOTCH signaling inhibitor genes *TSPAN15* and *DLK1* were significantly downregulated. HSPGs regulate NSC self-renewal, differentiation, and overall central nervous system homeostasis and several studies have indicated a role for proteoglycans in MDD. Overall, our results show that rs73352950 in *ADD3* affects NSC self-renewal and differentiation capacity and is plausibly associated with MDD risk.

Session Title: Molecular Effects of Genetic Variation Poster Session I

PB1114 † Functional genomics studies in Xenopus can be used to inform clinical interventions in rare genetic diseases

Authors:

D. Baralle[1], A. Godwin[2], S. Ennis[3], S. Martin[4], J. Lord[1], C. Jaramillo-Oquendo[1], T. Fletcher[4], E. Seaby[3], M. Guille[4]; [1]Univ. of Southampton, Faculty of Med., Southampton, United Kingdom, [2]Univ. of Portsmouth, European Xenopus Resource Ctr., Portsmouth, United Kingdom, [3]Univ. of Southampton, Southampton, United Kingdom, [4]Univ. of Portsmouth, Portsmouth, United Kingdom

High throughput, next generation DNA sequencing has revolutionised understanding of human genetic diseases by enabling clinical research to identify causal variants within a gene. For many patients, the link between the gene identified by sequence comparisons and their disease is insufficiently robust for clinical decision making. With the backing of the European *Xenopus* Resource Centre (https://xenopusresource.org/about-us) we have established a UK government funded multidisciplinary team of clinicians, geneticists, bioinformaticians and developmental biologists focused on discovering the genetic and molecular basis of rare diseases in the South of England. Here we report some recent findings arising from this collaboration using the model organism *Xenopus tropicalis* to test genotype-phenotype correlations in syndromic models of dysmorphology, intellectual disability and developmental syndromes. We show how F_0 CRISPR screens in *Xenopus tropicalis* can test the genotype-phenotype correlations of candidate variants of unknown significance and prioritise those suitable for precision modelling. We also compare findings between the F_0 and F_1 CRISPR generations in autosomal dominant disease presentation, demonstrating the creation of stable lines for elucidating the developmental origin of disease, studying cellular or molecular mechanisms in greater detail and screening novel or re-purposed therapeutics. This work includes developing a pipeline to screen phenotypes in these disease models including high-resolution imaging techniques and novel behavioural parameters of working memory, anxiety, and motility in tadpoles, that are comparable to higher vertebrates. To date we have strong evidence for more than 20 novel disease-causing genes, including recently published work supporting the description of novel intellectual disability syndromes (*copb1* and *gria1*). These results not only strengthen disease modelling in *Xenopus* tadpoles but also advance knowledge of human gene function, supporting the goal to both reveal further novel genotype-phenotype associations and strengthen evidence for genotype-phenotype associations within known disease genes.

Session Title: Molecular Effects of Genetic Variation Poster Session II

PB1115 Functional validation of non-coding GWAS variants in inflammatory bowel disease (IBD) using high-throughput Lenti-MPRA.

Authors:

S. Ghosh[1], C. Espinoza[1], C. Lu[1], F. Rahimov[1], A. Alsheikh[2], A. den Hollander[1]; [1]AbbVie, North Chicago, IL, [2]AbbVie, Cambridge, MA

The functionalization of genetic variants and pinpointing causal genes in disease-associated loci identified by genome wide association studies (GWAS) remain a significant challenge. To identify causal mechanisms behind non-coding variant associations in inflammatory bowel disease (IBD), we implemented a Lentivirus-based massively parallel reporter assay (Lenti-MPRA) to screen for SNPs that disrupt enhancer activity in IBD- associated GWAS loci in colonic epithelial cells. Lentivirus-based MPRA has the advantage of generating 'in-genome' readouts, is a scalable and cost-effective approach and can be executed in difficult-to-transfect cells such as immune cells. In this study, we screened SNPs in the primary human immortalized colon cell line T0570 and colonic epithelial adenocarcinoma cell line Caco2, in presence and absence of IBD stimulation conditions (IFNϒ and TNF-α). We profiled H3K27Ac marks to map colon epithelial enhancers in T0570 and Caco2, and overlayed the enhancer regions with SNPs in LD (> r2 0.6) with top SNPs in IBD-associated GWAS loci. This approach led to a selection of 226 IBD colon epithelial specific GWAS SNPs, in addition we screened 45 SNPs from the *TNRC18* locus identified as the strongest IBD-associated locus in the Finnish population (FinnGen), along with 56 positive controls and 79 negative controls. Among the 226 IBD SNPs we identified 22 SNPs in Caco2 and 29 SNPs in T0570 with strong allele-specific enhancer activity (FDR < 0.001). Several SNPs with the strongest effects are in introns of genes previously implicated in IBD, such as *ERRFI1* (rs34768550) and *CARD9* (rs 4078099). We also narrowed down the functional (causal) SNPs in the FinnGen *TNRC18* locus to rs79032618 and rs182361057, which overlap with enhancers between the *WIPI2* and S*LC29A4* genes. In conclusion, this study identified functional variants in IBD GWAS loci, and reinforces Lenti-MPRA platform as a solid framework for prioritization of non-coding GWAS variants.

Session Title: Molecular Effects of Genetic Variation Poster Session III

PB1116 Functional validation of variants in cardiac troponin T using stem cell-derived cardiac muscle bundles.

Authors:

A. Renberg, O. Merritt, B. Elder, S. Friedline, J. Meisner, E. Smith, Y-C. Tsan, A. Helms; Univ. of Michigan, Ann Arbor, MI

Hypertrophic cardiomyopathy (HCM) is a heart muscle disorder most often caused by variants in sarcomere genes, including cardiac troponin T (*TNNT2).* However, variant interpretation is a major hurdle in ascribing pathogenicity of rare variants in *TNNT2* and other sarcomere genes. Functional testing with induced pluripotent stem cell-derived cardiomyocytes (iPSC-CMs) may improve variant interpretation but has been limited by throughput and reproducibility. We hypothesized that high throughput contractile analysis of aligned, two-dimensional, micropatterned cardiac muscle bundles generated from iPSC-CMs (CMBs) could be used to functionally adjudicate pathogenicity of *TNNT2* variants. To test this hypothesis, we selected two *TNNT2* variants from the Sarcomeric Human Cardiomyopathy Registry (SHaRe). The *TNNT2* variant R286C is associated with highly-penetrant, severe HCM, while the R278C variant has been classified as a variant of unknown significance but is associated with low penetrance HCM (odds ratio for SHaRE vs general population, 9.9 [95% CI 6.9, 14.1]). These two variants and WT (control) *TNNT2* were cloned into lentiviral vectors. IPSC-CMs were transduced and selected (puromycin) to attain >95% replacement of endogenous *TNNT2*. These iPSC-CMs were then replated onto 8 kPA silicone polymer substrates micropatterned with human fibronectin to generate CMBs. CMBs were imaged using live cell microscopy at two calcium concentrations, and contractile analyses were performed. R278C CMBs exhibited increased maximal fractional shortening vs WT and R286C (5.1±1.5% vs 6.9±1.4% and 6.1±1.4%, $p<0.002$). R286C CMBs exhibited reduced relaxation velocity compared to both R278C and WT (3.5±0.7µm/sec vs 4.3±0.97µm/sec and 4.5±0.97µm/sec, $p<0.002$). The relaxation velocity of R278C increased less than WT when exposed to increased calcium (high/low 1.2±0.3 vs 1.5±0.5, $p<0.01$). R286C had no significant difference in increased relaxation velocity with higher calcium concentration compared to WT or R278C (high/low 1.3±0.4 vs 1.5±0.5 and 1.2±0.3, $p>0.5$). These data demonstrate that contractile analysis of iPSC-derived CMBs can be used to detect phenotypic differences due to cardiomyopathy causing variants. Additionally, different variants may cause distinct effects on contractile and relaxation parameters, as we found in the case of a strongly penetrant *TNNT2* variant (R286C) more profoundly affecting relaxation velocity than a lower penetrance variant (R278C). These findings have implications for future systematic use of functional testing to adjudicate pathogenicity of potential cardiomyopathy associated variants.

Session Title: Molecular Effects of Genetic Variation Poster Session I

PB1117 Gene Expression Predictions from Enformer Show Improved Patterns of Cross-population Portability over Existing Methods

Authors:

S. Gona[1], T. Adeluwa[2], B. Liu[3], R. Madduri[4], H. Im[1]; [1]Univ. of Chicago, Chicago, IL, [2]The Univ. of Chicago, Chicago, IL, [3]Natl. Univ. of Singapore, Singapore, Singapore, [4]Argonne Natl. Lab., Naperville, IL

Advances in prediction of disease risk based on genetics is clinically promising, but the poor portability of current population-based prediction models to non-European populations is a major health equity issue.Notably, much of the intuition of what drives poor portability in existing population-based models doesn't apply to alternative modeling strategies. Enformer, a deep learning model investigated here, is an example of an alternative strategy. Enformer is trained using information across a single genome, rather than across individuals in a population as is typical of current models. As poor portability in polygenic score models stems from cross-population genetic variation patterns such as linkage disequilibrium, it stands to reason that single genome deep learning models could show better portability. Still, no systematic examination has been done to confirm this.Addressing this gap, we examined portability of Enformer gene expression predictions between EUR and AFR ancestries on 448 individuals from the GEUVADIS cohort. We compared these results with the existing expression European population-based prediction method, PrediXcan. We first examined Enformer performance on a set of genes (n=3387) for which confident PrediXcan results can be obtained. We found that PrediXcan has higher overall personalized prediction accuracy in both EUR and AFR than Enformer with a mean R2 of 0.035 against mean R2 of 0.015 for Enformer across all genes (including lowly-heritable genes). This is expected since PrediXcan models are trained to maximize cross-individual correlation over hundreds of individuals, whereas Enformer has no cross-individual training. In terms of portability, defined as the log2 ratio of performance in AFR samples versus EUR samples, Enformer predictions appear less susceptible to European bias. Genes in the upper 50th percentile of prediction accuracy for PrediXcan, which represent the bulk of significantly heritable genes, show mean portability of -0.39 and -0.068 in PrediXcan and Enformer respectively.The difference in portability suggests that population-agnostic training strategies such as that used for Enformer can be used to improve prediction portability to non-European populations.

Session Title: Molecular Effects of Genetic Variation Poster Session II

PB1118 Genetic analysis of HLA polymorphism in tribal and non-tribal populations of Maharashtra, India: A pilot study.

Authors:

P. Mahajan[1], S. Das[2], K. V. Somasundaram[3]; [1]Drake Univ. Coll. of Pharmacy & HS, Des Moines, IA, [2]Ctr. for Biotechnology, Pravara Inst. of Med. Sci., Loni, Maharashtra State, India, [3]Sch. of Publ. Hlth.and Social Med., Pravara Inst. of Med. Sci., Loni, Maharashtra State, India

Our long-term objective is to understand the effects of genomic diversity on the health of different subpopulations in India. Currently we are focusing on employing genomic analysis for reducing adverse drug reactions (ADRs). Recent reports have documented incidence of Steven Johnson Syndrome (SJS)/toxic epidermal necrosis (TEN) in India is ranging between 3%-19.5%. Therefore, we investigated utility of a simple, low-cost genotyping method for HLA-A*31:01, a marker for carbamazepine induced SJS/TEN in the gene-rich HLA locus of human genome. Using the Amplicon Fragment Length Polymorphisms (AFLP) assay we assessed genetic diversity of HLA-A*31:01 in the populations served by the Pravara Institute of Medical Sciences Hospital Loni, India. One hundred family trios (300 individuals) from indigenous tribes (Mahadev Koli and Thakar) as well as non-tribal population from the Bhandardhara region of the Akole Tribal Block of Maharashtra were tested in this study. Primers specific for HLA-A*31:01 were used to amplify a 662 basepair fragment from individual genomic DNA samples and subjected to restriction digestion by BtsCI. Our results confirm utility of the AFLP genotyping assay as a rapid and relatively inexpensive method than the conventional HLA typing for identifying patient risk of carbamazepine related SJS/TEN. Interestingly, only 217 out of the total 300 or roughly 72% samples produced the expected 662bp amplicon. Moreover, only three of the 204 amplicons analyzed exhibited the expected restriction pattern in the BtsCI assay. Using beta globin primers as internal control, we confirmed that lack of amplification of HLA-A*31:01 in ~28% of samples was not due to the presence of some unknown PCR inhibitor in these samples nor a technical failure. Our preliminary results confirm utility of the simple AFLP assay for HLA-A*31:01 genotyping. However, our data strongly indicate substantial heterogeneity even at this single biomarker within the HLA locus in the study populations. Further in-depth studies including complete sequencing of the HLA Locus from the study population are needed to fully understand genetic implications of the potential heterogeneity to individualized medicine.

Session Title: Molecular Effects of Genetic Variation Poster Session III

PB1119 Genetic and Functional Activity Analyses of *ADA2* in Patients with Deficiency of Adenosine Deaminase 2.

Authors:

E. Everest[1], M. Ozkilinc Onen[1], T. Demirci[1], A. Timucin[2], S. Sahin[3], O. Kasapcopur[3], E. Tahir Turanli[2]; [1]Istanbul Technical Univ., Faculty of Sci. and Letters, Dept. of Molecular Biology and Genetics, Istanbul, Turkey, [2]Acibadem Univ., Faculty of Engineering and Natural Sci., Dept. of Molecular Biology and Genetics, Istanbul, Turkey, [3]Istanbul Univ.-Cerrahpasa, Cerrahpasa Med. Faculty, Dept. of Pediatric Rheumatology, Istanbul, Turkey

Deficiency of Adenosine Deaminase Type 2 (DADA2) is an autosomal recessive autoinflammatory disease caused by biallelic mutations in the *ADA2* gene. Here, we analyzed *ADA2* mutations in six patients with DADA2 and measured the total adenosine deaminase (ADA) activity in the peripheral blood mononuclear cells of eight patients, including those six, and five healthy controls using an ADA Activity Assay Kit (Abcam, ab212093). Additionally, we conducted a computational study involving stability analyses, 100 ns molecular dynamics simulations, and molecular mechanics Poisson-Boltzmann Surface Area analyses. Three patients had the homozygous p.G47R variant, two patients had the compound-heterozygous p.G47R/p.G321E variant, and one patient had a novel homozygous variant (c.467C>T) in the *ADA2* gene. Total ADA activity was significantly lower in patients than in healthy controls (p=0.0009). Moreover, ADA activity was significantly lower in patients possessing compound-heterozygous p.G47R/G321E than in patients with homozygous p.G47R (p=0.004). It was previously shown that dimerization of the ADA2 protein is essential for its catalytic activity and secretion and that p.G47R in the dimerization domain affects the stability of the homodimer. Our computational analysis investigating the combined impact of p.G47R and p.G321E on the catalytic activity of ADA2 showed that these mutations, particularly in their compound heterozygous form, can interfere with dimerization of ADA2. Altogether, these results suggest that the p.G321E variant in *ADA2* plays an important role in the catalytic activity of ADA2.

Session Title: Molecular Effects of Genetic Variation Poster Session I

PB1120 Genetic determinants of metabolic profiles of multiple dietary patterns and type 2 diabetes risk

Authors:

H. Yun[1], J. Hu[1], V. Sarsani[2], Z. Wang[1], F. Wang[1], O. Zeleznik[3], A. Eliassen[3], F. Hu[1], L. Liang[2], J. Li[4]; [1]Harvard T.H. Chan Sch. of Publ. Hlth., BOSTON, MA, [2]Harvard T.H. Chan Sch. of Publ. Hlth., Boston, MA, [3]Channing Div. of Network Med., Dept. of Med., Brigham and Women's Hosp. and Harvard Med. Sch., BOSTON, MA, [4]Brigham & Women s Hosp., Harvard Med. Sch., Boston, MA

Healthy dietary patterns are critical for type 2 diabetes (T2D) prevention, but individuals may respond differently to the same diet. Circulating metabolites may capture dynamics in metabolic homeostasis as a result of combined effects of dietary intakes and other risk factors. In prior studies examining 6 well-established dietary patterns (including a Mediterranean diet [AMED], alternate healthy eating index [AHEI], an anti-hypertensive diet [DASH], and 3 plant-based dietary indices [PDIs]), we identified a metabolomic signature (MS) that robustly assessed the adherence to each of the 6 dietary pattern and predicted T2D risk. However, a large proportion of variation in the MS cannot be fully explained by dietary intakes. Here, given the role of genetics in modulating dietary metabolism, we performed genome-wide association studies for the 6 MS among 6669 participants in the Nurses' Heath Studies and Health Professionals Follow-up Study. Genetic variants explained an average of 26.5% (h^2 =16.5-34.5%) of the total variance in the MS ($P < 0.05$). We identified 7 loci associated with the MS; these include, *FADS1/2* (encode the fatty acid desaturase; $P < 3.0\times10^{-11}$) associated with all dietary patterns, *ALMS1* and *ALMS1P1* (involve in microtubule/cilia function that are vital in energy metabolism; $P < 3.4\times10^{-14}$) associated with five dietary patterns, *CERS4* (involve in ceramide synthesis, $P = 1.4\times10^{-15}$) associated with plant-based diet, and *SLCO1A2*, *SLCO1A2B1*, *DGCR5* ($P < 2.8\times10^{-8}$). Lead variants on *FADS1/2*, *ALMS1/ALMS1P1*, and *CERS4* were significantly associated with expressions of the encoded genes in tissues relevant to glucose/insulin response (e.g., pancreas, liver, muscle, and adipose) and endocrine regulation (e.g., hypothalamus, thyroid, and pituitary) (all $P < 2.7\times10^{-5}$). Genome-wide gene-diet interaction analysis revealed a locus, *OLFM4* (involved in color mucosal defense), only influencing MS through interacting with AHEI ($P = 1.9\times10^{-8}$). Leveraging a recent large-scale GWAS for T2D (80154 cases and 853816 controls), a two-sample mendelian randomization analysis showed that, the genetically predicted MS of 4 dietary patterns (i.e., AMED, AHEI, PDI, and unhealthy PDI) were associated with 8-11% difference in T2D risk (adjusted $P < 0.05$). In summary, the MS of dietary patterns may reflect the combined effects of both dietary intakes and metabolic responses determined by genetics, and potentially point to causal metabolic pathways related to T2D pathophysiology. Multi-omics profiling could identify biological factors affecting individual dietary metabolism and inform personalized nutritional strategies for effective diabetes prevention.

Session Title: Molecular Effects of Genetic Variation Poster Session II

PB1121 Genetic determinants of the human plasma proteome in a Chinese population reveal substantial ancestry diversity and identify new drug development opportunities

Authors:

S. Said[1], A. Pozarickij[1], S. Morris[1], C. Kartsonaki[1], L. Liming[2], R. Clarke[1], I. Y. Millwood[1], Z. Chen[1], R. G. Walters[1], China Kadoorie Biobank Collaborative Group; [1]Univ. of Oxford, Oxford, United Kingdom, [2]Peking Univ., Peking, China

The proteome is fundamental to understanding human biology and disease. Genetic association analyses of plasma levels of 1,451 proteins in 3,974 Chinese adults were conducted and compared with the corresponding analysis in adults of European ancestry. In Chinese, we identified pQTLs for 1,081 proteins, of which 743 (69%) had at least one *cis*-pQTL. Fine-mapping defined credible sets for 3,343 independent signals, of which 25% had no overlap with pQTL signals identified for the same proteins in Europeans. We further assessed 741 *cis*-pQTLs in phenome-wide Mendelian randomization analyses and identified significant associations for 199 protein-phenotype pairs. Out of the 37 identified pQTL-disease pairs, 13 derived from East Asian-specific analyses. Evaluation of current drug development highlighted opportunities for repurposing of 13 drugs either under trial or already approved, and for exploration of 19 potential drug targets. These findings emphasise the importance of extending genome-wide plasma proteomic analyses to non-European ancestry populations, for identification of potential novel biomarkers and drug targets for major diseases.

Session Title: Molecular Effects of Genetic Variation Poster Session III

PB1122 Genetic effects on chromatin accessibility and nucleosome positioning using snATAC data across 281 skeletal muscle biopsies.

Authors:

X. Wang[1], C. C. Robertson[1,2], A. Varshney[1], N. Manickam[1], P. Orchard[1], M. Laakso[3,4], J. Tuomilehto[5], T. A. Lakka[3], K. L. Mohlke[6], M. Boehnke[7], L. Scott[7], H. A. Koistinen[5], F. S. Collins[2], S. C. J. Parker[1,7,8]; [1]Dept. of Computational Med. and Bioinformatics, Univ. of Michigan, Ann Arbor, MI, [2]Natl. Human Genome Res. Inst., NIH, Bethesda, MD, [3]Inst. of Clinical Med., Internal Med., Univ of Eastern Finland, Kuopio, Finland, [4]Dept. of Med., Kuopio Univ Hosp., Kuopio, Finland, [5]Dept. of Publ. Hlth.and Welfare, Finnish Inst for Hlth.and Welfare, Helsinki, Finland, [6]Dept. of Genetics, Univ North Carolina, Chapel Hill, NC, [7]Dept. of Biostatistics and Ctr. for Statistical Genetics, Univ. of Michigan, Ann Arbor, MI, [8]Dept. of Human Genetics, Univ. of Michigan, Ann Arbor, MI

Insulin resistance in skeletal muscle is a central process contributing to type 2 diabetes (T2D) and related metabolic traits. Among 13 cell types we identified in single-nucleus RNA (snRNA) and ATAC (snATAC) data from skeletal muscle, type 1 muscle fiber was the most abundant. Studying genetic variants that affect chromatin accessibility and nucleosome positioning in type 1 muscle fiber can provide mechanistic insights at GWAS signals for T2D and related traits, and allow for causal inference of local chromatin dynamics (e.g., which transcription factors (TFs) causally position nucleosomes?). We profiled the chromatin landscape using snATAC of muscle from 281 individuals from the FUSION project and focused our analyses on type 1 fibers. We used two methods to infer nucleosome free regions (NFR), which represent TF-occupied regions, and nucleosomal positioning, which represents nucleosomal phasing that may be induced by a primary TF-binding event. Using a hidden Markov model (HMMRATAC), we inferred NFR (N=44,763 peaks) and their adjacent nucleosomal chromatin regions (N=89,526 peaks). Using the NucleoATAC method, which used the 2D fingerprint from the ATAC data to generate nucleosome maps, we found corresponding NFR (N=66,041 peaks) and nucleosomal positions (N=239,944 peaks). For each of the samples, we used Multi-Otsu statistical thresholding to categorize whether the ATAC reads originated from nucleosome-free or nucleosome-bound fragments. Subsequently, we utilized the reads from nucleosome-free fragments to create the NFR count matrix, and we used the reads originating from nucleosome-bound fragments to build the nucleosomal count matrix. Using QTLtools, we performed a cis-QTL analysis (cis window of 10kb) on both NFR (nfrQTL) and nucleosome positioning (nucQTL). From HMMRATAC, we identified 12,819 nfrQTLs and 6,544 nucQTLs at 5% FDR. Through the NucleoATAC method, we identified 14,417 nfrQTLs and 19,265 nucQTLs. We observed significant enrichment in both the nfrQTL HMMRATAC (ln-enrichment=2.03, CI=[0.967, 2.70]) and NucleoATAC (ln-enrichment=3.30, CI=[2.59, 3.83]) predicted NFR peaks when analyzing with T2D GWAS summary statistics. We are now performing formal colocalization analyses to assess whether the nfrQTLs and the nucQTLs are driven by the same variant or if nfrQTLs mediate the nucQTLs. Meanwhile, we will investigate the overlapping nfrQTL and nucQTL signals with T2D GWAS signals to nominate locus-specific chromatin mechanisms. Overall, these analyses provide us with a deeper understanding of regulatory mechanisms underlying T2D genetic risk in skeletal muscle, which could inform T2D diagnosis and treatment in the future.

Session Title: Molecular Effects of Genetic Variation Poster Session I

PB1123 Genetic epistasis between *RET* and *ELP1* explains gastrointestinal motility dysfunction in Familial Dysautonomia patients.

Authors:

S. Chatterjee[1], M. Chaverra[2], F. Lefcort[2]; [1]NYU Sch. of Med., New York, NY, [2]Montana State Univ., Bozeman, MT

Familial dysautonomia (FD) is a rare disorder in the Ashkenazi-Jewish population that is characterized by developmental failure of the autonomic nervous systems; key phenotypes include cardiovascular instability, autonomic crises and gastrointestinal motility (GI) dysfunction. Almost all cases are due to a founder point mutation in the *ELP1* gene though the phenotypic diversity points to the presence of undiscovered modifier alleles. Pathology studies have revealed a severe reduction of enteric neurons in FD patients, raising the possibility that there is an overlap between the mechanisms leading to GI motility dysfunction in FD and a genetically distinct disorder, Hirschsprung disease (HSCR). HSCR is a developmental disorder marked by aganglionosis of the GI tract resulting from the proliferative failure of the enteric nervous system (ENS). HSCR is associated with pathogenic variants in 24 genes comprising a gene regulatory network (GRN) controlling the development of the ENS, but ~50% of patients contain pathogenic variants in the gene *RET*. Studies in HSCR families have demonstrated linkage to the *ELP1* locus, along with coding variants in *RET* providing evidence of genetic interaction between the 2 genes. To tease apart the cellular changes in the ENS in FD and HSCR, we created knockout mouse models of *Elp1* and *Ret* followed by single cell RNA-seq in E14.5 embryonic gut. We show that both *Ret* and *Elp1* deficiency leads to significant loss of inhibitory motor neurons driven by proliferative defects in these cells marked by loss of expression of DNA replication (*Pcna*, *Mcm3*), and cell cycle regulation (*Rgcc*, *Mad2l1*) genes. Other major neuronal and glial cells are largely unaffected. There are 131 common transcriptionally affected genes in both disease models among them 52 transcription factors including *Sox10* a ENS regulator. To recapitulate the effects in human cells, we independently knocked out *ELP1* and *RET* in the neuroblastoma cells SK-N-SH using CRISPR/Cas9 followed by RNA-seq. *ELP1* heterozygous cells have 20% and 15% reduction of *RET* and *SOX10* expression (both $p \leq 0.01$) along with expression changes in multiple ENS expressed genes. Conversely, *RET* heterozygous cells show a 50% ($p=1.8\times10^{-3}$) reduction in *ELP1* expression along with significant ($p \leq 0.01$) drop in expression of *SOX10* (80%) and *EDNRB* (25%) another HSCR gene, along with multiple ENS genes. These mutant cells also display significant proliferative defects measured by WST1 assay. This work provides a conceptual framework for the discovery of modifier genes with functional roles in rare diseases whose individual mutational burden might not cross statistical threshold in genome scale sequencing studies.

Session Title: Molecular Effects of Genetic Variation Poster Session II

PB1124 † Genetic insights into causal genes and directionality through large-scale colocalization of metabolite and protein QTLs

Authors:

J. Chiou[1], X. Yin[2], B. Ryan[3], M. Laakso[4], X. Wen[3], M. Boehnke[3], E. Fauman[1]; [1]Pfizer, Cambridge, MA, [2]Nanjing Med. Univ., Nanjing, China, [3]Univ. of Michigan, Ann Arbor, MI, [4]Univ. of Eastern Finland, Kuopio, Finland

Translating findings from large-scale studies of molecular QTLs into biological insight is challenging because it requires a thorough understanding of genetic mechanisms, including the causal genes and their direction of effect. The intersection of metabolite QTLs (metabQTLs) and protein QTLs (pQTLs) provides a unique opportunity to interrogate metabolite genetic mechanisms at scale because metabolite regulation is intrinsically rooted in biochemistry, enabling direct comparisons between proteins encoding enzymes and their metabolic substrates and products. To systematically identify shared genetic signals for metabQTLs and pQTLs, we first applied fine-mapping methods (SuSiE and DAP-G) to detect 4,486 conditionally independent signals for 1,391 metabolites assayed in plasma from 10k Finnish METSIM cohort individuals and 44,749 signals for 2,682 proteins assayed in plasma from 35k European UK biobank participants. We performed colocalization tests with coloc for 441,431 pairs of metabQTL and pQTL signals and identified 75,106 colocalized pairs at $PP_{H4} > 0.8$ and $r^2 > 0.6$ between lead variants, nearly all of which involved *trans* pQTLs. Of all colocalized pairs, 76.2% (57,208) were also found using fastENLOC (LCP > 0.5). For 459 colocalized pairs involving *cis* pQTLs, we compared the observed directionality from pQTLs with predicted directionality based on expert knowledge in biochemistry. In a blinded analysis, the predicted biochemical and observed pQTL directionality were concordant for 82% (189/231) of pairs involving *cis* pQTLs where the metabolite was predicted as a substrate or product. For example, decreased CDA protein levels were also associated with increased 2'-O-methycytidine (substrate) and decreased uridine (product) levels, concordant with the biochemical activity of CDA as a cytidine deaminase. Across 29 pairs of proteins and metabolites with multiple independent colocalized *cis* signals, 26 were directionally consistent across all signals. As a novel example, decreased MRI1 protein levels were consistently associated with increased levels of an uncharacterized metabolite (X-12411) across 5 independent signals. The observed directionality, combined with the known function of MRI1, suggested X-12411 was likely a substrate similar to methylthioribose-1-phosphate. We are currently conducting additional analyses to map the causal genes and directionality for colocalized pairs involving *trans* pQTLs. In sum, our results reveal genetic mechanisms for hundreds of metabolites, and more broadly, how genetics can be used in combination with biochemistry to uncover novel findings.

Session Title: Molecular Effects of Genetic Variation Poster Session III

PB1125 Genetic susceptibility to inflammatory bowel disease impacts on epithelial barrier integrity.

Authors:

I. Hébert-Milette[1,2], C. Lévesque[2], J. Paquette[2], M-È. Rivard[2], P. Goyette[2], G. Charron[2], J. Rioux[2,1]; [1]Université de Montréal, Montréal, QC, Canada, [2]Montreal Heart Inst., Montreal, QC, Canada

Introduction: Intestinal permeability is increased in unaffected 1st degree relatives of patients with inflammatory bowel disease (IBD), and is considered a risk factor for the development of IBD, likely increasing the interactions between intestinal microorganisms and the immune system. We recently reported that *C1ORF106*, a gene located within a genomic region associated with IBD, regulates epithelial permeability. We further demonstrated that a rare coding variant within *C1ORF106* (p.Y333F) decreases protein stability and that lower levels of C1ORF106 protein leads to altered stability of adherens junctions and to an increase in intestinal epithelial permeability.
Hypothesis: In addition to altering adherens junctions, we believe that C1ORF106 is also involved in the regulation of tight junction (TJ) formation, which impacts on epithelial permeability and cell polarity establishment.
Objectives: The objectives of the project were to (a) validate the impact of C1ORF106 on TJ formation and cell polarity establishment and (b) verify the impact of *C1ORF106* IBD-associated variants on intestinal barrier integrity.
Results: We observed that knocking down the expression of *C1ORF106* in Caco-2 cells leads to a number of phenotypes in human epithelial monolayer (2D) and spheroid (3D) cultures that are associated with alterations in TJs. Specifically, when studying the dynamic reformation of TJs in 2D cultures after transient withdrawal of calcium, which is required for TJ stability, we observed that lower levels of *C1ORF106* resulted in (1) decreased recovery of barrier function as measured by transepithelial electrical resistance (TEER) and (2) an alteration of TJ protein localization. Moreover, in 3D cultures, we observed an altered spheroid formation associated with an impaired epithelial cell polarization. We also observed an impaired spheroid permeability suggesting an altered TJ function. In addition, our studies in human induced pluripotent stem cell-derived epithelial cultures support that Y333F heterozygotes also have an altered spheroid formation.
Conclusion: Our observations indicate an important role of C1ORF106 in epithelial barrier integrity through the regulation of TJ formation and of cell polarity establishment. TJ formation is important for epithelial repair after an injury and its dysregulation impairs the formation of an impermeable epithelial barrier, which likely facilitates the passage of microorganisms and the induction and maintenance of intestinal inflammation.

Session Title: Molecular Effects of Genetic Variation Poster Session I

PB1126 Genetic variants associated with within-individual gene expression dispersion at single-cell resolution reveal a new mechanism of genome regulation

Authors:

A. Xue[1,2], S. Yazar[1], J. Alquicira-Hernandez[1,3], A. Cuomo[1,4], J. Powell[1,5]; [1]Garvan-Weizmann Ctr. for Cellular Genomics, Garvan Inst. of Med. Res., Sydney, Australia, [2]Sch. of BioMed. Sci., Univ. of New South Wales, Sydney, Australia, [3]Inst. for Molecular BioSci., The Univ. of Queensland, Brisbane, Australia, [4]Ctr. for Population Genomics, Garvan Inst. of Med. Res., Sydney, Australia, [5]UNSW Cellular Genomics Futures Inst., Univ. of New South Wales, Sydney, Australia

The expression levels of a gene can vary substantially across cells, even when they are the same cell type. Understanding the relationship between genetic variation and gene expression is vital to dissecting mechanisms of genome regulation. However, the genetic control of gene expression variability among the cells within individuals has yet to be extensively examined, primarily due to limited samples and statistical challenges such as adjusting mean-variance dependence. Here, we present a genome-wide association analysis to identify genetic variants associated with the intra-individual variation of single-cell gene expression (intra-veQTLs). Using single-cell RNA-seq data of ~1.2 million peripheral blood mononuclear cells from ~1,000 human donors (OneK1K cohort), we identified 14 ~ 3,488 genes (referred to as vGenes) with significant intra-veQTLs (q-value < 0.05) across different blood cell types, 2,103 of which were shared across more than one cell type. We further detected 55 intra-veQTLs (in 34 unique genes) showing significant association with intra-individual dispersion (intra-deQTLs) regardless of the mean-variance dependence, and these dGenes were enriched in the biological process relevant to immune response and viral infection. We highlight the SNP rs1131017, a significant intra-deQTL in the 5'UTR of *RPS26*, which shows a consistent direction of dispersion effect in all 14 cell types and higher intra-individual dispersion level is associated with lower auto-immune disease risk, including rheumatoid arthritis and type 1 diabetes. Our study demonstrates an efficient and robust statistical method to identify genetic variants associated with gene expression variability and how these associations and their involved pathways confer auto-immune disease risks. This is also the first time that genetic effects on the gene expression dispersion level across single cells were identified. We anticipate our analytical framework will open a new gate to unravelling the genetic regulation of gene expression at the single-cell resolution, thereby advancing our understanding of complex biological processes.

Session Title: Molecular Effects of Genetic Variation Poster Session II

PB1127 Genetic variation in a distal enhancer modifies cardiac transcription factor HAND2 expression and function

Authors:

N. Xie[1], Y. Feng[1], S. Tishkoff[1,2,3]; [1]Dept. of Genetics, Perelman Sch. of Med., Univ. of Pennsylvania, Philadelphia, PA, [2]Dept. of Biology, Univ. of Pennsylvania, Philadelphia, PA, [3]Ctr. for Global Genomics and Hlth.Equity, Univ. of Pennsylvania, Philadelphia, PA

Genome-wide association studies (GWAS) have identified a large number of genetic variants that are associated with increased risk of cardiovascular disease. However, the mechanism by which these variants contribute to disease risk is not well understood. Expression of cardiac genes is under the regulatory control of distal enhancers, and genetic variation in enhancers may modulate their activity. To investigate the functional impacts of genetic variants associated with cardiovascular diseases, we overlapped genetic variants from published GWAS datasets with genomic regions predicted to be functional regulatory regions, based on human heart-derived DNase-seq, ATAC-seq, and ChIP-seq data for the cardiac transcription factors GATA4, TBX5, and NKX2.5. Intersection of these data sets identified GWAS variants located within predicated enhancer regions close to genes important for heart development. Specifically, we identify a variant located in a predicted heart-specific enhancer of the HAND2 (heart and neural crest derivatives expressed 2) gene, a key transcription factor for cardiac morphogenesis. This variant is located 100 kb upstream of HAND2 gene and is associated with atrial fibrillation. Using a luciferase reporter assay, we found this predicted enhancer displayed significant regulatory activity in human cardiomyocytes derived from embryonic stem cells. Moreover, the variant markedly reduced enhancer activity, indicating its essential role in modulating enhancer function in cardiomyocytes. CRISPR gene editing to inhibit or remove the enhancer that harbors this variant markedly reduced HAND2 expression in human cardiomyocytes. We are further investigating the effect of this variant on cellular function and cardiac development. This study demonstrates an integrated approach for identifying functionally important variants that may play a role in cardiovascular disease. Funded by NIH grant 1U01HG012047-01, 1R35GM134957, R01AR076241, and ADA 1-19-VSN-02.

Session Title: Molecular Effects of Genetic Variation Poster Session III

PB1128 Genetic, transcriptomic, and proteomic perturbations linked to schizophrenia converge on neuron-specific protein interactomes of rare variant schizophrenia risk genes.

Authors:

Y-H. Hsu[1,2], G. Pintacuda[1,2], P. Páleníková[1,2], J. Biagini[1,2], D. Mena[1,2], J. Ching[1,2], E. Crouse[1], H. Gebre[1], T. Botts[1,2], T. Werge[3,4,5], A. Kirkeby[6], O. Pietilainen[1,7], N. Fornelos[1,2,8], R. Nehme[1], K. Lage[1,2,3,8]; [1]Stanley Ctr. for Psychiatric Res., Broad Inst. of MIT and Harvard, Cambridge, MA, [2]Novo Nordisk Fndn. Ctr. for Genomic Mechanisms of Disease, Broad Inst. of MIT and Harvard, Cambridge, MA, [3]Inst. of Biological Psychiatry, Mental Hlth.Ctr. Sct. Hans, Mental Hlth.Services Copenhagen, Roskilde, Denmark, [4]iPSYCH, Lundbeck Fndn. Initiative for Integrative Psychiatric Res., Aarhus, Denmark, [5]Dept. of Clinical Med., Faculty of Hlth.and Med. Sci., Univ. of Copenhagen, Copenhagen, Denmark, [6]Novo Nordisk Fndn. Ctr. for Stem Cell Med. and Dept. of NeuroSci., Univ. of Copenhagen, Copenhagen, Denmark, [7]NeuroSci. Ctr., Helsinki Inst. of Life Sci., Univ. of Helsinki, Helsinki, Finland, [8]Dept. of Surgery, Massachusetts Gen. Hosp., Boston, MA

Schizophrenia (SCZ) is a severe psychiatric disorder with hundreds of genetic variants contributing to its risk across the allele frequency spectrum. The Schizophrenia Exome Sequencing Meta-Analysis (SCHEMA) study recently reported 32 rare variant SCZ risk genes at a 5% false discovery rate and highlighted overlapping genetic signals between SCZ and other neurodevelopmental disorders. Characterizing the functional roles and pathways of the high-confidence SCHEMA genes in relevant cell types of the human brain could elucidate the connections between these disorders and facilitate development of effective treatments and therapeutics. Therefore, we performed interaction proteomics for 14 SCHEMA genes in human iPSC-derived neural progenitor cells (NPCs) and excitatory neurons (ExNs). The resulting protein-protein interaction (PPI) networks contain ~89% newly reported PPIs, reflect cell-type-specific biology in NPCs vs. ExNs, and are the convergence point for genetic, transcriptomic, and proteomic perturbations associated with SCZ. In terms of genetics, both the NPC-specific and ExN-specific networks are enriched for rare variant risks of developmental and autism spectrum disorders, while only the ExN-specific networks are enriched for rare variant risk of schizophrenia. In terms of transcriptomics, the ExN-specific networks are enriched for differentially expressed genes identified in superficial and deep layer cortical neurons of individuals with SCZ vs. controls. To assess whether this type of convergence is also present at the proteomic level, we performed proteomic profiling of iPSC-derived neurons from 22q11.2 deletion carriers with SCZ diagnosis vs. controls. Strikingly, several ExN-specific networks that were down-regulated at the RNA level in postmortem brains of SCZ individuals with varied genotypes were also down-regulated at the protein level in iPSC-derived neurons from SCZ individuals carrying the 22q11.2 deletion. Overall, our findings recapitulate the known genetic correlations between SCZ and neurodevelopmental disorders using cell-type-specific PPI data, implicate the ExN-specific networks as a convergent point that concentrates both transcriptional and translational dysregulation in SCZ, and prioritize numerous genes and interactions in these networks for further functional investigation of SCZ-related biology.

Session Title: Molecular Effects of Genetic Variation Poster Session I

PB1129 Genetically predicted efficacy and safety profiles of anticoagulant targets.

Authors:

E. Gagnon; Unversité Laval, Quebec, QC, Canada

Introduction. Anticoagulant therapies allow prevention of life-threatening thrombosis at the expense of increased bleeding risk. Optimal anticoagulant targets must be safe (cause low bleeding risk) and efficacious (prevent thrombosis risk). **Objective.** To prioritize coagulation cascade proteins with the most favourable efficacy and safety profiles using Mendelian randomization. **Methods**. We investigated the association of genetically predicted reductions of 27 coagulation factors levels with risk of venous thromboembolism, cardioembolic and ischemic stroke, bleeding outcomes, and lifespan. We leveraged genetic data from three blood proteins genome-wide association study (GWAS) datasets and performed Mendelian randomization and genetic colocalization. We also include bulk RNA sequencing data from explanted liver samples obtained from 246 participants of the IUCPQ Obesity Biobank and GWAS summary statistics on 1589 diseases (>1000 disease cases) from FinnGen. **Results.** We found converging evidence that thrombin and Factor 11 (FXI) had the most optimal efficacy versus safety profiles. Thrombin triggers FXI activation which ultimately promotes coagulation. Thrombin inhibition was more efficacious to prevent venous thromboembolism than FXI inhibition (odds ratio per 1 standard deviation (SD) lower blood protein: 0.45 95% CI=0.38-0.54, p=1.9e-19 for thrombin vs. 0.67 95% CI=0.61-0.73, p=3.4e-19 for F11). Thrombin inhibition was also mor efficacious to prevent cardioembolic stroke than FXI inhibition (0.58 95% CI=0.50-0.68, p=3.2e-12 for thrombin vs. 0.81 95% CI=0.77-0.85, p=2.5e-17 for FX1). However, thrombin inhibition increased bleeding risk (1.09 95% CI=1.00-1.17, p=4.1e-02), whereas FXI inhibition did not (1.00 95% CI=0.97-1.02, p=7.8e-01). To further assess thrombin and FXI inhibition safety, we performed a phenom-wide Mendelian randomization on the effect of genetically predicted thrombin and FXI on 1,589 diseases in the FinnGen cohort. Thrombin and FXI genetic inhibition appeared to be overall safe. Both proteins are specifically expressed in the liver and are then secreted in the bloodstream. We found that hepatic gene expression was causally associated with thrombotic diseases but had lower effect than blood protein levels. **Conclusions.** Thrombin inhibitors are already being used for prophylaxis and treatment of venous thromboembolism and acute coronary syndrome. Phase 3 trials investigating the health effects of FXI inhibition are underway. This Mendelian randomization study supports thrombin and FXI inhibition as efficacious and safe therapeutic targets.

Session Title: Molecular Effects of Genetic Variation Poster Session II

PB1130 Genetics of Long QT Syndrome: iPSC-Derived Cardiomyocyte Approach

Authors:

A. Shen, D. W. Mitchell, B. Kroncke; Vanderbilt Univ. Med. Ctr., Nashville, TN

Long QT syndrome (LQTS) is associated with sudden unexplained death in young populations. LQTS is an inherited heart arrhythmia syndrome characterized by a prolonged corrected QT interval on an electrocardiogram. Individuals with LQTS are at high risk for life-threatening arrhythmia torsades de pointes. Although LQTS is responsible for a significant proportion of sudden unexplained deaths in the young and sudden infant death syndrome cases, its genetic basis and underlying mechanisms remain poorly understood. LQTS often presents with incomplete penetrance and variable expressivity, indicating a complex interplay of monogenic and polygenic factors. This study aims to generate human induced pluripotent stem cell-derived cardiomyocytes from individuals with extreme polygenic liability for QT interval duration to gain insights into the mechanisms underlying QT variability and the interaction between polygenic and monogenic variation in driving LQTS. To address the challenge of incomplete penetrance and heterogeneous expressivity, we investigated iPSC lines derived from individuals with LQTS liabilities significantly deviating from the mean (3+ standard deviations). We collected and reprogrammed peripheral blood mononuclear cells (PBMCs) from these patients into iPSCs. Specifically, we generated iPSC lines: two with low LQTS risk scores of 0.35% and 0.1%, two with mean scores of 50%, and two with high scores of 99.8% and 99.18%. The resulting iPSCs were differentiated into cardiomyocytes and interrogated for conduction properties. We validated pluripotency for all cell lines. We compared RNA transcript changes between high and low QTc-polygenic risk cell lines, focusing on genes associated with QT interval GWAS loci. We also compare in vitro correlates of prolonged QT interval, action potential recordings, extracellular field potential using CardioExcyte96. By interrogating the biological basis of underlaying mechanisms and the impact of small effect-size genetic variants, we aim to unravel mechanisms underlying the presentation of LQTS. This study expands our understanding and provides a valuable template for investigating other disorders with significant polygenic and monogenic contributions. Through the use of iPSC-derived cardiomyocytes, this study offers insights into the complex genetic and mechanistic interplay of Long QT syndrome. By elucidating the influences of various factors, we aim to enhance risk assessment, identify therapeutic targets, and enable early interventions for patients. Our findings will contribute to a deeper understanding of disease mechanisms and advance personalized medicine strategies.

Session Title: Molecular Effects of Genetic Variation Poster Session III

PB1131 Genome Wide CRISPR Screen in a*TUBB* Hypomorphic Background Reveals Novel Genetic Interactions

Authors:

D. Thomson Taylor[1,2], H. J. Lim[1], P. Mero[2], K. Chan[2], K. Brown[2], M. Costanzo[1], M. Billmann[3], C. Myers[4], B. Andrews[1], C. Boone[1], J. Moffat[2,1]; [1]Univ. of Toronto, Toronto, ON, Canada, [2]The Hosp. for Sick Children, Toronto, ON, Canada, [3]Univ. of Bonn, Bonn, Germany, [4]Univ. of Minnesota, Minneapolis, MN

Background: GWAS have successfully linked genetic variation to disease phenotypes however there remains a large percentage of heritability that cannot be explained. The interaction between multiple genetic variants has been able to explain the presence of phenotypes that are otherwise absent when only a subset of the genetic variants is present. Using a genetic interaction (GI) screening approach, a global GI network of the *Saccharomyces Cerevisiae* model organism has been constructed. This demonstrated that genes essential for cell fitness (EGs) act as GI hubs and can predict the functions of uncharacterized genes with highly correlated GI profiles. The human genome is larger and more complex than lower eukaryote genomes and therefore to condense the human GI network we are profiling GIs of EGs in a model human cell line. **Methods:** As a pilot study we selected *TUBB*, a common essential gene, for GI profiling. We utilized fitness data from genome wide CRISPR fitness screens to identify guide RNAs (gRNAs) that can generate hypomorphic variants of *TUBB* in the haploid chronic myelogenous leukemia cell line HAP1. We used the CRISPR-Cas9 gene disruption system to construct stable HAP1 cell lines harbouring hypomorphic mutations in *TUBB.* We then performed a fitness screen using the genome wide Toronto Knockout Library which contains gRNAs that target 18,053 human protein coding genes to identify GIs of *TUBB.* We also performed chemical GI screens with tubulin-destabilizing drugs nocodazole and colchicine to analyze the quality of the *TUBB* GI profile. **Results:** Our *TUBB* mutant cell lines are highly sensitive to the tubulin-destabilizing drugs nocodazole and colchicine. By RT-qPCR analysis we confirmed that all *TUBB* mutant cell lines had decreased transcript levels and the *TUBB* paralogue, *TUBB4B*, had increased transcript levels. This reliance on *TUBB4B* for cell fitness was confirmed by analysis of the GI profile of the *TUBB* mutant cell line where *TUBB4B* was one of the top negative GIs of *TUBB.* Additionally, analysis of the top GIs showed strong enrichment for mitosis and microtubule-related functions. We also discovered *TUBB* GIs with genes not reported to be functionally associated with *TUBB* in the literature as well as genes with uncharacterized or poorly characterized functions. **Conclusion:** This pilot study demonstrated the feasibility of constructing and profiling GIs of EG mutants in a human model cell line. Our screen uncovered known functional interactors of *TUBB* and identified an uncharacterized gene that is synthetic lethal with our *TUBB* mutation.

Session Title: Molecular Effects of Genetic Variation Poster Session I

PB1132 Genome-wide annotation of fitness-linked regulatory elements

Authors:

M. ter Weele[1], T. Klann[1], A. Barrera[1], A. Ettyreddy[1], R. Rickels[1], J. Bryois[2], S. Jiang[1], S. Adkar[1], N. Iglesias[1], P. Sullivan[3], T. Reddy[1], A. Allen[4], G. Crawford[1], C. Gersbach[1]; [1]Duke Univ., Durham, NC, [2]Univ Geneva, Geneva, Switzerland, [3]Univ North Carolina, Chapel Hill, NC, [4]Duke Univ, Durham, NC

Noncoding regulatory elements control gene expression and govern all biological processes. Epigenomic profiling assays have identified millions of putative regulatory elements, but systematically determining the function of each of those regulatory elements remains a substantial challenge. Here we adapt CRISPR-dCas9-based epigenomic regulatory element screening (CERES) technology to screen all 111,619 putative non-coding regulatory elements defined by open chromatin sites in human K562 leukemia cells for their role in regulating essential cellular processes and proliferation. In an initial screen with 1,084,704 gRNAs, we implemented a new analysis framework to quantify element-level significance and effect size and discovered approximately 11,000 regulatory elements with evidence of impact on cell fitness. These essential elements were, on average, closer to essential and more highly expressed genes and were enriched for stronger chromatin contacts, H3K27ac signal, and chromatin accessibility. However, fitness-linked elements were also diverse, spanning a wide range of chromatin states and chromHMM annotations including candidate enhancers, insulators, and Polycomb repressed elements. RNA sequencing of significant gRNAs revealed examples of perturbations affecting fitness through various pathways, including differentiation, cell metabolism, and immune activation. We validated over 1,000 of the element hits in a secondary screen in K562 cells, evaluated cell-type specificity in a second cancer cell line, and identified target genes of 350 distal regulatory elements using epigenetic perturbations combined with single cell RNA-seq. This comprehensive and quantitative genome-wide map of essential regulatory elements represents a framework for extensive characterization of noncoding regulatory elements that drive complex cell phenotypes and for prioritizing non-coding genetic variants that may contribute to common traits and disease risk.

Session Title: Molecular Effects of Genetic Variation Poster Session II

PB1133 Genome-wide meta-analysis of methylation quantitative trait loci from 1,721 placentas: characterization and implications in human complex disease.

Authors:

A. Cilleros Portet[1], S. Marí[1], J. Lepeule[2], D. Czamara[3], K. M. Bakulski[4], L. Dieckmann[3], C. Breeze[5], J. M. Tuhkanen[6], J. F. Dou[4], M. F. Fernández[7], B. Heude[8], M. Bustamante[9], T. M. Everson[10], J. Chen[11], C. Marsit[10], J. Bilbao[1], C. Lesseur[11], N. Fernandez-Jimenez[1], Pregnancy And Childhood Epigenetics Consortium (PACE); [1]Univ. of the Basque Country (UPV/EHU), Leioa, Spain, [2]Univ. Grenoble Alpes, Inserm, CNRS, Team of Environmental Epidemiology Applied to Reproduction and Respiratory Hlth.(IAB), Grenoble, France, [3]Max-Planck-Inst. of Psychiatry, Dept. of Translational Res. in Psychiatry, Munich, Germany, [4]Dept. of Epidemiology, Sch. of Publ. Hlth., Univ. of Michigan, Ann Arbor, MI, [5]Natl. Cancer Inst., Rockville, MD, [6]Dept. of Psychology and Logopedics, Univ. of Helsinki, Helsinki, Finland, [7]BioMed. Res. Ctr. & Sch. of Med., Univ. of Granada, Inst. de Investigación Biosanitaria de Granada (ibs.GRANADA), Consortium for BioMed. Res. in Epidemiology and Publ. Hlth.(CIBERESP), Granada, Spain, [8]Université de Paris, Ctr. for Res. in Epidemiology and Statistics (CRESS), INSERM, INRAE, Paris, France, [9]Inst. for Global Hlth.(ISGlobal), Barcelona, Spain, [10]Gangarosa Dept. of Environmental Hlth., Rollins Sch. of Publ. Hlth.at Emory Univ., Atlanta, GA, [11]Dept. of Environmental Med. and Publ. Hlth., Icahn Sch. of Med. at Mount Sinai, New York, NY

The placenta is of utmost importance in maternal-fetal interactions, and the central organ to study in the context of the Developmental Origins of Adult Health and Disease hypothesis that postulates that early-life environments can impact later-life health. Growing evidence indicates that the associations among an adverse intrauterine environment, genetics, and health outcomes could be driven by placental epigenetic marks like DNA methylation (DNAm). In this study, we aimed to elucidate genetic factors contributing to disease susceptibility through placental DNA methylation. We performed the largest meta-analysis of placental methylation Quantitative Trait Loci (mQTLs) including more than 1,700 samples across seven cohorts from the Pregnancy and Childhood Epigenetics Consortium. In each cohort, placenta DNA methylation array and imputed genotype data were used to map mQTLs in *cis* (0.5 Mb window) using tensorQTL. The quality control of mQTLs was performed with EasyQC and the meta-analysis with GWAMA. We discovered more than 25 million associations between fetal genotypes and placental DNA methylation levels, including 3,531,259 SNPs and 210,476 CpGs. The vast majority of the SNP-CpG pairs in *cis*-mQTLs were located close to each other with a median distance of 250 bp, indicating that genetically modulated DNAm is typically close to the implicated variant. Furthermore, most of the mQTLs showed consistent effect direction across cohorts. DNAm sites in mQTLs were enriched in genomic features with intermediate methylation values, including CpG island shelves and shores, as well as open sea regions. Additionally, we observed a significant depletion from gene promoters and CpG islands. With the eFORGE software, we were able to detect enrichment in fetal placenta-specific DNase I hotspots and H3K4me1 broadPeaks, that mark accessible chromatin and enhancer regions, respectively. Finally, we performed Summary data-based Mendelian Randomization (SMR) analyses to study the genetic susceptibility acting through placental methylation in more than 40 GWAS traits including anthropometric measurements, and metabolic, cardiovascular, immune, digestive, respiratory, and neurologic diseases. Our results suggest that placental mQTLs may have a widespread impact in disease-associated loci with relevant implications in the prenatal origins of human complex disease.

Session Title: Molecular Effects of Genetic Variation Poster Session III

PB1134 Genome-wide microRNA QTL mapping and integration with GWAS links brain microRNAs to neuropsychiatric disorders

Authors:

S. Vattathil[1], E. Gerasimov[1], S. Canon[1], A. Lori[1], Y. Liu[1], E. C. Lai[2], D. A. Bennett[3], T. S. Wingo[1], A. P. Wingo[1,4]; [1]Emory Univ., Atlanta, GA, [2]Sloan Kettering Inst., New York, NY, [3]Rush Univ. Med. Ctr., Chicago, IL, [4]VA Atlanta Hlth.Care System, Atlanta, GA

MicroRNAs (miRNAs) are small non-coding miRNAs that function as post-transcriptional regulators. Given their broad influence on gene expression, characterizing the genetic regulation of miRNA variation can shed light on the role miRNAs play in the complex biological networks that link population genetic variation to brain traits and disease. To this end, we mapped brain miRNA eQTls (miR-QTLs) using genome-wide small RNA sequencing profiles from 604 donors of European ancestry. We detected miR-QTLs for 224 miRNAs (48% of 470 tested miRNAs) at false discovery rate < 1%, demonstrating that miRNAs are subject to extensive genetic regulation. The miR-QTLs were more likely to be upstream than downstream of the miRNA gene and were enriched in brain promoters. We then integrated the brain miR-QTLs with GWAS results from 16 psychiatric and neurodegenerative diseases using multiple independent approaches (miRNome-wide association analysis with FUSION, colocalization analysis with COLOC, and summary statistic-based Mendelian randomization with SMR). This approach prioritized five miRNAs (miR-92b-3p, miR-1908-5p, miR-499a-5p, miR-1255a, miR-190b-5p) that may mediate the effect of genetic variation on complex brain traits including bipolar disorder, major depression, neuroticism, post-traumatic stress disorder, schizophrenia, and Parkinson's disease. The results we report provide a basis not only for follow-up studies aimed at clarifying the role of brain miRNAs in relation to brain traits, but also for future analyses aimed more generally at understanding the regulation of miRNA expression.

Session Title: Molecular Effects of Genetic Variation Poster Session I

PB1135 Harnessing HT-SELEX and CAP-SELEX data for predictive modeling of transcription factor binding.

Authors:

R. Chung[1], G. Loeb[2], N. Ioannidis[1]; [1]Univ. of California, Berkeley, Berkeley, CA, [2]UCSF, San Francisco, CA

Despite being an in vitro assay, high throughput systemic evolution of ligands by exponential enrichment (HT-SELEX) remains one of the main experiments used for discovering transcription factor (TF) binding preferences in TF motif databases such as JASPAR. Previous works modeling DNA binding preferences of TFs using HT-SELEX datasets (e.g. DeepBind and DeepSELEX) have shown limited transferability to in vivo TF binding assays, such as ChIP-seq, for some transcription factors. The more recent CAP-SELEX assay identifies pairs of TFs that cooperatively bind to a DNA sequence.In order to understand the different types of information captured by HT-SELEX and CAP-SELEX data, we modeled both data sets jointly using a shared convolutional neural network (CNN). We found that a model trained on SOX2-OCT4 CAP-SELEX data outperformed a model trained on SOX2 HT-SELEX data at identifying SOX2 ChIP-nexus peaks (0.777 vs. 0.628 AUROC, respectively). Similarly, a model trained on E2F1-ELK1 CAP-SELEX data outperformed a model trained on E2F1 HT-SELEX data in predicting ChIP-seq peaks (AUROC 0.732 vs 0.533).These preliminary results suggest that the gap between in vitro and in vivo data is partially explained by TF cooperativity. Furthermore, a CNN that jointly models HT and CAP-SELEX datasets may learn shared DNA binding motifs between experiments. Our goal is to use this model, trained on all publicly available SELEX data, to augment existing genome interpretation models, like Basenji, which have only been trained on the human genome which lacks sequence diversity.

Session Title: Molecular Effects of Genetic Variation Poster Session II

PB1136 High-resolution 3D reconstruction of spatial conformations for the human genome decodes the context-specific mechanisms of long-range genetic associations.

Authors:

J. Wang, H. Wang, J. Yang, S. D. Adhikari; Michigan State Univ., East Lansing, MI

One of the fundamental challenges in disease genetics is to delineate the molecular mechanisms linking the disruptive effects of genetic variants to complex phenotypes, especially for non-coding SNPs that are located distal to target genes. While the spatial chromosomal conformations in 3D space have been demonstrated to play pivotal roles in modulating long-range genetic associations, current experimental data of chromatin contacts (*e.g.* Hi-C and Capture-C) are highly limited to a few cellular-contexts. The datasets have extremely high rates of missing data, making high-resolution characterization of detailed genome folding challenging. Excessive missing data of chromatin contacts also makes the interpretation of long-range genetic associations infeasible. Here **we develop a family of computational models, FLAMINGO and its variants, which are able to reconstruct 1kb-resolution spatial configurations of the human genome (*i.e.* the highest resolution to date) across diverse cell types and even at single-cell specific levels**. FLAMINGO consistently demonstrates superior accuracy, orders-of-magnitude boost in scalability, and strong robustness against high rates of missing contacts. Integrative analysis of reconstructed 3D genome structures with context-specific multi-omics data and genetic association studies (such as eQTLs, hQTLs and GWAS SNPs) revealed a series of novel discoveries, including the detection of super long-range QTLs (>900kb), single-cell specific genetic associations, spatial 3D hubs of orchestrated regulatory activities around SNPs, and improved prioritization of causal SNPs. **Beyond 1D genomic analysis and traditional association studies, our predictive and analytical framework opens up a new paradigm to decipher the molecular mechanisms of SNPs associated with complex human diseases from the systems-level view of high-resolution 3D genomes across diverse cellular contexts**.

Session Title: Molecular Effects of Genetic Variation Poster Session III

PB1137 High-throughput identification of deep intronic splice-disrupting variants in an undiagnosed rare disease cohort using a novel massively parallel minigene splicing assay

Authors:

S. Younger, A. L. Martinez-Bengochea, J. C. Means, T. Pastinen; Childrens Mercy Kansas City, Kansas City, MO

Improved DNA sequencing technologies and decreases in the cost of DNA sequencing have enabled the use of whole genome sequencing in a clinical setting. While the application of this technology has facilitated the identification of disease-associated genetic variants, the rate of diagnosis for rare disease remains below 50%. A major factor in these poor diagnostic rates is that most disease-associated genetic variants occur in noncoding regions of the genome with no prior annotation and no predictable functional consequences. Noncoding genetic variants within intronic sequences, which constitute ~35% of the human genome, can disrupt proper RNA splicing and lead to disease. However, the molecular consequences of intronic variants are difficult to predict and the potential contribution of these variants to disease etiology is commonly overlooked. Rapid, robust, and scalable experimental methods for deciphering the impact of intronic variants on RNA splicing and gene function would hold the promise of dramatically expanding the sequence space of the human genome that is amenable to clinical interpretation. To achieve this goal, we have developed a massively parallel minigene splicing assay that enables the empirical validation of the molecular consequences of deep intronic variants at scale. Our assay utilizes a GFP minigene composed of two exons separated by a short human intron that contains a central cloning site. Individual genomic sequences of interest can be cloned into the assay vector and, following introduction into cultured cells, splicing outcomes can be evaluated using RT-PCR. Alternatively, we have engineered a multiplexed barcoding system that facilitates the simultaneous assessment of thousands of genomic sequences in a pooled format using a high-throughput sequencing-based readout. Notably, this sequencing-based approach enables the characterization of splicing aberrations at single nucleotide resolution. To validate our assay, we leveraged the pediatric data repository generated by our institute's rare disease initiative, Genomic Answers for Kids, to curate a selection of >4500 rare deep intronic variants identified in undiagnosed rare disease patients. Our assay illuminated multiple variants that reproducibly disrupted RNA splicing across diverse cellular contexts. For a subset of variants, we generated patient-derived cellular models and performed RNA-seq to confirm the presence of splicing aberrations matching those observed in our assay. We anticipate that the assay described here will eventually contribute to diagnoses for many children that are impacted by rare disease and have long been searching for answers.

Session Title: Molecular Effects of Genetic Variation Poster Session I

PB1138 † Human embryoid bodies reveal dynamic genetic regulation in diverse differentiation trajectories

Authors:

J. Popp[1], K. Rhodes[2], K. Barr[2], R. Jangi[1], M. Li[1], B. Strober[3], Y. Gilad[2], A. J. Battle[1]; [1]Johns Hopkins Univ., Baltimore, MD, [2]Univ. of Chicago, Chicago, IL, [3]Harvard Sch. of Publ. Hlth., Brookline, MA

Large-scale efforts to identify genetic variants associated with nearby gene expression levels (cis expression quantitative trait loci, or cis-eQTLs) have focused primarily on healthy adult tissues. The scarcity of data from transient cell states such as those arising during differentiation presents a key bottleneck to the identification of target genes and mechanisms underlying disease-associated genetic loci, some of whose regulatory effects may present only during dynamic processes. In this work we examined genetic regulation of gene expression across diverse differentiating cell states with an efficient, unified experimental framework based on embryoid bodies. Embryoid bodies are three-dimensional aggregates derived from induced pluripotent stem cells (iPSCs) that spontaneously differentiate into a wide variety of cell types, including derivatives of all three germ layers. We generated embryoid bodies from 53 donors, and collected single cell RNA-sequencing data for over 900,000 cells. We called cell type-specific eQTLs across 29 cell types, and identified thousands of eQTLs displaying varying degrees of context specificity. While a majority of these loci overlap with eQTLs previously identified by GTEx, the eQTL genes without overlap were enriched for multiple developmental processes. We next interrogated context specificity using unsupervised machine learning, which revealed hundreds of variants with regulatory effects that vary across the differentiation landscape. The asynchronicity of differentiation across individual cells in this system enables us to reconstruct entire differentiation trajectories, from stem cells to developmentally relevant intermediate stages to differentiated cells. We therefore additionally called dynamic eQTLs along the differentiation trajectories from iPSCs to neurons, to cardiomyocytes, and to hepatoblasts, characterizing regulatory dynamics across all three germ layers in a single experimental framework. We intersected these newly discovered dynamic regulatory effects with known disease loci, and highlight an example dynamic eQTL along the neuronal trajectory which overlaps a previously identified schizophrenia risk locus. This work expands the identification of genetic regulatory effects to underexplored cell states arising during cellular differentiation for a wide array of cell types, accelerating the identification of regulatory mechanisms underlying disease risk loci.

Session Title: Molecular Effects of Genetic Variation Poster Session II

PB1139 Human Leukocyte Antigen-C and Haplotypes Associations with Resistance and Susceptibility to HIV-1 Infection among Serodiscordant Couples in Nigeria

Authors:

N. Otuonye[1], L. Luo, Dr[2], N. N. Odunukwe, Prof[1]; [1]Nigerian Inst. of Med. Res., Yaba, Nigeria, [2]Natl. Microbiol. Lab., Winnipeg, WA, Canada

Introduction The Human Leucocyte Antigen (HLA) class-1 is known to play a significant role in mediating resistance or susceptibility to HIV infection in the clinical course of AIDS. Recent studies have identified HLA-C as a key molecule that affects HIV disease progression. However, the role of HLA class 1 in heterosexual HIV-1 susceptibility or resistance in serodiscordant couples is not known in Nigeria. Therefore, this study evaluated the association between HLA-C susceptibility and resistance in HIV-1 transmission amongst heterosexual serodiscordant couples in Nigeria. **Methods** A total of 271 serodiscordant, concordant HIV positive and negative couples who gave informed consent were enrolled into this study. Extracted genomic DNA was sequenced for high resolution HLA-C class 1 genotypes using allele-specific primers (on exons 2 and 3) for HLA-C sequencing and typing. **Results** The highest frequency distribution of high-resolution HLA-C alleles observed in the HIV positive subjects were: HLAC*040101 178 (35.0%) followed by C*0701 124 (24.9%) compared with HIV negative subjects: C*040101 108 (39.0%) followed by C*0701 64 (24.7%). Alleles C*070201 (OR = 4.19, $P < 0.05$) and C*0804 (OR = 3, $P < 0.045$) were found to be independently associated with HIV-1 susceptibility in the cohort of serodiscordant couples. HLA-C*0802 (OR = 0.5. $P < 0.005$) and C*0304 (OR = 0.34. $P < 0.002$) were significantly associated with HIV-1 resistance to HIV-1 infection among the cohort. **Conclusion** The result has contributed to the importance of how host HLA-C genetic factors can influence HIV-1 disease susceptibility (HLA-C*070201; C*0804) and resistance (HLA-C*0802; C*0304) in serodiscordant couples. This information may contribute to the development of future effective HIV vaccine in Nigeria. **Keywords** Serodiscordant Couples, HLA-C Class 1 Resistance Allele, HLA-C Class I Susceptible Allele, HIV-1 Plasma Viral Load, and Nigeria

Session Title: Molecular Effects of Genetic Variation Poster Session III

PB1140 Hundreds of repeat polymorphisms influence splice site usage via diverse mechanisms.

Authors:

R. Mukamel[1], R. Handsaker[2], M. Sherman[3], A. Barton[4], M. Hujoel[5], S. McCarroll[1], P-R. Loh[6]; [1]Harvard Med. Sch., Boston, MA, [2]Broad Inst., Cambridge, MA, [3]Massachusetts Inst. of Technology, Cambridge, MA, [4]Harvard Univ., Cambridge, MA, [5]Brigham & Women s Hosp./Harvard Med. Sch., Boston, MA, [6]Brigham and Women s Hosp. / Harvard Med. Sch, Boston, MA

Gene expression and splicing are strongly regulated by nearby sequence variation, but studies of gene regulation have mostly been limited to SNP and indel variants. Here, we examined thousands of regions in the genome that vary in length due to variable numbers of tandem repeats (VNTRs) but are challenging to genotype from short-read sequencing. We analyzed sequencing depth-of-coverage together with SNP haplotypes to estimate the lengths of 9,561 autosomal VNTRs in 838 GTEx donors and test them for association with expression and splicing quantitative traits. Statistical fine-mapping analyses identified 702 repeat polymorphisms that appeared to influence expression or splicing of a nearby gene ($P<1 \times 10^{-10}$, FINEMAP PIP>0.5 in at least one tissue). These VNTRs were enriched near relevant genomic features: 25% (CI=21-29%) of expression-altering VNTRs overlapped an annotated promoter or enhancer (compared to 8% of analyzed VNTRs), while 18% (CI=14-22%) of splicing-altering VNTRs were within 100bp of an affected splice site (compared to <0.1% of all tested VNTRs). Moreover, VNTRs were 7-fold enriched among lead variants at quantitative trait loci, indicating that they contribute substantially more to gene regulation than SNPs on a per-variant basis. Associations in GTEx suggested transcriptional roles for a sizable fraction (18/58) of the VNTRs we recently linked to complex traits in UK Biobank (Mukamel et al.; medRxiv preprint), including a VNTR 1kb downstream of *GPIHBP1* (which encodes an HDL binding protein) that associated with increased HDL ($P=2.6 \times 10^{-41}$; PIP=0.99), apparently by influencing *GPIHBP1* expression ($P<1 \times 10^{-10}$, PIP>0.5 in 13 tissues).

Examining the genomic context of splice-altering VNTRs together with sequence-level information about splice junctions available in RNA-seq reads suggested that repeat polymorphisms influence splicing by a diverse set of mechanisms, with VNTRs apparently capable of modulating usage of both proximal and distal splice sites. For example, VNTR allele length in *UPF3A* and *TUBGCP2* appeared to modulate the size of arrays of within-repeat alternative splice sites. At *PLIN5*, the length of a 24bp-repeat 5bp upstream of a canonical splice acceptor associated with intron retention. The pyrimidine-rich composition of this VNTR (75% pyrimidines) suggested that VNTR alleles influence branch point recognition by modulating the distance between the branch point and the splice acceptor. Repeats at *CUL4A* and *CNN2*, by contrast, appeared to influence usage of distal splice sites (>800bp away). These results indicate a substantial and previously underappreciated role of VNTRs in regulating gene expression and splicing.

Session Title: Molecular Effects of Genetic Variation Poster Session I

PB1141 Hypothesis-free Mendelian Randomisation Identifies New Metabolites Linked to Risk of ALS

Authors:

E. Alhathli[1], T. Julian[2], C. Harvey[1], T. Moll[1], D. Wang[3,4], Project MinE sequencing consortium, P. Shaw[1], J. Cooper-Knock[5]; [1]Univ. of Sheffield, Sheffield, United Kingdom, [2]The Univ. of Manchester, Manchester, United Kingdom, [3]Univ. of Sheffield, Sheffield, South Yorkshire, United Kingdom, [4]Imperial Coll. London, London, United Kingdom, [5]Univ. Of Sheffield, Sheffield, United Kingdom

ALS, a rapidly developing and invariably fatal neurodegenerative disorder, is characterised by the progressive loss of motor neurons. ALS is an archetypal complex disease, with 10% of patients suffering from a monogenic disease; the vast majority of cases are sporadic, caused by a combination of genetic and environmental factors. With two-sample Mendelian randomisation (MR), causal inference can be made between various exposures and disease risks, such as serum concentrations of the entire set of metabolites. Summary statistics from genome-wide association studies (GWAS) of 575 metabolites were compared with those from a GWAS of amyotrophic lateral sclerosis (ALS) consisting of 29,612 ALS patients and 122,656 controls. Unbiased MR using the inverse variance weighted (IVW) estimate and weighted median causally associated five metabolites with risk for ALS after stringent Bonferroni multiple testing correction. Two hits are in the carnitine synthesis pathway - carnitine has been previously linked to the severity of ALS via a role in energy metabolism within motor neurons; this is the subject of an ongoing clinical trial (ALSUntangled). The remaining hits are being investigated with a combination of protein, metabolite and rare variant analyses. Follow-up studies will include evaluation within in vitro and in vivo models of ALS.

Session Title: Molecular Effects of Genetic Variation Poster Session II

PB1142 IBD-focused eQTL analysis identifies an increased rate of IBD GWAS colocalizations and novel target genes using diseased tissue compared to normal tissue eQTL

Authors:

N. Nishiyama[1], S. Silverstein[1], A. Bharadwaj[1], R. Ganesan[2], K. Clough[1], K. Darlington[1], G. Lian[1], C. Beasley[1], M. Schaner[1], S. Sheikh[1], T. Furey[3]; [1]UNC Chapel Hill, Chapel Hill, NC, [2]Carnegie Mellon Univ., Pittsburgh, PA, [3]Univ North Carolina, Chapel Hill, Chapel Hill, NC

While genome-wide association studies (GWAS) have identified over 200 loci associated with inflammatory bowel diseases (IBD), the functional mechanism and target gene of most loci remains unknown. Furthermore, eQTL mapped using any normal tissue such as the Genotype-Tissue Expression (GTEx) project only explain about 20% of GWAS loci, including IBD. Context-specific eQTL have identified regulatory mechanisms that may only be detected under specific conditions or stimuli and have provided additional explanation of GWAS loci. Therefore, we have performed the largest single-tissue diseased-focused eQTL analysis using colon tissue from IBD patients to increase our power to detect IBD GWAS colocalization and identify novel target genes. We hypothesize that eQTL mapped using diseased tissue will be enriched for disease-relevant molecular effects compared to normal tissue eQTL.

We performed paired genotyping and bulk RNA-sequencing on uninflamed colonic mucosa samples from 190 IBD patients (CD n=137; UC n=53). We mapped eQTL using IBD tissue for 30,173 genes using 6.7M genetic variant sites with a MAF > 0.05 +/- 1 Mb around the TSS of each gene. eQTL were compared to normal colon eQTL reported by GTEx. We then colocalized eQTL with 241 reported IBD GWAS loci to identify likely target genes using linkage disequilibrium (LD R2 > 0.4) and a Bayesian colocalization method, *coloc*. We compared eQTL-GWAS colocalizations between IBD and normal colon tissue to identify novel IBD GWAS colocalizations and target genes.

We have mapped 293,064 significant eQTL in diseased colonic tissue, associated with 2961 genes. Twenty-two lead eQTL signals using IBD tissue colocalized with 18 distinct IBD GWAS loci (*coloc* H4 Posterior Probability > 0.5). These colocalizations included well-known IBD GWAS candidate genes such as *ERAP2*, *FUT2* and multiple HLA MHC class II genes. eQTL mapped in IBD tissue were >2 times more likely to colocalize with IBD GWAS loci compared to eQTL mapped using normal tissue (Fisher's exact test p = 6.72x10-3), despite an IBD sample size one-half smaller (GTEx transverse colon n=368). Using diseased tissue, 60% of eQTL colocalizations are novel. Furthermore, we can detect distinct eQTL signals that colocalize with IBD GWAS loci for reported eGenes that do not colocalize using normal tissue eQTL. We report the first candidate target gene for the *CAMK2A* locus through a novel colocalization for *CAMK2A*. Additionally, our results implicate functionally relevant candidate target genes such as *TNFRSF14* for IBD GWAS loci that are missed by normal tissue eQTL, highlighting the advantages of using diseased tissue eQTL to prioritize trait-relevant candidate genes for GWAS loci.

Session Title: Molecular Effects of Genetic Variation Poster Session III

PB1143 Identification and characterization of the SNPs in the human *ZP2* gene and their role in female infertility.

Authors:

G. Gahlay, N. Rajput; Guru Nanak Dev Univ., Amritsar, India

ZP2 is an important component of the zona matrix which surrounds mammalian oocytes and facilitates fertilization. Recently, some studies have linked SNPs in genes encoding the zona matrix to female infertility. Single nucleotide polymorphisms (SNPs) are the most common type of genetic variations observed in a population. According to the dbSNP database, approximately 8,013 SNPs are reported in the human *ZP2* (*ZP2*) gene. In our previous research, we conducted an extensive *in silico* analysis of the SNPs in the *ZP2* gene, which had been reported in the dbSNP database, using six different computational tools, including SIFT and PolyPhen-2. This analysis predicted 18 non-synonymous SNPs (nsSNPs) as deleterious, with 12 of them likely to affect the structure or functional properties. These nsSNPs were predominantly located either in the N-terminal region crucial for sperm-zona interaction or within the zona domain which is responsible for the formation of zona matrix. To validate our *in silico* findings, we analyzed genomic DNA from female patients (n=15) with idiopathic infertility. The aim was to investigate the presence of these predicted nsSNPs in the *ZP2* gene, as well as other novel SNPs. Our analysis identified 8 nsSNPs, including 7 novel ones and 1 which has already been documented in the dbSNP database. Similar to the nsSNPs from our *in silico* studies, these mutations were also predominantly located in the N-terminal region or the zona domain. Four of these nsSNPs were predicted to be deleterious by *in silico* analysis and have structural/functional effects. To further investigate the impact of these changes, we have expressed (4 each from patient samples and *in silico* studies) the shortlisted deleterious SNPs and are currently conducting sperm binding assays to assess their effects on sperm interaction. This comprehensive approach will provide invaluable insights into the functional consequences of these SNPs and their implications on fertility. Understanding how these SNPs affect sperm-zona interaction and the zona domain can guide us to enhance the diagnosis of female infertility and/ improve the efficiency of in vitro fertilization (IVF). This, in turn, may help improve the success rates of fertility treatments, leading to higher pregnancy rates.

Session Title: Molecular Effects of Genetic Variation Poster Session I

PB1144 Identification of an elastin long-range *cis*-regulatory element.

Authors:

S. S. Procknow[1,2], K. M. Tsang[3], T. R. Luperchio[2], C. J. Billington[4], K. Small[5], A. E. Scheuerle[6], D. Tegay[7,8], P. Fernandez-Ferri[9], J. Meck[10], E. Totten[11], S. R. Braddock[11], K. Hornick[12], N. Redekar[12], B. Kozel[2]; [1]Washington Univ. in St. Louis, Saint Louis, MO, [2]Natl. Heart Lung and Blood Inst., NIH, Bethesda, MD, [3]NICHD, NIH, Bethesda, MD, [4]Univ. of Minnesota, Minneapolis, MN, [5]Northwell Hlth., Manhasset, NY, [6]Univ. of Texas, Southwestern Med. Ctr., Dallas, TX, [7]Northwell Hlth., Mount Sinai, NY, [8]IQVIA, Durham, NC, [9]GeneDX, Gaithersburg, MD, [10]Gene Dx, Gaithersburg, MD, [11]Saint Louis Univ., Saint Louis, MO, [12]Natl. Inst. of Allergy and Infectious Diseases, NIH, Rockville, MD

Supravalvar aortic stenosis (SVAS), a narrowing of the aorta just distal to the aortic valve, is often caused by elastin insufficiency. Elastin insufficiency is commonly seen in patients with Williams-Beuren Syndrome (WBS), a recurrent microdeletion syndrome spanning 26-28 genes including elastin (*ELN*), and in patients with loss-of-function single nucleotide variants and exon level deletions in the *ELN* gene. However, in some patients, and even families with SVAS, genetic evaluation fails to identify a causative variant. To investigate the genetic basis of SVAS in the absence of *ELN* variation, we identified a cohort of 4 individuals with SVAS and atypical 7q11.2 deletions that spare the *ELN* gene. The deletions share a region of approximately 200 kb that extends to 90 kb upstream of the *ELN* gene. Using bulk RNAseq (Illumina), we evaluated allele balance using known heterozygous single nucleotide polymorphisms (SNPs) to overcome differences in *ELN* expression with age and culture conditions. In contrast to fibroblasts from unaffected individuals which demonstrated biallelic *ELN* expression, a single predominant *ELN* allele was detected in affected fibroblasts, suggesting transcription off a single allele. Genomic DNA from patient fibroblasts was assessed by long-read sequencing (PacBio) which demonstrated that the suppressed allele was in *cis* with the upstream deletion. These data suggest the presence of a long-range *cis*-regulatory element (CRE) contained within the patients' deletions affecting *ELN* transcription. Using data available from ENCODE and GeneHancer, we identified key areas within the shared deletion region that have chromatin marks consistent with CREs including high H3K27ac and H3K4me1 in *ELN* expressing cell lines. We subsequently identified a fifth patient with SVAS and a smaller, novel ~47 kb deletion located within the shared deletion region upstream of *ELN*, containing one such putative CRE. This putative CRE was removed from IMR-90 cells using CRISPR technology. Further studies to characterize *ELN* expression in these cells and potential disruption of chromatin structure are underway. These findings collectively suggest the presence of a novel long-range CRE upstream of *ELN* that may identify the genetic etiology of as yet undiagnosed patients with SVAS. Further characterization of this region may identify potential therapeutic targets in patients with elastin-mediated disease. Moreover, this approach may be more broadly applicable to identifying the mechanism of disease in patients with suspected genetic disease in which current testing methods fail to identify causative variants.

Session Title: Molecular Effects of Genetic Variation Poster Session II

PB1145 Identification of bidirectional regulatory regions for *FADS1*and *FADS2*

Authors:

S. Yang[1], H. Zhang[2], K. Ye[2]; [1]Univ. OF GEORGIA, Athens, GA, [2]Univ. of Georgia, Athens, GA

Long-chain polyunsaturated fatty acids (LC-PUFAs) play critical roles in human health, and the *FADS1* and *FADS2* genes encode the two rate-limiting enzymes in the biosynthesis process of LC-PUFAs. These two genes are located adjacent to each other in a head-to-head orientation, suggesting a shared regulatory region. GTEx data revealed that in this region, two previously experimentally supported causal regulatory variants, rs174557 for *FADS1* and rs968567 for *FADS2*, have opposite effects on the two genes. However, the underlying mechanism is still unclear. We performed bidirectional reporter assays for the regions harboring each causal variant, but no bidirectional regulatory activity was observed. ChIP-seq data for HepG2 cells from ENCODE revealed that these two causal variants both lie in binding sites for transcription factor SP1 and SREBP1c. Therefore, we conducted CRISPRi to block each binding site and overexpressed SP1 and SREBP1c, respectively, to see if there is any competing regulation. We found that overexpression of SREBP1c could increase the expression of *FADS1* and *FADS2*, while SP1 only increased *FADS2 expression*. Meanwhile, when we blocked the binding site in the *FADS1* intron 1, both *FADS1* and *FADS2* decreased expression, and *SREBP1c* increased while *SP1* decreased. To further confirm how these two causal variants in the shared region regulate *FADS1* and *FADS2*, we generated HEK293 cells with homozygous single-nucleotide edits for both variants using prime editing. Similar single-nucleotide editing will be performed in HepG2 and their effects on the expression levels of *FADS1* and *FADS2* will be examined. Taken together, findings from this study will enhance our understanding of *FADS* expressions and facilitate the development of genome-informed LC-PUFAs supplementation for preventing associated complex diseases.

Session Title: Molecular Effects of Genetic Variation Poster Session III

PB1146 Identification of genetic variation associated with cell state regulatory networks in human induced pluripotent stem cells.

Authors:

T. Arthur[1], J. Nguyen[1], A. D'Antonio-Chronowska[1], H. Matsui[1], N. Silva[1], I. Joshua[1], iPSCORE Consortium, W. Greenwald[1], M. D'Antonio[1], M. Pera[2], K. Frazer[1]; [1]Univ. of California, San Diego, La Jolla, CA, [2]The Jackson Lab., Bar Harbor, ME

Human induced pluripotent stem cells (hiPSCs) exhibit a range of interconvertible pluripotent cell states that resemble different stages of early embryonic development. Most hiPSC lines are primarily comprised of cells in the primed state which has a diminished differentiation capacity relative to the smaller fraction of cells in the formative state. Unraveling the genetic and epigenetic processes governing transitions between these states is crucial for understanding early embryonic development and advancing hiPSC applications. Here, we analyzed RNA-seq and ATAC-seq data from 143 hiPSC lines using a machine learning algorithm, which resulted in the identification of 24 gene network modules (GNMs) and 20 regulatory network modules (RNMs). Comprehensive characterization revealed strong correlations between GNMs and RNMs, indicating that the RNMs underlie the coordinated co-expression of genes in the GNMs and allowing us to elucidate their pivotal roles in pluripotency and self-renewal. Remarkably, we discovered non-canonical epigenetic profiles within RNMs, unveiling novel regulatory networks active during formative and primed states. Genetic investigations further identified 2,234 regulatory variants that disrupted transcription factor binding, leading to altered co-accessibility of regulatory elements and increased stability of specific pluripotency states. Notably, we highlighted the significant impact of rs9350250, located within an *E2F3* intronic regulatory element, which completely disrupted co-accessibility and was associated with increased prevalence of the formative state of pluripotency. These findings reveal novel pluripotency and self-renewal regulatory mechanisms and provide a valuable resource for future stem cell research to generate genetic signatures that can potentially identify hiPSC lines that are likely to have improved differentiation outcomes.

Session Title: Molecular Effects of Genetic Variation Poster Session I

PB1147 Identification of multiple *cis*-regulatory enhancer variants underlying *NOS1AP* QT interval GWAS locus.

Authors:

K. Vickery[1], E. Sanchez[1], D. Lee[2], A. Kapoor[1]; [1]UTHlth.Houston, Houston, TX, [2]Harvard Med. Sch. & Boston Children s Hosp., Boston, MA

Electrocardiographic QT interval, an index of ventricular de- and re-polarization, is a clinically relevant heritable quantitative trait as prolongation or shortening of the QT interval is associated with increased risk of cardiac arrhythmias and sudden cardiac death. Genome wide association studies (GWAS) have identified the *NOS1AP* locus as a major regulator of the QT interval variation. Although, in our prior work we identified a single enhancer polymorphism that affects human cardiac *NOS1AP* expression and underlies trait association, causal variants underlying multiple independent GWAS signals at the *NOS1AP* locus remain unknown. Following a *cis*-regulatory mechanistic hypothesis, we performed an enhancer screen for all QT interval-associated *NOS1AP* locus variants that overlap cardiac open chromatin regions, a hallmark of *cis*-regulation. First, we identified all bi-allelic variants (*n*=788) observed in 1000 Genomes European ancestry populations within 500 kb of the sentinel hit at the NOS1AP locus that are common (minor allele frequency>1%) and in high linkage disequilibrium (r^2>0.9) with any of the genome-wide significant ($P<5\times10^{-8}$) GWAS hits. Next, we used variant overlap with human cardiac left ventricle DNase-seq regions (extended 1 kb on either side of peak summits) to filter 160 candidate variants. Following variant-centered PCR amplicon design, where 154 variants in 95 amplicons passed the primer design parameters (amplicon size range 414-785 bp; median 624 bp), and amplification from genotyped 1000 Genomes/HapMap reference DNA samples, both alleles of 138 variants along with flanking sequence in 84 amplicons were successfully cloned upstream of a minimal promoter-driven firefly luciferase gene in pGL4.23. Mouse cardiomyocyte HL1 cells, seeded in 24-well plates, were transfected with test constructs and *Renilla* luciferase vector (for transfection normalization) in quadruplicate and luciferase assays were performed 48 h later. Compared to the empty vector control, we identified five amplicons encompassing six variants with enhancer activity and with significant allelic difference (*t* test $P<0.05$), including the GWAS sentinel variant rs12143842. These six enhancer variants are being evaluated further, including assessing their combined impacts on *NOS1AP* cardiac expression using expression quantitative trait locus datasets, assessing *in vivo* activities using zebrafish transient enhancer assays, and assessing the dependency of enhancer activity on variant and flanking sequence using variant-centered small deletion (± 5 bases) constructs.

Session Title: Molecular Effects of Genetic Variation Poster Session II

PB1148 Identifying Cell and State Specific Genetic Effects on Gene expression through Comprehensive Analysis of Human Microglia Transcriptomics.

Authors:

B. Jang[1,2], A. Ravi[1], E. Brophy[1], M. Rosen[1], R. Kubler[1], K. BP[1], W. Cuddleston[1], H. Seo[1], M. Park[2,3], J. Humphrey[1], H-H. Won[2,4], T. Raj[1]; [1]Icahn Sch. of Med. at Mount Sinai, New York, NY, [2]Sungkyunkwan Univ., Seoul, Korea, Republic of, [3]Genome Inst. of Singapore, Agency for Sci., Technology and Res. (A*STAR), Singapore, Singapore, [4]Samsung Genome Inst., Samsung Med. Ctr., Seoul, Korea, Republic of

Background: Research into neurodegenerative diseases such as Alzheimer's disease (AD) and Parkinson's disease (PD) has made significant progress through genome-wide association studies (GWAS), identifying over a hundred of genetic loci associated with disease. Human microglia have emerged as a crucial cell type in understanding the underlying mechanisms of neurodegeneration. However, despite these findings, the causal variants, genes, and pathways specific to microglia underlying disease pathogenesis remain largely elusive, hindering our ability to develop a comprehensive mechanistic understanding of the disease.
Methods: To address this challenge, we conducted a large-scale meta-analysis of primary human microglia from published studies (Lopes, Kosoy, and Young) from 391 donors. Additionally, we harmonized microglia nuclei data from Fujita, Gabitto (unpublished) and Bryois (published) single-nucleus RNA-seq datasets, encompassing a total of about 180,000 microglia nuclei from 678 donors. We performed eQTL mapping and employed statistical fine-mapping, transcriptome-wide association study (TWAS), colocalization analysis and co-expression network analysis. We are performing state eQTL (st-QTL) and co-expression QTL (coQTL) analysis, which considers different microglial cell states at single-cell resolution.
Results: Our meta-analysis of human microglia datasets, incorporating both sorted and single-nuclei resolutions identified 2,542 cis-eQTLs at a false discovery rate (FDR) <0.05. Among these, 34 and 18 eQTLs colocalized with AD risk loci and PD (PP.H4 > 0.7), respectively, and many of these were also prioritized using a TWAS approach. Moreover, our analysis revealed that many AD/PD-associated eGenes are specifically expressed in specific microglia states including disease-associated microglia (DAM I and II), lipid metabolism, phagocytosis, and homeostatic, as evidenced by single-cell/nuclei microglia data.
Conclusion: Our study represents the most extensive collection of human microglia transcriptomic data to date, encompassing both bulk and single-cell resolutions. By identifying specific eQTLs, prioritizing functional variants and genes, and considering microglial cell states, we have taken a significant step forward in understanding the underlying mechanisms of AD and PD.

Session Title: Molecular Effects of Genetic Variation Poster Session III

PB1149 Identifying cell and tissue types where host genetic variation acts to impact gut bacterial abundance

Authors:

N. Huntley, E. Davenport; Pennsylvania State Univ., University Park, PA

Our microbiome influences our health and therefore characterizing how the microbiome is determined is an important step towards understanding disease treatments. Environmental factors such as diet, medication, and exercise influence the microbiome, but it is becoming increasingly clear that host genetics also plays a role in determining microbiome composition. Recently there has been a growth of genome-wide association studies (GWAS) to identify variants in the human genome that are associated with the relative abundances of common taxa found in feces. While we are starting to gain insight into the genomic variants associated with microbial composition, we lack an understanding of where in the body those variants may act to affect microbial abundance and function. Methods have been developed to identify 'pathogenic tissues' for diseases with and without known etiology by integrating GWAS results with gene expression data. Integrating GWAS with bulk-RNA sequencing has shown that there are differences in expression of genes across tissues that are not intuitively associated with disease, and single-cell RNA sequencing (scRNA-Seq) has taken this a step further by providing a more detailed picture of the variability of expression at the cellular level within tissues. These approaches have not been applied to microbiome-GWAS data, therefore we lack an understanding of the relevant host tissues and cell types in which host genetic variation acts to influence the abundances of microbial taxa in the gut. To address this gap, we intersected publicly available microbiome GWAS summary statistics from greater than 18,000 individuals with scRNA-Seq datasets from 16 different studies for a total of 1.3 million cells from 31 tissues collected from mouse and human subjects by applying single cell disease relevance score methodology to identify host cell types with enrichment of microbiome associated quantitative trait loci. This study is allowing us to identify which cell types host genetic variation acts to influence gut microbiome composition.

Session Title: Molecular Effects of Genetic Variation Poster Session I

PB1150 Identifying deleterious noncoding variation through gain and loss of CTCF binding site activity.

Authors:

C. Tubbs[1], M. Benton[2], E. McArthur[3], J. Capra[4], D. Ruderfer[1]; [1]Vanderbilt Univ., Nashville, TN, [2]Baylor Univ., Waco, TX, [3]Univ. of Washington, Seattle, WA, [4]Univ. of California San Francisco, San Francisco, CA

Background: CCCTC-binding factor (CTCF) regulates 3D genome organization by binding its DNA sequence motif at thousands of sites throughout the genome. Genetic variation in these CTCF Binding Sites (CBSs) have clinical impact; their disruption occurs at high frequency in cancers and has been linked to the etiology of rare disease. However, while many variants affecting CBSs exist, only a small number are characterized, and few approaches exist to systematically interpret them. Here, we introduce new methodology to functionally annotate single nucleotide variants (SNVs) in CBSs. We do this by explicitly connecting changes in predicted affinity for the CTCF sequence motif to the experimentally observed probability of DNA binding. **Methods:** For each CBS in the human genome, we related its predicted binding affinity to its observed DNA binding activity by calculating the likelihood its precision-weight-matrix (PWM) score will be detected in experimentally supported motifs from the UniBind database. We call this measure *CBS Activity*. For each SNV overlapping a CBS in gnomAD, we annotated its impact by taking the difference in activity scores (*ΔCBS Activity*) between the reference and alternate alleles. We then characterize the relationship between a variant's function (*ΔCBS Activity*), mutability adjusted proportion of singletons (MAPS), and the impact on gene expression in 773 post-mortem human brain samples from the CommonMind Consortium (CMC). **Results:** We calculated *ΔCBS Activity* for 2.7 million SNVs (gnomAD) overlapping 657,593 CBSs. *ΔCBS Activity* scores range from negative one (complete loss of binding activity) to one (complete gain of binding activity). We discover 2,402 high-impact variants (*ΔCBS Activity* > +/- 0.75). These variants are under stronger constraint (MAPS=0.09) than missense variants (MAPS=0.03) in gnomAD (MWU, $p < 1x10\text{-}6$). Next, we assessed each variant's impact on gene expression in the CMC by calculating the mean absolute standard deviation (MAD) of expression z-scores for carriers. Carriers of high-impact CBS variants associated with more aberrant patterns of expression among genes within 1 Mb compared with other CBS variants (high-impact MAD=0.36, other-CBS MAD=0.32, MWU, $p = 5.5x10\text{-}4$). **Conclusion:** Our approach identifies genetic variants that create or destroy functional CBSs genome-wide. These variants are more constrained than other functional classes (missense) and induce detectable changes in gene expression in the human brain. This work represents a critical step in understanding a previously uncharacterized class of variation and will enable improved prioritization of non-coding variants in future disease studies.

Session Title: Molecular Effects of Genetic Variation Poster Session II

PB1151 Identifying differential protein binding of Noncoding AMD risk-associated Single Nucleotide Polymorphisms

Authors:

E. Santi[1], J. Dunn[2], C. Berlinicke[1], H. Zhu[1], J. Qian[1], D. J. Zack[1]; [1]Johns Hopkins Univ. Sch. of Med., Baltimore, MD, [2]Johns Hopkins Univ., Baltimore, MD

Age-related Macular Degeneration (AMD) is a retinal neurodegenerative disorder that causes central vision loss in the elderly population throughout the world; currently there are 20 million known cases in the United States alone. Anti-VEGF treatments have proven to be effective in managing and slowing the progression of neovascular ("wet") AMD and have been in use for almost 20 years. On the other hand, for the far more prevalent atrophic ("dry") AMD, pegcetacoplan, a drug that targets the complement cascade, has only just been approved by the FDA and its wide-scale effectiveness is yet to be determined. A better understanding of the underlying genetics that contribute to AMD etiology could potentially lead to the design of complementary therapeutic options, and in fact data from genome-wide association studies (GWAS) demonstrated the connection between AMD and the complement system. The connection between the many other AMD risk-associated single nucleotide polymorphisms (SNP) to AMD, though, remains poorly understood. Determining the mechanism of action of AMD risk associated SNPs is especially difficult, as many of the identified SNPs fall within noncoding genomic regions. Methods normally used to investigate the role of non-coding disease-associated SNPs include expression quantitative trait loci (eQTL) studies as well as chromatin immunoprecipitation sequencing (ChIP-seq), but such methods have not been done in AMD-relevant cell types in addition to only accounting for a small number of human TFs in a limited number of cell types, respectively. Therefore, to determine the potential functional relevance of non-coding AMD associated SNPs, we are performing a proteome-wide analysis of disease-associated SNPs (PWAS) study. PWAS can survey the entire human TF repertoire by probing protein-chip arrays, containing up to 1700 human proteins and isoforms, with SNP-carrying fluorescently labelled DNA oligomers. By assaying risk SNP probes vs Non-risk SNP probes, we will determine whether there are any proteins in which there is differential binding. The rationale of this approach stems from the hypothesis that functional AMD-related DNA SNPs are likely to execute their function via allele-specific interactions with specific proteins. PWAS-identified differential protein to SNP-allele interactions will then be functionally assessed using in vitro retinal pigmented epithelia (RPE) and photoreceptor (PR) cells derived from human stem cell lines containing the two different alleles. Functional assays to be used include the Massively Parallel Reporter Assay (MPRA) for downstream transcriptional expression and ChIP-seq for binding verification.

Session Title: Molecular Effects of Genetic Variation Poster Session III

PB1152 Identifying key underlying regulatory networks and predicting targets of orphan box C/D *SNORD116* snoRNAs in Prader-Willi Syndrome.

Authors:

R. Gilmore, Y. Liu, G. Carmichael, J. Cotney; UConn Hlth., Farmington, CT

Prader-Willi Syndrome (PWS) is a neurodevelopmental disorder linked to loss of paternal expression of an imprinted region on chromosome 15 (15q11-q13). PWS is characterized by initial hypotonia and failure-to-thrive, followed by hyperphagia and obesity. Other symptoms include mild intellectual disability, short stature, and a range of behavioral and sleep problems. Any therapeutic intervention for this disorder is focused on ameliorating symptoms as there is no cure. Most PWS cases exhibit megabase-scale deletions encompassing the imprinted 15q11-q13 locus which harbors multiple protein coding genes including *MAGEL2*, *NDN*, and *SNRPN*, and various non-coding RNA species, including tandem arrays of small nucleolar RNAs, *SNORD116* and *SNORD115*. Recently, several PWS patients have been identified harboring a much smaller deletion that encompasses primarily the *SNORD116* gene array. These genetic findings suggest that this gene cluster may be a direct driver of PWS phenotypes. The SNORD116 gene cluster is composed of 30 copies of individual *SNORD116* box C/D snoRNAs. These snoRNAs are classified as 'orphans' because no known targets have been identified and their sequences show no significant complementarity to rRNA. It is crucial to identify the targets and functions of *SNORD116* snoRNAs because all reported PWS cases lack their expression. To address this gap in knowledge, we engineered two different deletions at the 15q11-13 locus modeling deletions similar to those identified in PWS patients. The deletions were performed in two distinct human embryonic stem cell lines to control for effects of genetic background. To study the effects of loss of these loci in a neuronal context, we utilized an inducible Neurogenin-2 (*NEUROG2*) expression system to enable quick and reproducible differentiation into neurons. Performing bulk RNA-sequencing on these resulting neurons allowed us to identify a novel list of approximately 40 genes consistently transcriptionally dysregulated in our PWS-like systems. Importantly, our results showed that it is critical to compare differential expression from isogenic pairs of multiple cell lines with distinct genetic backgrounds as this enabled us to eliminate a large number of spuriously affected genes. By employing the recently described box C/D snoRNA binding prediction tool snoGloBe, we found our list of dysregulated genes is significantly enriched for predicted *SNORD116* snoRNA targeting versus multiple levels of control analyses. Our results indicate a gene regulatory network likely controlled by *SNORD116* is perturbed in PWS patients.

Session Title: Molecular Effects of Genetic Variation Poster Session I

PB1153 Identifying non-coding causes of congenital heart defects: Abnormal RNA splicing with multiple novel isoforms as a mechanism for heterotaxy.

Authors:

J. Wells, M. Padua, A. Haaning, S. Ware; Indiana Univ. Sch. of Med., Indianapolis, IN

Heterotaxy is a disorder characterized by severe cardiovascular defects and abnormal left-right patterning in other thoracic and/or abdominal organs. Previous clinical and research-based genetic testing has mainly relied on functional evaluation of coding variants to identify causative heterotaxy variants. The potential for non-coding variants causing congenital heart defects (CHDs), including heterotaxy, has been largely neglected. Mutations in Zinc finger of the cerebellum 3 (*ZIC3*), a transcription factor with 5 zinc finger DNA binding domains, cause X-linked heterotaxy. Our lab identified an X-linked heterotaxy pedigree without a coding mutation in *ZIC3*. Whole genome sequencing in this family revealed a novel, intronic variant (*ZIC3* c.1224+3286A>G) predicted to alter RNA splicing. An *in vitro* minigene splicing assay confirmed this variant can act as a cryptic splice acceptor. We then used CRISPR/Cas9 to introduce the *ZIC3* c.1224+3286A>G variant into human embryonic stem cells (hESCs) and demonstrated it acts as a cryptic splice acceptor resulting in pseudoexon inclusion. Intriguingly, cDNA amplification and Sanger sequencing of *ZIC3* amplicons revealed that this variant triggers multiple abnormal splicing events, resulting in several novel exons and isoforms. Many of these isoforms lack a portion of a DNA binding domain and the nuclear localization signal. Short read mRNA sequencing confirmed the initial, abnormal splicing patterns detected, but also identified additional novel exons not seen in control hESCs. In total, our results suggest a minimum of six novel exons: three alternative splicing patterns of a canonical exon, one pseudoexon with the *ZIC3* c.1224+3286A>G variant acting directly as a splice acceptor, and two additional pseudoexons created by the activation of cryptic splice sites. Long read mRNA sequencing of *ZIC3* c.1224+3286A>G and wildtype hESCs is ongoing to determine the relative abundance of the full-length novel transcripts. In addition, functional evaluation of the most abundant novel transcripts is in progress using *in vitro* (subcellular localization) and *in vivo* (RNA rescue in *Xenopus laevis*) methods. These results are the first reported instance of pseudoexon inclusion associated with heterotaxy. Our results suggest the importance of non-coding variants in heterotaxy and the need for improved methods to identify and classify non-coding variation that may contribute to CHDs.

Session Title: Molecular Effects of Genetic Variation Poster Session II

PB1154 Immunolocalization and functional characterization of PFE1445C: a novel integrin alpha-like cell adhesion protein in *Plasmodium falciparum.*

Authors:

C. Donkeu[1,2], R. Keumoe[1,2], M. Biabi[3,2], B. Fogang[1,2], E. Essangui[3,2], S. Kemleu[4,2], C. Else Eboumbou[3,2], R. Grailhe[5], F. Njiokou[1], L. Ayong[2]; [1]Univ. of Yaounde 1, Yaounde, Cameroon, [2]Ctr. Pasteur Cameroon, Yaounde, Cameroon, [3]Univ. of Douala, Douala, Cameroon, [4]Univ. of Buea, Buea, Cameroon, [5]Technology development division, Inst. Pasteur of Korea, institute Pasteur of Korea, Cameroon

Background: The cytoadherence/sequestration process of Plasmodium parasites is organ selective and thought to be mediated by a combination of specific ligands and receptors that are located in trans on both the parasite or infected erythrocyte surface membrane and the associated host cells. Whereas the cytoadherence protein, PfEMP-1 had been identified as the most important ligand for mature asexual stage parasite cytoadherence to microvasculature endothelia, the protein ligands responsible for sexual stage sequestration in collagen-rich bone marrow tissues have remained unknown. Here, we describe the initial characterization of a novel Plasmodium falciparum integrin-like protein with potential to mediate both asexual and sexual parasite sequestration in collagen-rich tissue linings. **Methods:** *Plasmodium falciparum* integrin-like proteins were identified by structural classification of proteins (SCOP) hierarchical search of integrin superfamily protein domains in malaria parasites. Polyclonal antibodies (anti-PfINT) reactive to the Plasmodium integrin alpha-like protein was elicited by immunization of rabbit with synthetic peptides derived from the membrane-proximal extracellular domain of PFE1445C, and used in both Western immunoblot analysis and fluorescence microscopy examination of the protein in fixed cells. Protein function was further analyzed by investigating the effect of anti-PfINT on parasite growth *in vitro* as well as determining the association of anti-PfINT antibody carriage with protection against malaria. **Results:** Bioinformatics analysis of protein domains revealed a single protein in the Plasmodium falciparum 3D7 protein database with domain combinations consistent with those of classical integrin alpha proteins. No integrin beta tail domain proteins were observed using SCOP. Western blot analyses detected a single expected protein band at 82 kDa. This protein was predominantly expressed in Plasmodium trophozoites and schizonts where it appeared to localize to the developing merozoite surface. Anti-PfINT antibodies exhibited significant growth inhibitory effects against *P. falciparum* 3D7 culture, suggesting an effect on merozoite invasion *in vitro*. Anti-PfINT antibody levels were significantly high in sub-microscopically infected and asymptomatic endemic individuals compared to malaria patients, further indicating a role of anti-PfINT antibodies in natural protection against malaria. **Conclusion:** Taken together, our data suggest a role for PFE1445C in asexual development of *P. falciparum* parasites both in vitro and in vivo in humans.

Session Title: Molecular Effects of Genetic Variation Poster Session III

PB1155 Impact of environmental air composition on immune responses and cardiometabolic risks: insights from a Canadian cohort study

Authors:

M-J. Fave[1], E. Bader[2], E. Gbeha[2], V. Bruat[3], M. Agbessi[2], D. Soave[4], P. Awadalla[2]; [1]Ontario Inst. for Cancer Res., Montreal, QC, Canada, [2]Ontario Inst. for Cancer Res., Toronto, ON, Canada, [3]MaRS Ctr., 661 Univ. Avenue, suite 510, Toronto, ON, Canada, [4]Wilfrid Laurier Univ., Waterloo, ON, Canada

Climate change, in conjunction with various natural and anthropogenic health stressors, will greatly impact human health and disease. However, humans exhibit remarkable variability in their individual immune responses to environmental challenges. Cytokines, key regulators of inflammation and immune responses, act at the interface of physiology and environmental exposures and have been implicated in diseases such as asthma, cardiometabolic disorders, and cancer. The World Health Organization estimates that nearly the entire human population resides in areas with high levels of air pollutants. Episodes of harmful air composition can arise from specific atmospheric conditions, excessive motor vehicle usage, residential or industrial emissions, or wildfires.

In a study involving over 7000 participants from the Canadian Partnership for Tomorrow's Health (CanPath), we have demonstrated an association between serum cytokine protein profiles and harmful air composition. Elevated levels of cytokines, including IL1b, members of the VEGF cascade, and CCL5, involved in angiogenesis and macrophage recruitment, were observed in response to high levels of PM2.5. These findings align with our observations linking IL1b to lung adenocarcinoma following PM2.5 exposure in mice and humans. Aged individuals exposed to deleterious air composition exhibit increased levels of these pro-angiogenic proteins, potentially contributing to elevated health risks. Notably, after accounting for age and BMI, air pollution explains most of the remaining variation in cardiometabolic risk within our cohort.

By analyzing data from 1400 CanPath individuals using either single-cell RNA sequencing or bulk RNA sequencing, we have identified cell type-specific eQTLs in genes encoding cytokine levels associated with air quality. This indicates that individual genetic variation and cell type specificity can modulate health risks and responses to environmental challenges. Specifically, we found an eQTL for *CCL5* in monocytes that is specific to aged individuals, as well as eQTLs affecting *VEGF* levels in plasmablasts. Additionally, we have replicated the dysregulation of cytokines in response to air quality for several biomarkers, including *VEGFB* and *IL1b*, using bulk RNA sequencing across regions in Canada that experience varying air quality profiles.

Our work shed light on the complex interplay between environmental factors, individual genetic variation and function, and human health, emphasizing the need for targeted interventions to mitigate the adverse effects of air pollution.

Session Title: Molecular Effects of Genetic Variation Poster Session I

PB1156 Improved multi-ancestry fine-mapping of gene expression data identifies more cis-regulatory mechanisms

Authors:

N. Mancuso[1], Z. Lu[1], J. Jung[1], D. Yuan[1], S. Gazal[2]; [1]Univ. of Southern California, Los Angeles, CA, [2]USC, Los Angeles, CA

Expression quantitative trait (eQTL) analyses have identified numerous variants associated with gene expression levels, but distinguishing causal *cis*-eQTLs from nearby tagging variants is challenging due to linkage disequilibrium (LD). Multiple works have shown that multi-ancestry fine-mapping approaches improve fine-mapping precision by leveraging LD heterogeneity. However, existing methods are computationally expensive and assume that causal variant effect sizes are independent across ancestries, which fails to reflect shared genetic architectures.

To solve these limitations, we present the Sum of Single Shared Effects (SuShiE) model to fine-map eQTL data across multiple ancestries. We note that SuShiE 1) improves fine-mapping precision by leveraging LD heterogeneity under shared causal assumptions; 2) learns the prior correlation of effect sizes across ancestries; 3) estimates ancestry-specific effect sizes under this correlation structure, which boosts statistical power in downstream transcriptome-wide association studies (TWAS); 4) does so using a highly-scalable variational inference framework. Through simulations, SuShiE places higher Posterior Inclusion Probabilities (PIPs) on causal eQTLs, outputs smaller credible sets, and exhibits greater precision ($P<5E-9$) than existing fine-mapping approaches. In addition, SuShiE estimates more accurate ancestry-specific effect sizes ($P<1E-7$), thus increasing the power for TWAS ($P<1E-31$).

Next, we analyze expression levels measured in PBMCs from individuals of EUR, AFR, HIS, and EAS ancestries (N=1,194) in the TOPMed MESA study together with LCL expression levels from individuals of EUR and AFR ancestries (N=813) from the GENOA study. Compared to a simple baseline, SuShiE identified more putative causal cis-eQTLs (18,631/35,764 eGenes vs 10,570) while outputting smaller credible set sizes, higher PIPs and credible sets with PIPs > 0.9 more frequently ($P<0.05$). SuShiE estimates 93% of genes exhibit 1-3 shared cis-eQTLs and infers highly correlated posterior effect sizes on average (rho=0.74) across ancestries, thus reflecting shared eQTL architectures.

To validate our results, we perform an enrichment analysis using candidate *cis*-regulatory element (cCRE) annotations from snATAC-seq in PBMCs. Focusing on MESA, PIPs inferred by SuShiE are more enriched in relevant regulatory features ($P=2.1E-6$). For example, we find 1.36x, 1.17x, and 1.15x relative enrichment of PIPs in TSS, native-T, and cytotoxic-NK cCREs compared with existing approaches.

Overall, our approach sheds light on understanding the genetic architecture of gene expressions across multiple ancestries.

Session Title: Molecular Effects of Genetic Variation Poster Session II

PB1157 Improving metabolite QTL discovery using metabolite ratios

Authors:

S. D. Slack[1], A. Romero[1], L. Vanderlinden[1], S. S. Rich[2], M. Rewers[3], O. Fiehn[4], J. M. Norris[5], R. K. Johnson[1]; [1]Univ. of Colorado Anschutz Med. Campus, Aurora, CO, [2]Univ. of Virginia, Scottsville, VA, [3]Barbara Davis Ctr. for Diabetes, Aurora, CO, [4]UC Davis Genome Ctr., Davis, CA, [5]Colorado Sch. of Publ. Hlth., Aurora, CO

Characterizing the genetic control of molecular traits through identification of quantitative trait loci (QTLs) can provide insight into the biochemical processes underlying disease. Ratios between metabolites more directly reflect flux through chemical reactions than individual metabolites, potentially offering a clearer connection between the metabolome and genome. As proof-of-concept, we compared heritability and QTL loci discovered through metabolome-wide association of individual features to those discovered from pairwise ratios of metabolites in a multi-national prospective study of children at high-risk for type 1 diabetes. We combined genotypes from a customized exome array, imputed to the TOPMed reference panel, with untargeted metabolomics data from The Environmental Determinants of Diabetes in the Young study for 338 islet autoimmunity cases and 921 controls matched on age, sex, country, and family history. SNP-based heritability (H) was calculated for 2,505 metabolites and 1,241,053 ratios from all pairwise combinations of metabolites using Genome-wide Complex Trait Analysis constrained restricted maximum likelihood, adjusted for case-control status and genetic relatedness. Heritability was similar (P = 0.13) for significantly heritable ratios (mean H = 0.37, IQR: 0.25-0.44) and individual metabolites (mean H = 0.38, IQR: 0.25-0.46). P-gain was calculated as a quantitative measure of the gain in heritability significance for each ratio compared to its more significant metabolite. 4,064 ratios (0.3%, mean H = 0.87, IQR: 0.77-0.99) had a Bonferroni significant p-gain at alpha 0.05, corresponding to an average heritability increase of 0.22 (IQR: 0.14-0.31) for the ratio over the metabolite. Matrix eQTL software identified metabolite QTLs for the prioritized 4,064 ratios and corresponding 496 individual metabolites, adjusted for case-control status and genetic relatedness. After Bonferroni correction, we identified 45,243 metabolite-SNP pairs and 12,949 ratio-SNP pairs (P < 1.1x10-11) which were combined into 21,301 loci of 287 individual metabolites and 3,887 loci of 450 ratios. 561 (14.4%) ratio loci, involving 153 ratios, were uniquely discovered in ratio QTL analyses; over 25% of these controlled pairs of ether phosphocholine levels. Our results suggest that analysis of metabolite ratios may provide novel insights into the genetic control of metabolism missed by QTL analyses of individual metabolites. Analysis of ratios may offer a more direct biological connection to genetic control of the metabolome via enzymatic reactions while potentially improving power by decreasing technical noise.

Session Title: Molecular Effects of Genetic Variation Poster Session III

PB1158 Incidental/secondary ACMG and non-ACMG genetic findings in a multiethnic cohort of 16,713 pediatric patients.

Authors:

A. Saeidian[1], M. March[1], D. Watson[1], X. Wang[1], L. Youssefian[2], E. Bhandari[1], F. Aliazami[3], G. Kazemioula[4], A. Saeidian[5], H. Vahidnezhad[1], M. Harr[1], F. Mentch[1], M. Akbarzadeh[6], L. Tsoi[7], J. Gudjonsson[8], J. Glessner[1], H. Hakonarson[1]; [1]Children's Hosp. of Philadelphia, Philadelphia, PA, [2]UCLA, Los Angeles, CA, [3]Azad Univ., Tehran Med. Sci., Dept. of Genetics, Tehran, Iran, Islamic Republic of, [4]Univ. of Social Welfare and Rehabilitation Sci., Tehran, Iran, Islamic Republic of, [5]Iran Univ. of Med. Sci., Tehran, Iran, Islamic Republic of, [6]Shahid Beheshti Univ. of Med. Sci., Tehran, Iran, Islamic Republic of, [7]Univ. of Michigan, Ann Arbor, MI, [8]Univeristy of Michigan, Ann Arbor, MI

Purpose: Clinical next-generation sequencing can identify pathogenic sequence variants that are medically actionable for patients and their families but are not associated with the patient's primary diagnosis. These variants are incidental findings (IFs), secondary findings (SFs), or unsolicited findings (UFs). To date, there is no report of types and frequencies of SFs in a large pediatric cohort including a substantial percentage of African-American children. We sought to investigate the types, frequencies, and rates of SFs as well as the effects of SF disclosure on patients and families at the Center for Applied Genomics at The Children's Hospital of Philadelphia. **Methods:** We searched for pathogenic and likely pathogenic variants in disease-causing genes, specifically in 78 ACMG recommended and 93 non-ACMG genes that met the definition of SFs/IFs in patients of four pediatric cohorts. The 93 non-ACMG genes were compiled from literature reviews of additional proposed genes to the ACMG list. Variants were filtered by P/LP classification as per ACMG/AMP recommendations, CADD vs Mutation Significance Cutoff (MSC) scores, variant type, presence in disease-causing variant databases, and IGV visualization for manual variant quality inspection. We divided the patient's P/LP variants into four disease categories: metabolic, cancer, cardiovascular, and other. **Results:** We report a distinctive distribution of types and frequency of 1901 P/LP variants of SF/IFs in 16,713 subjects. There were 702 (4.2%) previously reported P/LP variants in 35 ACMG genes and 294 (1.8%) in 31 non-ACMG genes. In addition, 543 (3.2%) previously not reported P/LP variants in 56 ACMG genes and 543 (3.2%) in 46 non-ACMG genes. The most frequent variants among the ACMG and non-ACMG genes were TTR and G6PD, respectively. In our patients, we found 177 P/LP variants in 12 ACMG and 189 P/LP variants in six non-ACMG cancer-predisposing genes. Additionally, we found 442 P/LP variants in 12 ACMG and 41 P/LP variants in 10 non-ACMG cardiac disease-causing genes. Variants of potential medical importance were identified in 8% of participants. Challenges in reporting SFs included variants associated with adult-onset disease within a pediatric cohort, loss of an "open future," pre-symptomatic testing, variable expressivity, incomplete penetrance, and unknown family history of phenotype. **Conclusions:** Our data on SF findings in a large pediatric cohort provide several implications, such as medical actionability with pharmacogenetic implications, family planning, additional family genetic testing, and genetic counseling strategies to manage the impact of SF disclosure to patients and families.

Session Title: Molecular Effects of Genetic Variation Poster Session I

PB1159 Initial screening of large >10 Kb deletions from long-read sequencing of 1,000 Emiratis

Authors:

S. Elavalli[1], G. Katagi[1], L. Paulin[2], F. Sedlazeck[3], J. Quilez[1], T. Magalhaes[1], The Emirati Genome Program; [1]G42 Hlth.care, Abu Dhabi, United Arab Emirates, [2]Baylor Coll. of Med., Houston, TX, [3]Baylor Coll. Med., Houston, TX

The Emirati Genome Program (EGP) aims to whole-genome sequence at ~30X the ~1 million Emirati nationals in the United Arab Emirates (UAE). Two key features of EGP are the sequencing of tens of thousands of participants using long-read Oxford Nanopore Technologies (ONT), and the aim to translate genomics into clinical practice. Amongst the advantages of ONT is the capacity to detect structural variants (SV), which are important to better understand their contribution to biology and disease.

Here we present an initial characterization of SV in the Emirati population based on ~1,000 (1K) EGP participants whole-genome sequenced on ONT. We first assessed the quality of such dataset. We achieved satisfactory median read length (4.79 Kb), mapping coverage (41X) as well as percentage of mapped bases (94%) and genome covered (>99%). Using Sniffles2 and in line with previous studies, we called ~24,000 SV per genome irrespective of coverage. Besides, we further demonstrated the power of Sniffles2 to merge SV across all genomes and thus recall variants which had been missed in the individual calling (increasing to 24,500 SV per genome). Across all 1K individuals, we identified a total of unique 131,570 SV regions (46.2% deletions, 3.2% duplications, 49.1% insertions and 0.5% inversions) - this split between SV types is in line with previous studies.

We restricted our subsequent analysis to large (>10 Kb) deletion regions (LDR). We argue that (i) it is easier to predict the effect of deletions on the genes they overlap, compared to other SV types, and (ii) large events are more likely to have an impact on gene function. We detected ~45 LDR per genome (range = ~30-60) and a total of 2,300 LDR across the 1K genomes. We observed the expected Poisson-like minor allele frequency (MAF) distribution, with many singletons (n=868, 37.74%) and few common (MAF>5%) deletions (n=88, 3.83%). In between were events detected in >1 individual but with MAF<5%, which represented >58% (n=1,344) of the LDR. Of all LDR, 25% exceeded 25 Kb and ~200 were >100 Kb. As expected, relatively common LDR were <100 Kb and none >1 Mb reached a MAF>10%.

Finally, we screened our cohort of 1,000 Emirati genomes for homozygous common gene-overlapping LDR. We defined them as LDR with >1 individual homozygous for the alternate allele, with a MAF>5% and overlapping >=1 gene. We identified 32 of such regions each present in 0.10-71.90% of the cohort (median = 6.95%) and 5.10-44.70% MAF. On average, these regions were ~16 Kb (IQR = 10.30-31.40 Kb) and mostly comprised a single gene. We propose these regions as candidates for validation, further investigation and matching with clinical information.

Session Title: Molecular Effects of Genetic Variation Poster Session II

PB1160 Integrating genetics, metabolomics, and gene expression to dissect the sphingolipid pathway in Parkinson's disease.

Authors:

R. De Paula[1], C. Shaw[1], S. Elsea[1], J. Shulman[2,1]; [1]Baylor Coll. of Med., Houston, TX, [2]Duncan Neurological Res. Inst., Houston, TX

Parkinson's disease (PD) is a common, complex, and incurable condition that causes progressive motor disability and increased risk for cognitive impairment. This disease is neuropathologically defined by the formation of alpha-synuclein (aSyn) aggregates and loss of dopaminergic neurons in the substantia nigra. Accurate PD biomarkers remain to be developed to better stratify PD subtypes. Rare variants in genes associated with lysosomal storage disorders (LSD) have been linked to increased PD risk and may help to explain PD molecular heterogeneity, but metabolic evidence linking most LSD genes and PD is unavailable. Many LSD genes have key roles on the sphingolipid pathway, and mutations in those genes are potentially linked to disruption of sphingolipids. For example, *GBA* loss of function may cause the accumulation of glucosylceramide, disrupt lysosomal function, and affect aSyn accumulation. Importantly, aSyn aggregates seem to inhibit GBA enzyme, highlighting the existence of *GBA*-aSyn interactions. With that in mind, we hypothesize that the cumulative effect of LSD gene variants disrupts the sphingolipid pathway in PD via gene dysregulation and interactions with aSyn. We leveraged genome sequencing and metabolomic profiles from the Religious Orders Study and Rush Memory and Aging Project (n = 514 brain autopsies). We examined putative damaging, single nucleotide variants in 54 LSD genes (minor allele frequency (MAF) <3% and CADD scores >12.37). We used the likelihood-ratio test to compute dysregulated metabolites between subjects with and without subclinical PD pathology, and partition around medoids clustering to find multivariate patterns of sphingolipid dysregulation and their association with LSD gene variants. Our analysis revealed some significant dysregulated sphingolipids in PD brains, and clustering of subjects based on levels of 9 ceramide species revealed that *GBA*, *GALC*, and *SMPD1* variants are associated with higher ceramide levels. Consistent results were obtained in an analysis considering overall LSD genetic variant burden. Further integration with brain transcriptome and proteome data from the same human cohort reveals significant dysregulation of sphingolipid-related enzymes in PD. Similar result was observed in our lab using a *Drosophila melanogaster* model of aSyn pathology. Our results are potentially consistent with a model in which gene expression changes and metabolic perturbations mediate the impact of LSD genetic variants on aSyn pathology. Such findings increase our understanding of PD molecular mechanisms and may aid the development of more accurate PD risk scores in the future.

Session Title: Molecular Effects of Genetic Variation Poster Session III

PB1161 † Integration of cerebrospinal fluid pQTL and GWAS prioritizes novel candidate genes involved in Alzheimer's disease

Authors:

D. Western[1], J. Timsina[1], L. Wang[1], C. Wang[1], A. Ruiz[2], M. Marquié[2], M. Boada[2], I. de Rojas[2], I. Alvarez[3], M. Aguilar[3], P. Pastor[4], D. J. Pulford[5], Y. Sung[1], C. Cruchaga[1]; [1]Washington Univ. in St. Louis Sch. of Med., St. Louis, MO, [2]ACE Alzheimer Ctr. Barcelona, Barcelona, Spain, [3]Memory Disorders Unit, Univ. Hosp. Mutua Terrassa, Terrassa, Spain, [4]Univ. Hosp. Trias i Pujol and the Germans Trias i Pujol Res. Inst., Badalona, Spain, [5]GlaxoSmithKline, Stevenage, United Kingdom

Background: The integration of quantitative trait loci (QTLs) with disease genome-wide association studies (GWAS) has proven successful at prioritizing candidate genes at disease-associated loci. While tissue-specific genetic regulation of RNA transcripts is robustly analyzed, most studies assaying the genetic regulation of proteins have been limited to plasma. We have previously demonstrated that pQTL associations are highly tissue-specific. Given the relevance of cerebrospinal fluid (CSF) to Alzheimer's disease (AD), we developed a CSF-specific pQTL atlas and integrated it with AD GWAS to identify novel candidate genes for AD. **Methods:** We generated CSF proteomics (Somalogic, 7,584 analytes) data for 3,107 individuals. We integrated genomic and proteomic data through pQTL analysis using a three-stage approach: discovery, replication, and meta-analysis. We analyzed the tissue- and molecule-specificity of our associations by comparing to plasma & brain pQTLs and various eQTLs. We performed a proteome-wide association study (PWAS) and Mendelian randomization (MR) to prioritize proteins affecting AD. We determined shared protein-disease genetic regulation using colocalization. Using AD-associated proteins, we built prediction models for AD biomarker status. **Results:** Our meta-analysis identified 2,316 significant pQTLs (1,247 in *cis* and 1,069 in *trans*) for 1,960 proteins, of which 1,228 were not observed in plasma. Through PWAS, we identified 473 proteins associated with AD risk. MR prioritized 40 proteins as causal and colocalization identified 158 proteins that share genetic etiology. 42 of these overlap between at least two of these methods and are enriched in immune and lysosomal pathways. Multiple (including PILRA, PRSS8, and SIRPA) represent novel candidate proteins. A prediction model using the PWAS-identified proteins outperforms a PRS at predicting CSF amyloid/tau status across ages (AUC $\geq$ 0.972) and *APOE* genotypes (AUC $\geq$ 0.989). **Conclusions:** We have developed the largest CSF pQTL analysis to date and confirmed that CSF pQTLs are largely tissue-specific. Through a rigorous approach combining three methods, we have identified high-confidence proteins involved in AD that confirm previously reported candidate genes and prioritize new ones at GWAS loci. We developed highly accurate prediction models using prioritized proteins. Our findings offer insights into Alzheimer's disease biology that were missing when using plasma analyses, supporting the development of tissue-specific proteomics databases in neurologically-relevant tissues.

Session Title: Molecular Effects of Genetic Variation Poster Session I

PB1162 Integration of Genomics and Epigenetics Data IdentifyNovel Variants for Isolated Cleft Palate.

Authors:

A. Alade[1], P. Kumari[2], E. Wentworth[3], T. Busch[4], W. Awotoye[4], A. Oladayo[5], E. Aladenika[6], M. Olujitan[1], T. Naicker[7], V. Sule[8], P. Mossey[9], L. Gowans[10], M. Eshete[11], W. Adeyemo[12], E. Zeng[4], E. Van Otterloo[6], M. O'Rorke[1], A. Adeyemo[13], J. Murray[14], J. Cotney[15], R. Cornell[16], P. Romitti[17], A. Butali[18]; [1]Univ. of Iowa, Iowa, IA, [2]Univ. of Washington, Seattle, WA, [3]Univ. of Connecticut, Farmington, CT, [4]Iowa Inst. for Oral Hlth.Res., Iowa, IA, [5]Univ. of Iowa, Iowa city, IA, [6]Univ. of Iowa, Iowa City, IA, [7]Kwazulu Natal Univ., Durban, South Africa, [8]Uiowa, Coralville, IA, [9]Univ. of Dundee, Dundee, United Kingdom, [10]Kwame Nkrumah Univ. of Sci. and Technology, Kumasi, Ghana, [11]Addis Ababa Univ., Addis Ababa, Ethiopia, [12]Univ. of lagos, Lagos, Nigeria, [13]NIH, Bethesda, MD, [14]U of Iowa, Iowa City, IA, [15]UConn Hlth., Farmington, CT, [16]Univ of Washington, Seattle, WA, [17]Univ Iowa, Iowa City, IA, [18]Univ. of Iowa, Coralville, IA

Background Non-syndromic orofacial clefts (NSOFCs) are the most prevalent craniofacial birth defects, affecting approximately 1 in 1000 live births worldwide. Despite previous genome-wide association studies (GWASs) identifying risk loci for NSOFCs, these variants explain only a small portion (20-30%) of the estimated heritability, and their functional mechanisms remain largely unknown. Therefore, we meta-analyzed suggestive genome-wide signals and integrated epigenetics data to discover and functionally characterize novel loci for NSOFCs. **Methods** We genotyped 158 single nucleotide variations (SNVs) with suggestive associations (E-05 ≤ P > E-08) from the African NSOFCs GWAS in an independent cohort of 824 cases and 1038 controls from Nigeria, Ghana, and Ethiopia. This was followed by performing allelic association adjusting for sex and the top principal components (PCs). Subsequently, we meta-analyzed the results with the discovery GWAS results. For non-coding significant SNVs, we constructed a haplotype block around the SNVs and overlayed it on chromatin mark datasets from human craniofacial tissues to predict the impact of the SNV(s) on regulatory regions. Further, we performed a dual luciferase reporter assay to test the allele-specific effect of the predicted regulatory variant(s) on enhancer activity.
Results and Discussions We found significant associations for SNVs in the 8q24 region with nonsyndromic cleft lip/palate, as well as associations with the 9q22, chr10q24 [*OPALIN*], and 2q22 regions with non-syndromic cleft palate (NSCP). Chromatin mark datasets from developing human craniofacial tissues showed that the SNVs at 8q24 and 9q22 were in haplotype blocks within, annotated craniofacial enhancers predicted to regulate the expression of the *MYC* and *PTCH1* genes; respectively. Evidence from dual luciferase reporter assay in relevant cell lines further supported the regulatory function of the SNV at 9q22 on enhancer activity. The 8q24 region is a widely studied cleft candidate region, and deletion of this region has been reported to cause NSOFCs in mice. The *PTCH1* functions in the hedgehog signaling pathway, and variants in this gene cause nevoid basal carcinoma syndrome, a syndromic form of OFCs. Additionally, the *PTCH1* gene is expressed in mesenchyme and subclusters of ectoderm during facial development. **Conclusions** Our study highlights the utility of epigenetic data in unraveling the functional mechanism underlying GWAS associations. Further, our result suggests that non-coding variants in regulatory regions near the *PTCH1* gene increase the risk for NSCP.

Session Title: Molecular Effects of Genetic Variation Poster Session II

PB1163 Integration of multi-omic data to investigate the pathogenesis of early onset Parkinsonism.

Authors:

A. Gil Martinez[1], L. Wu[1], T. Georgiades[1], E. Stafford[1], K. Levine[2], Z-H. Fang[3], A. Martinez Carrasco[1], D. Gveric[4], N. Wood[1], C. Smith[1], A. King[5], S. Love[6], A. Robinson[7], L. Parkkinen[8], C. Morris[9], B. Traynor[10], S. Scholz[11], M. Nalls[10], C. Blauwendraat[11], A. Singleton[10], G. Halliday[12], C. Shepherd[13], S. Gentleman[4], T. Warner[1], Z. Jaunmuktane[1], J. Hardy[1], T. Lashley[1], R. Real[1], M. Ryten[14], H. Morris[1]; [1]Univ. Coll. London, London, United Kingdom, [2]Data Tecnica, Glen Echo, MD, [3]German Ctr. for Neurodegenerative Diseases (DZNE), Göttingen, Germany, [4]Imperial Coll. London, London, United Kingdom, [5]King's Coll. Hosp., London, United Kingdom, [6]Univ. of Bristol, Bristol, United Kingdom, [7]Univ. of Manchester, Manchester, United Kingdom, [8]Univ. of Oxford, Oxford, United Kingdom, [9]Newcastle Univ., Newcastle, United Kingdom, [10]Natl. Inst. on Aging, Bethesda, MD, [11]NIH, Bethesda, MD, [12]Univ. of Sydney, Sydney, Australia, [13]Univ. of New South Wales, Sydney, Australia, [14]UCL Great Ormond Street Inst. of Child Hlth., London, United Kingdom

Early-onset parkinsonism (EOP) can be monogenic. However, emerging studies indicate that rare and common variants, individually or jointly, together with known mutations, can influence age at onset, suggesting a more complex genetic aetiology of EOP. Here, we aimed to integrate clinical, pathological, genetic, and transcriptomic data from clinically diagnosed EOP cases to determine putative causal variants and genes as well as likely pathogenic biological pathways. We define early onset by motor age-at-onset below 50 years old following the recommendations from the International Parkinson and Movement Disorder Society Task Force. We evaluated 182 EOP cases with summary clinicopathological data, including 116 clinically diagnosed Parkinson's disease cases (PD), 8 postencephalitic parkinsonism cases (PEP), 11 progressive supranuclear palsy (PSP) cases, and 47 multiple system atrophy cases (MSA) from brain bank cases. We studied the genetic features using the Illumina Neurobooster array (NBA) and whole genome sequencing with 150 base-pair paired-end sequencing with 30x coverage generated in the Global Parkinson's Genetic Program (gp2.org). 33% of cases had mutations in defined Parkinsonism genes and further evaluation of this dataset is underway. We investigated the gene expression profiles of 56 early-onset PD cases compared to 269 late-onset PD cases and 242 controls with blood-derived RNA available samples at baseline from the Parkinson's Disease Biomarkers Program (PDBP). We accessed data from the AMP-PD portal and used the DESeq2 package to run differential gene expression analysis and clusterProfiler package for the functional enrichment analysis. We identified 216 up-regulated genes and 27 down-regulated when comparing EOPD cases to LOPD and 1088 genes up-regulated and 273 down-regulated when comparing EOPD cases against controls. In addition, we observed significantly enriched terms related to the regulation of RNA metabolic processes and the suppression of mitochondrial transport. These results suggest different gene expression profiles for early-onset PD cases. We are correlating the results of the gene expression data with the genetic analysis of the brain bank data, and further brain RNA analysis is underway. This work highlights an integration of clinicopathological data and multi-omic strategies in understanding the pathogenesis of EOP.

Session Title: Molecular Effects of Genetic Variation Poster Session III

PB1164 Integrative analysis of multi-omics data identifies *PSMB9* as a gene responsible for TNF inhibitor response in patients with rheumatoid arthritis.

Authors:

Y. Kim[1], S. Lee[2], J. Lee[2], H-H. Won[1]; [1]Samsung Advanced Inst. for Hlth.Sci. & Technology (SAIHST) , Sungkyunkwan Univ., Seoul, Korea, Republic of, [2]Dept. of Med., Samsung Med. Ctr., Sungkyunkwan Univ. Sch. of Med., Seoul, Korea, Republic of

Rheumatoid arthritis (RA) is an inflammatory disease characterized by joint inflammation and pain. Tumor necrosis factor (TNF) causes the inflammation related to RA. TNF inhibitors (TNFi) are used to treat RA patients by effectively blocking the activity of TNF. However, each RA patient can have different response to TNFi treatment. To detect genetic biomarkers of TNFi response and understand the underlying mechanism, we investigated expression quantitative trait loci (eQTLs) using the INTERVAL database, influencing to gene expression involved in TNFi response. We identified 31,873 *cis*-eQTLs for 29 differentially expressed genes (DEGs) in TNFi responders compared to non-responders. We tested associations of the 31,873 variants with TNFi response in 2,712 RA patients of European ancestry. We identified two variants, rs3793127 and rs1391373 that were significantly associated with TNFi response (p = 9.23E-05 and p = 0.000241) and with expression levels of *PSMB9* (p = 0.00131 and p = 3.74E-08). The up-regulation of *PSMB9* was observed in responders ($p < 0.05$, Mann and Whitney's test) after three months of TNFi treatment. Furthermore, single-cell RNA analysis showed that *PSMB9* was abundantly expressed in subsets of B cells and $CD8^+$ T cells, characterized by plasmablasts and $GZMK^+$, respectively. They might play a significant role in antibody response and cytokine activation. This study may help elucidate the molecular mechanism for the response to TNFi treatment.

Session Title: Molecular Effects of Genetic Variation Poster Session I

PB1165 Integrator complex and gene expression dysregulation in schizophrenia and intellectual disability.

Authors:

N. Jalkanen[1], L. Urpa[2], N. Räsänen[1], M. Kurki[2,3,4,5], E. Hämäläinen[2], O. Kuismin[6], M. Daly[2,7,4,3], A. Palotie[2,3,4,5], H. Kilpinen[1], O. Pietiläinen[1]; [1]Helsinki Inst. of Life Sci. HiLIFE, Helsinki, Finland, [2]Inst. for Molecular Med. Finland FIMM, Helsinki, Finland, [3]Program in Med. and Population Genetics, Broad Inst. of Harvard and MIT, Cambridge, MA, [4]Stanley Ctr. for Psychiatric Res., Broad Inst. of MIT and Harvard, Cambridge, MA, [5]Analytic and Translational Genetics Unit, Massachusetts Gen. Hosp., Boston, MA, [6]Dept. of Clinical Genetics, PEDEGO Res. Unit and Med. Res. Ctr. Oulu, Oulu Univ. Hosp. and Univ. of Oulu, Oulu, Finland, Oulu, Finland, [7]Analytic and Translational Genetics Unit, Dept. of Med., Massachusetts Gen. Hosp., Boston, MA

Genetic discoveries in schizophrenia (SZ) and other neurodevelopmental disorders have highlighted transcriptional regulation as one key point of convergence. Biological modeling of these genetic high-risk variants in human induced pluripotent stem cell (hiPSC) -derived neuronal models can help elucidate how transcriptional misregulation leads to impaired neuronal function. With whole exome sequencing, we have discovered deleterious variants in the transcriptional regulator *INTS6* gene in a family with 6 patients with intellectual disability (ID) and one with SZ from the Northern Finnish Intellectual Disability study (NFID). Intriguingly, in the Schizophrenia Exome Sequencing Meta-Analysis (SCHEMA) consortium, 3 additional patients and no controls were reported with deleterious variants of *INTS6* (OR=16.1, p=0.0092). *INTS6* encodes for a subunit of the integrator complex, which is involved in regulating transcription and processing of RNA polymerase II transcribed 3' small nuclear RNAs (snRNAs). Transcription initiation and attenuation are crucial for neuron function and development. We hypothesise that *INTS6* haploinsufficiency leads to reduced integrator complex function and therefore to alternative splicing and gene expression dysregulation in neurons. To study *INTS6*, we reprogrammed hiPSC lines from 6 cases, 6 non-carrier family members and 6 controls to discover functional consequences of the mutation in a patient-relevant genetic background. The variant carriers had reduced INTS6 peptide levels consistent with a loss of function effect. We will present results from transcriptional profiling of the differentiated neurons with RNA sequencing from cases and controls to illuminate the influence of *INTS6* in transcription, mRNA processing, and neuronal function. For validation of causality, we have generated CRISPR Cas9-edited isogenic rescue and knock-in neurons. Our study highlights the utility of patient-specific neuronal models in illuminating biological mechanisms associated with genetic risk variants in SZ and ID.

Session Title: Molecular Effects of Genetic Variation Poster Session II

PB1166 Interrogation of functional variants in five COPD GWAS loci by massively parallel reporter assays

Authors:

A. Zhou[1], L. Gong[1], F. Guo[1], D. Qiao[2], V. Malik[3], y. yu[1], C. Benway[4], J. Liu[1], R. Chase[1], M. Cho[5], E. Silverman[6]; [1]channing lab, Brigham and Women's Hosp., Boston, MA, [2]Brigham and Women s Hosp., Needham Heights, MA, [3]Brigham and Women s Hosp., Boston, MA, [4]Boston Children s Hosp., Boston, MA, [5]Brigham and Women s Hosp., Duxbury, MA, [6]Brigham & Women's Hosp., Boston, MA

Most human complex disease-associated genetic variants identified from genome-wide association studies (GWAS) are in non-coding regions of the genome, indicative of potential undiscovered gene regulatory function. Integrative and high-throughput methods have been developed to identify functional variants associated with disease susceptibility by regulating causal gene transcription. We applied massively parallel reporter assays (MPRA) to identify functional variants in five COPD GWAS loci in three lung-relevant cell types: epithelial cells (16HBE), fibroblasts (MRC5), and endothelial cells (HUVEC). Among 1120 fine-mapped SNPs in five COPD GWAS regions included in the MPRA screening, we identified 25 SNPs with potential allele-specific effects on transcriptional activity. Four of these variants were confirmed by luminescence reporter assays in 16HBE and/or MRC5 cells demonstrating allele-specific enhancer activity. We further performed CRISPRi-based genome editing with targeted deletion of the genomic regions spanning potential functional variants in human bronchial epithelial cells (16HBE), followed by RNA-sequencing and qPCR validation. Two genes (*RUVBL1* and *RAB7A*) related to the WNT pathway, known to be important for lung repair regeneration and regulated by the previously identified COPD GWAS gene *FAM13A*, were identified as plausibletarget genes of potential functional COPD GWAS SNPs. In summary, through unbiased screening with MPRA in three cell types in combination with confirmatory reporter assays and CRISPR-based genome editing, we not only identified and validated top functional genetic variants associated with COPD but also pinpointed key target genes possibly implicated in COPD genetic susceptibility for future in depth mechanistic studies.

Session Title: Molecular Effects of Genetic Variation Poster Session III

PB1167 Investigating ancestry-dependent cis-regulatory elements in *TOMM40* that modulate *APOE*-ε4 expression in Alzheimer's disease patient-derived microglia

Authors:

B. DeRosa[1], A. Ramirez[1], K. Nuytemans[1,2], L. Coombs[1], O. Oron[1], M. Vasquez[1], S. Simon[1], D. Van Booven[1], F. Rajabli[1,2], K. Celis[1], P. Whitehead[1], L. D. Adams[1], T. D. Starks[3], P. Mena[1], S. Tejada[1], G. Byrd[3], A. Griswold[1,2], L. Wang[1,2], J. Young[1,2], M. Pericak-Vance[1,2], D. Dykxhoorn[1,2], J. Vance[1,2]; [1]John P. Hussman Inst. for Human Genomics, Univ. of Miami Miller Sch. of Med., Miami, FL, [2]Dr. John T. Macdonald Fndn. Dept. of Human Genetics, Univ. of Miami Miller Sch. of Med., Miami, FL, [3]Maya Angelou Ctr. for Hlth.Equity, Wake Forest Univ., Winston-Salem, NC

The strongest genetic risk factor for developing late onset Alzheimer's disease (AD) and sporadic early onset AD is the ε4 variant of the Apolipoprotein E gene (*APOE*). However, the risk conveyed by carrying *APOE*-ε4 varies between ancestral populations, with those of African ancestry showing lower ε4 risk compared to those of European ancestry. Furthermore, recent studies have shown that the variability in *APOE*-ε4 risk in admixed populations can be attributed to ancestry-dependent genetic variation in the region surrounding *APOE* (local ancestry; LA). *TOMM40*, a gene found immediately upstream of *APOE*, contains two intronic single nucleotide polymorphisms (SNPs), namely rs2075650 and rs59007384, with significant frequency differences between the African and the European *APOE*-ε4 haplotypes. Using massively parallel reporter assays (MPRAs) and chromatin structure analysis (Capture-C analyses), we previously observed higher enhancer activity of the intronic regions of *TOMM40* containing rs2075650 and rs59007384 in the European haplotype. To investigate the regulatory role of rs2075650 and rs59007384 in mediating ancestry-dependent *APOE*-ε4 expression, we used CRISPR/Cas9 to delete a 500 bp intronic region surrounding each SNP in induced pluripotent stem cell (iPSC) lines derived from African and European AD patients with an ε4/ε4 genotype. The isogenic sets of African and European iPSC lines (unedited and with deleted intronic regions) were validated for pluripotency, genomic stability, and the lack of off-target genome-editing resulting from the CRISPR-based gene correction. Ongoing studies in microglia generated from these lines will broaden our understanding of the contribution of these *TOMM40* regions in regulating the expression of *APOE*-ε4 in the context of African and European ancestral genetic backgrounds. The generation of isogenic sets of African and European iPSC lines with rs2075650 and rs59007384 reference and alternative alleles will help to comprehensively examine whether these *TOMM40* SNPs are the driving factors for the observed difference in ancestry-dependent enhancer activity.

Session Title: Molecular Effects of Genetic Variation Poster Session I

PB1168 Investigating the interplay between variant pathogenicity and frequencies across populations

Authors:

R. Estanboulieh, A. K. Manrai; Harvard Med. Sch., Boston, MA

In order to provide clinical-grade variant pathogenicity classifications, genetic testing laboratories employ guidelines that combine several criteria and heuristics including knowledge about implicated molecular pathways, protein biochemistry, case-control cohorts, and population allele frequencies. Current guidelines generally classify variants with minor allele frequency (MAF) greater than 5% in any of the 5 continental populations as benign, but there have been limited systematic investigations of how to precisely tune this threshold to balance inclusion of pathogenic and exclusion of benign variants. Previous work has proposed frequency bounds for dominant and recessive Mendelian disorders that account for estimates of variant penetrance, disease prevalence, and other factors. If the Poisson-based filtering allele frequency of a variant exceeds that bound, it is automatically classified as likely benign. In this study, we investigate multiple definitions of a variant's filtering allele frequency on pathogenicity prediction, as well as their robustness to smaller data from understudied or admixed populations. These definitions include Poisson-based estimates in addition to other estimates that aggregate frequencies across populations motivated by information theory. We explore the performance of this method on a wide range of simple and complex genetic diseases, using high-confidence ClinVar variants as a gold standard, and validate our findings against individual-level genetic and clinical data from the UK Biobank. With the latter, we also study whether it is useful to account for the ancestry of an individual, such as in the case of context- or ancestry-dependent pathogenicity. Our analysis reveals that frequency bounding, while capturing most pathogenic variants, can still be sensitive to frequency estimates from understudied populations. In summary, this work contributes to the practical challenge of classifying variants by furthering our understanding of the interplay between a variant's pathogenicity and its frequency across different populations.

Session Title: Molecular Effects of Genetic Variation Poster Session II

PB1169 IPSC-derived cerebral organoids as a neurological model of pediatric DPD deficiency.

Authors:

B. Bembenek, K. J. Bouchonville, R. E. Schmidt, S. M. Offer; Mayo Clinic, Rochester, MN

Dihydropyrimidine dehydrogenase (DPD) deficiency is caused by inactivating genetic variations in the *DPYD* gene, resulting in a metabolic disorder that has been correlated with neurologic and developmental conditions that arise during childhood. Case studies have reported associations between DPD deficiency and neurological symptoms, including epileptic seizures, microcephaly, delayed motor skill development, and autism spectrum-associated behaviors. The goal of this study was to characterize the cellular and molecular changes that are caused by DPD deficiency using patient- and control-derived cerebral organoids that were differentiated from human induced pluripotent stem cells (iPSC). Patient iPSC lines were obtained from affected individuals with three different genotypes associated with full or partial DPD deficiency; control lines were from unaffected individuals. Differential gene expression analysis with DESeq2 was performed comparing iPSCs, embryoid bodies, and mature organoids by patient phenotype and genotype. Overall, dysregulation of genes related to synaptic function, neurotransmitter signaling, cell-type markers, and neurological structural components was observed at all stages of development between cases and controls. To further investigate dysregulation at a single cell level, single cell RNA sequencing was performed on mature organoids, and immunofluorescence and functional analyses were performed to mechanistically investigate observed gene expression changes. Based on our results, we hypothesize that the upregulation of glutamate receptors results from a reduction of β-alanine availability via DPD deficiency, leading to reduced modulation of glutamatergic signaling, which provides a potential explanation for the epileptic-type seizures that have been reported in patients. This excitotoxic signaling is also hypothesized to inhibit synaptic development and promote demyelination through excessive superoxide production, which may be associated with delayed neurological development. Collectively, our data provide key insight into the etiology of this condition.

Session Title: Molecular Effects of Genetic Variation Poster Session III

PB1170 Iterative improvement of deep learning models using synthetic regulatory genomics

Authors:

A. Ribeiro dos Santos, G. Ellis, H. J. Ashe, R. Ordoñez, M. T. Maurano; NYU Sch. of Med., New York, NY

Generative deep learning models can accurately reconstruct genome-wide epigenetic tracks from the reference genome sequence alone. But it is unclear what predictive power they have on sequence diverging from the reference, such as disease- and trait-associated variants. We previously characterized dozens of deletions, inversions, and rearrangements of DNase I hypersensitive sites (DHSs) at the Sox2 Locus Control Region (LCR) in mouse embryonic stem cells (mESC). Here, we use the state-of-the-art model Enformer to predict DNA accessibility across these engineered sequences when delivered in place of the Sox2 LCR. Overall, we observe good correlation between accessibility predicted by Enformer and experimentally measured Sox2 expression. But model performance was better for deletions or inversions spanning multiple DHSs than for smaller deletions of individual transcription factor recognition sequences. Predictive power was particularly poor for rearrangements affecting DHS order. To explore model performance on sequences diverging even further from the reference, we designed and delivered composite enhancer payloads containing Sox2 DHSs paired with DHSs from other mESC super-enhancers. We found that their activity in vivo significantly diverged from predictions in silico. Our results suggest that current deep learning models perform poorly when presented with novel sequence diverging in certain critical features from their training set. This suggests an iterative approach incorporating profiling of synthetic constructs can improve model generalizability, and ultimately enable functional classification of regulatory variants identified by population studies.

Session Title: Molecular Effects of Genetic Variation Poster Session I

PB1171 Large single cell level analysis of subcutaneous adipose tissue in Mexicans discovers distinct adipocyte sub-cell-types and genes associated with sex, Native American ancestry, obesity, and ancestry-obesity interactions.

Authors:

A. Kar[1], S. H. T. Lee[1], M. Alvarez[1], S. Das[1], L. L. Muñoz-Hernandez[2,3], M. Herrera-Hernandez[2,3], I. Cruz Bautista[2,3], B. Balliu[4,5,6], C. Aguilar-Salinas[2,3], M. Tusié-Luna[7,8], P. Pajukanta[1,9,10]; [1]Dept. of Human Genetics, David Geffen Sch. of Med. at UCLA, Los Angeles, CA, [2]Unidad de Investigación de Enfermedades Metabólicas, Inst. Natl. de Ciencias Médicas y Nutrición Salvador Zubirán, Mexico City, Mexico, [3]Departamento de Endocrinología y Metabolismo, Inst. Natl. de Ciencias Médicas y Nutrición Salvador Zubirán, Mexico City, Mexico, [4]Dept. of Pathology and Lab. Med., David Geffen Sch. of Med. at UCLA, Los Angeles, CA, [5]Dept. of Biostatistics, UCLA Fielding Sch. of Publ. Hlth., Los Angeles, CA, [6]Dept. of Computational Med., David Geffen Sch. of Med. at UCLA, Los Angeles, CA, [7]Unidad de Biología Molecular y Med. Genómica, Inst. Natl. de Ciencias Médicas y Nutrición Salvador Zubirán, Mexico City, Mexico, [8]Departamento de Med. Genómica y Toxiología, Ambiental Inst. de Investigaciones Biomédicas, UNAM, Mexico City, Mexico, [9]Bioinformatics InterDept.al Program, UCLA, Los Angeles, CA, [10]Inst. for Precision Hlth.at UCLA, David Geffen Sch. of Med. at UCLA, Los Angeles, CA

Mexicans, a population characterized by genetic admixture, are critically underrepresented in omics studies of obesity, despite their increased susceptibility and prevalence of obesity and cardiometabolic disease. To address this health disparity and bridge the knowledge gap, we profiled subcutaneous adipose tissue, a key fat depot, at the single cell level in 48 Mexicans (17% male). Our cohort comprised individuals with varying body mass index (BMI) (33% classified as overweight and 37.5% as obese), with a high proportion of Admixed American (AMR) ancestry (mean=67.1%), which serves as a proxy of the Native American ancestry. We first performed adipose single-nucleus RNA sequencing and careful quality control to create the largest to date Mexican single-cell subcutaneous adipose reference with 126,123 nuclei profiled. Focusing on adipocytes, a central adipose cell-type, we used Milo to identify 1,961, 84, 235, and 231 subcellular neighborhoods of transcriptionally similar cells that show differential abundance (DA) and 654 (225), 109 (10), 1,580 (54), and 378 (9) genes (adipocyte markers) in these neighborhoods that are differentially expressed (DE) by sex, AMR ancestry, BMI, and AMR ancestry-by-BMI interaction (FDR<5%), respectively, thus discovering functional subtypes within Mexican adipocytes. Intriguingly, in the neighborhoods with lower global AMR ancestry, we observed higher (p=0.036) average *cis*-regional (±500kb) local AMR ancestry in the downregulated vs upregulated adipocyte marker genes, thus showing influences of local ancestry in these DE genes. To test for genetic regulation of the DE and adipocyte expressed genes, we performed a *cis*-eQTL mapping in the adipocyte pseudobulk expression data, adjusting for ancestry. At FDR<5% and MAF>10%, we identified 35,617 eQTL SNPs, regulating 830 genes, including 4.05%, 3.44%, 5.81%, and 2.75% of the DE genes in AMR ancestry, AMR-by-BMI interacting, sex, and BMI-associated neighborhoods. We also found significant eQTL SNPs for 67 (8.7%) adipocyte marker genes, including key mitochondria and adipose function genes, *ETFDH*, *MGST1*, and *AQP7*. Of these *cis*-eQTLs, 24.2% were not reported in bulk adipose tissue in GTEx. Collectively, our study represents the first endeavor to examine Mexican subcutaneous adipose tissue at the single cell level, revealing distinct subpopulations of adipocytes influenced by sex, AMR ancestry, BMI, and interactions between AMR ancestry and BMI. Furthermore, we discovered *cis*-eQTLs regulating the DE genes at the cell-type level. Taken together, our study suggests that Native American ancestry transcriptionally influences the high obesity rates in Mexicans.

Session Title: Molecular Effects of Genetic Variation Poster Session II

PB1172 Large-scale testing of AGXT missense variant effects.

Authors:

A. De Souza[1,2], N. Kishore[3,2], Z. Dai[1,2], J. Weile[4,2], W. van Loggerenberg[3,2], R. Li[2], A. Rayhan[2], C. L. Tucker[5], M. Gebbia[3,2], A. G. Cote[3,2], F. P. Roth[3,2]; [1]Univ. of Toronto, Toronto, ON, Canada, [2]Lunenfeld-Tanenbaum Res. Inst., Sinai Hlth., Toronto, ON, Canada, [3]Donnelly Ctr., Univ. of Toronto, Toronto, ON, Canada, [4]Donnelly Ctr., Univ. of Toronto, Ontario, Canada, Toronto, ON, Canada, [5]Dept. of Pharmacology, Univ. of Colorado Sch. of Med., Aurora, CO

Variants in the AGXT gene, encoding the liver enzyme alanine:glyoxylate aminotransferase (AGT), can cause kidney stones and other symptoms of Primary Hyperoxaluria Type 1 (PH1). Over 27% of clinically-observed AGXT missense variants are deemed variants of uncertain significance. Saccharomyces cerevisiae lacking key genes in glycine metabolism (SHM1, SHM2, GLY1, and AGX1) cannot grow in ethanol media without human AGXT or supplemented glycine. First, we validated this functional assay for assessing variants in the context of both the reference allele and the common minor allele Pro11Leu and Ile340Met (global minor allele frequency 10%), which sensitizes AGT to the effects of other mutations. Tests of variants in both the major and minor alleles backgrounds are useful in exploring the effects of genetic interactions on protein fitness. Next, we attempted to measure the functional impact of all possible amino acid substitutions in AGT using large-scale mutagenesis, experimental evolution of a lab population of yeast cells expressing these variants, and determination of variant frequencies before and after selection. This yielded functional scores for 6758 (86%) of 7820 possible amino acid changes, collectively representing a missense variant effect map for AGT. The map recapitulates known biochemical features of AGT and successfully separates pathogenic from benign variants. Thus, our map for AGXT/AGT represents a resource for discovering new sequence/structure/function relationships and potentially expanding the fraction of kidney stone patients who can obtain a definitive genetic diagnosis and thereby benefit from already-approved PH1 therapies.

Session Title: Molecular Effects of Genetic Variation Poster Session III

PB1173 Large-scale trans-eQTL meta-analysis in lymphoblastoid cell lines reveals functional consequences of GWAS variants.

Authors:

K. Freimann[1,2], A. Brümmer[3,4], E. Holzinger[5], J. Maranville[6], N. Nakic[7], H. Ongen[7], L. Stefanucci[8], M. Turchin[6], G. Trynka[9], S. Bergmann[3,4,10], K. Alasoo[1,2]; [1]Univ. of Tartu, Tartu, Estonia, [2]Open Targets, South Building, Wellcome Genome Campus, Cambridge, United Kingdom, [3]Dept. of Computational Biology, Univ. of Lausanne, Lausanne, Switzerland, [4]Swiss Inst. of Bioinformatics, Lausanne, Switzerland, [5]BMS, Cambridge, MA, [6]Bristol Myers Squibb, Cambridge, MA, [7]GSK, Brentford, United Kingdom, [8]Wellcome Trust Sanger Inst., Hinxton, United Kingdom, [9]Wellcome Sanger Inst., Hinxton, United Kingdom, [10]Dept. of Integrative BioMed. Sci., Univ. of Cape Town, Cape Town, South Africa

Genome-wide association studies (GWAS) have provided valuable insights into the genetic basis of complex traits and diseases. However, the interpretation of GWAS findings can be challenging, particularly when the functions of associated genes are unknown. To unravel the molecular and regulatory consequences of GWAS variants, trans-eQTL studies have emerged as a promising approach. Unfortunately, most existing eQTL studies lack the statistical power to detect robust associations. In this study, we conducted a large-scale trans-eQTL meta-analysis using data from lymphoblastoid cell lines (LCLs) from over 3700 samples collected across nine cohorts. Our study is the largest of its kind in a single homogenous human cell type. This helps us to avoid the confounding effects associated with cell type composition in bulk tissues such as whole blood. We detected ten loci where the same genetic variant was associated with five or more target genes at a conservative p-value threshold ($p < 1e-11$). Three of these loci near CIITA, SP140 and USP18 also overlapped GWAS loci for several autoimmune diseases. A trans-eQTL variant near the USP18 gene was associated with both increased risk for systemic lupus erythematosus (SLE) and increased expression of up to 40 interferon-inducible genes. USP18 encodes a ubiquitin-specific peptidase that is a known negative regulator of interferon signalling, suggesting that the trans-eQTL variant likely impairs the ability of USP18 to effectively limit the interferon response. This observation is consistent with existing literature on rare diseases, where the dysfunction of USP18 is linked to the development of severe autoimmune disorders. We find that trans-eQTL analysis in a purified cell type is a powerful approach to interpret GWAS associations, but very large sample sizes are likely required for this to be practical. Furthermore, technical confounders such as mis-mapping RNA sequencing reads need to be carefully considered to avoid false positive associations.

Session Title: Molecular Effects of Genetic Variation Poster Session I

PB1174 Leveraging AlphaFold2 for the classification of genetic Variants of Uncertain Significance

Authors:

H. Murali[1,2], E. C. Liao[1], K. Wang[1,2]; [1]Children's Hosp. of Philadelphia, Philadelphia, PA, [2]Univ. of Pennsylvania, Philadelphia, PA

Introduction: Next-generation sequencing has revolutionized genetic testing in the clinic, revealing numerous rare disease-associate variants. However, the sheer volume of data poses new challenges in determining the clinical significance of variants, often resulting in their classification as Variants of Uncertain Significance (VUS). Current *in silico* variant pathogenicity prediction tools rely on evolutionary conservation, phylogeny, and ensemble methods, but overlook variant impact on 3D protein structure and related features. The emergence of AlphaFold2 has transformed the field of protein structure determination, so we aim to use insights from structural biology to tackle the "VUS classification challenge". **Objectives:** This study outlines a method that leverages AlphaFold2 and genetic variant databases to enhance VUS reclassification. We used the gene *IRF6* as a case study, since rescue experiments in *irf6*$^{-/-}$ zebrafish have validated several mutations as deleterious or benign. New variants continue to emerge and present new opportunities for prediction and functional validation. We aim to evaluate the accuracy of computational tools in predicting the pathogenicity of missense mutations. Key questions regarding consensus among tools, correlations between residue position and pathogenicity, distribution of predictions for each tool, and comparison with functional assays are addressed. **Methods:** We compared results from over 30 pathogenicity prediction tools on 37 *IRF6* variants, using ANNOVAR for inter-tool comparisons and experimental results for tool-experiment comparisons. As IRF6 lacks an experiment-derived structure, we used predicted structures from AlphaFold2 to explore correlations between mutation clustering and pathogenicity within its protein structure. **Results and Conclusion:** Among the analyzed variants, 19 of 37 were unanimously predicted as deleterious by computational tools. Comparing *in silico* predictions with experimental findings, 12 variants predicted as pathogenic were experimentally determined as benign. Mapping variants to the protein revealed deleterious mutation clusters around the protein binding domain, underscoring its significance. N-terminal variants showed contrasting results, with most being benign despite being predicted as deleterious by some tools. In general, these tools favor labeling variants as pathogenic, highlighting the need for improvement in current predictive tools. Incorporating protein structural features and analyzing mutation neighborhoods may enhance classification and provide meaningful insights into pathogenicity predictions.

Session Title: Molecular Effects of Genetic Variation Poster Session II

PB1175 Leveraging cell type-specific epigenomic data to dissect adult-onset and childhood-onset asthma GWAS loci.

Authors:

X. Zhong[1], C. Billstrand[1], N. Sakabe[1], I. Aneas[1], J. Gu[1], I. Salamone[1], W. Wentworth-Sheilds[1], R. Mitchell[1], A. I. Sperling[2], N. Schoettler[1], M. A. Nóbrega[1], X. He[1], C. Ober[1]; [1]Univ. of Chicago, Chicago, IL, [2]Univ. of Virginia, Charlottesville, VA

Asthma is a heterogenous disorder with age of onset being an important discriminating factor in differentiating asthma subtypes. GWAS of adult-onset asthma (AOA) and childhood-onset asthma (COA) have mapped more than 60 independent risk loci, yet the biological mechanisms that contribute to asthma at most risk loci remain unknown. Recent advances in epigenome profiling provides unprecedented opportunity to functionally characterize non-coding GWAS variants across trait-relevant cell types. Here, we applied a computational pipeline that integrates open chromatin regions (OCRs) and chromatin interactions in blood and lung cell types to dissect AOA and COA GWAS loci. We used a Bayesian hierarchical model (TORUS) to estimate the genome-wide enrichment of asthma risk variants in OCRs and performed functionally-informed fine-mapping with sum of single effects (SuSiE) regression, assigning higher prior causal probability to SNPs overlapping with highly-enriched OCRs. Across 52 fine-mapped independent LD blocks, we found 76 credible sets, 25 of which contained a SNP with high posterior inclusion probability (PIP), i.e., ≥ 0.8. Leveraging OCRs in blood and lung cell types, we further identified 21 candidate *cis*-regulatory elements harboring asthma risk variants with moderate to high PIPs. Using luciferase assays in a human bronchial epithelial cell (BEC) line (16HBE14o-), we validated allele-specific enhancer activity for two putative enhancer sequences, one on chromosome 19 which looped to the promoter of *CEBPA*, an enhancer-binding protein involved in airway inflammation, by PCHi-C in freshly isolated BECs. The second validated allele-specific enhancer resided in a gene-dense locus at 5q31.1 and interacted with the promoters of multiple genes by PCHi-C in primary blood cells. Further studies of these and other candidate *cis*-regulatory elements are currently underway. Taken together, our study revealed intricate regulatory landscapes underlying many asthma risk loci and demonstrated the utility of integrating epigenomic annotations in post-GWAS variant-to-function studies.

Session Title: Molecular Effects of Genetic Variation Poster Session III

PB1176 Leveraging plasma proteomics and supervised machine learning to interpret complex-trait genetics

Authors:

J. Irudayanathan[1], C. Benner[1], X. Xie[2], R. Pendergrass[1], A. Mahajan[1], M. McCarthy[1]; [1]Genentech Inc., South San Francisco, CA, [2]Genentech Inc, South San Francisco, CA

Genome-wide association studies (GWAS) have identified thousands of variants that are associated with complex traits and diseases. Prioritizing trait-relevant cis-genes that are downstream effectors of non-coding variants remains challenging. Here we introduce a supervised machine learning model designed to prioritize cis-genes given a GWAS variant, by training on a truth set derived from cis-acting protein quantitative loci (pQTL). Specifically, cis-pQTL identified in the discovery cohort of the UK Biobank Pharma Proteomics Project (UKB-PPP) consortium (n=35571; 2941 proteins assayed) were fine-mapped to resolve credible sets of causal variants with SuSie and FINEMAP. This resulted in 4589 variant-gene pairs mapping to 1400 unique autosomal proteins (credible sets with at least one variant satisfying PIP > 0.7). Variant-gene pairs were annotated with 10626 features derived from the Enformer predicted variant effects and genomic tracks. Linear genomic distance between the transcriptional start site (TSS) of the gene and the variant as well as the number of genes within 1MB of the signal were included as additional features. We utilized the extreme gradient boosted tree algorithm and trained the model using 3616 of the cis-pQTLs: we then evaluated model performance using withheld data.We consistently achieved auROC ~ 0.8 for correct prediction of the associated protein in withheld data. As expected, given that almost 60% of the cis-pQTLs in the training set mapped to the nearest gene, linear distance to TSS scored consistently as an important feature. To assess whether the model might be overfitted to selecting the nearest gene, we tested model performance in the subset of cis-pQTL signals (n=2579 variant-gene pairs) where the cognate protein-coding gene is not the nearest gene to the variant, and observed only a modest drop in accuracy (auROC ~ 0.74) in this subset.To assess if the cis-pQTL predictor model would generalize to prioritize genes driving phenotypes in disease-based GWAS settings, we utilized a compendium of 841 paired noncoding-coding signals derived from analyses in UK Biobank. These were instances where a regulatory variant associated with a given trait was co-located with, but independent of, a coding burden signal for the same trait. The model achieved auROC ~ 0.71 in the validation data set.Using this modeling framework, we were able to demonstrate the transferability of cis-pQTL predictions to the challenging task of predicting effector genes at complex-trait GWAS loci. Ongoing analyses seek to explore the value of integrating these predictions with those from other orthogonal methods.

Session Title: Molecular Effects of Genetic Variation Poster Session I

PB1177 Limited overlap of eQTLs and GWAS hits due to systematic differences in discovery

Authors:

H. Mostafavi, J. Spence, S. Naqvi, J. Pritchard; Stanford Univ., Stanford, CA

Most signals in genome-wide association studies (GWAS) of complex traits point to noncoding genetic variants with putative gene regulatory effects. However, currently identified expression quantitative trait loci (eQTLs) explain only a small fraction of GWAS signals. By analyzing GWAS hits for complex traits in the UK Biobank, and cis-eQTLs from the GTEx consortium, we show that these assays systematically discover different types of genes and variants: eQTLs cluster strongly near transcription start sites, while GWAS hits do not. Genes near GWAS hits are enriched in numerous functional annotations, are under strong selective constraint and have a complex regulatory landscape across different tissue/cell types, while genes near eQTLs are depleted of most functional annotations, show relaxed constraint, and have simpler regulatory landscapes. We describe a model to understand these observations, including how natural selection on complex traits hinders discovery of functionally-relevant eQTLs. Our results imply that GWAS and eQTL studies are systematically biased toward different types of variants, and support the use of complementary functional approaches alongside the next generation of eQTL studies.

Session Title: Molecular Effects of Genetic Variation Poster Session II

PB1178 Local Admixture Mapping of Proteins in African Americans

Authors:

D. Cruz[1], L. Raffield[2], S. Deng[1], J. Wilson[1], R. Gerszten[1]; [1]Beth Israel Deaconess Med. Ctr., Boston, MA, [2]UNC - Chapel Hill, Chapel Hill, NC

Introduction: There is increasing evidence that protein levels in African Americans (AA) are influenced by population differentiated genetic variation. We investigated how local genetic ancestry may be associated with levels of circulating plasma proteins using proteomic profiling in the Jackson Heart Study, an AA cohort, in an attempt to identify additional genetic determinants of protein levels. **Methods**: We performed admixture mapping using local ancestry estimates (probabilities of whether an individual has 0, 1 or 2 alleles of African ancestry at each site in the genome from RFMix, based on similarity to 1000G reference panel) from TOPMed sequencing data for autosomal SNPs. Associations were then tested for 1,300 proteins measured using the SOMAscan proteomic platform. Each local ancestry regression model was adjusted for age, sex, and estimated global ancestry. We used a previously established significance threshold for local African ancestry analysis, $p<2.1X10^{-5}$. We further adjusted findings for previously reported GWAS variants in conditional analysis. **Results**: There were 58 local admixture mapping signals for 55 proteins (three proteins had two signals each). Twenty-nine protein signals survived conditional analysis, proteins of note include Apolipoprotein L1 (whose cognate gene contains known major risk variants for chronic kidney disease in AA) and Apolipoprotein A-1 (a component of high-density lipoprotein cholesterol). **Conclusion**: We present local admixture mapping results providing further evidence that population differentiated genetic variants influence circulating levels of plasma proteins. For a large proportion of our findings, admixture associations were not attenuated by conditioning on previously reported SNPs from GWAS studies. These analyses suggest that admixture mapping may identify association signals that are missed by GWAS, and may help to elucidate population differences affecting key biologic pathways.

Session Title: Molecular Effects of Genetic Variation Poster Session III

PB1179 Long-read proteogenomics to connect disease-associated sQTLs to the protein isoform effectors of disease

Authors:

G. Sheynkman[1], A. Abood[2], M. Murali[1], C. Farber[1]; [1]Univ. of Virginia, Charlottesville, VA, [2]Univ. of Virginia, Raleigh, NC

A major fraction of loci identified by genome-wide association studies (GWASs) lead to alterations in alternative splicing, but interpretation of how such alterations impact proteins is hindered by the technical limitations of short-read RNA-seq, which cannot directly link splicing events to full-length transcript or protein isoforms. Long-read RNA-seq represents a powerful tool to define and quantify transcript isoforms, and recently, infer protein isoform existence. Here we present an approach that integrates information from GWAS, splicing QTL (sQTL), and PacBio long-read RNA-seq in a disease-relevant model to infer the effects of sQTLs on the ultimate protein isoform products they encode. We demonstrate the utility of our approach using bone mineral density (BMD) GWAS data, as well as other complex diseases such as Coronary Artery Disease (CAD) and Idiopathic Pulmonary Fibrosis (IPF). For BMD, we identified 1,863 sQTLs from the Genotype-Tissue Expression (GTEx) project in 732 protein-coding genes which colocalized with BMD associations (H4PP $\geq$ 0.75). We generated deep coverage PacBio long-read RNA-seq data (N=~22 million full-length reads) on human osteoblasts, identifying 68,326 protein-coding isoforms, of which 17,375 (25%) were novel. By casting the colocalized sQTLs directly onto protein isoforms, we connected 809 sQTLs to 2,029 protein isoforms from 441 genes. Overall, we found that 74 sQTLs influenced isoforms likely impacted by nonsense mediated decay (NMD) and 190 that potentially resulted in the expression of new protein isoforms. We will highlight several examples in which functional validation points to distinct isoform-level (not gene-level) effects, including TPM2. This long read RNA-seq and proteogenomics integrative approach will be described for various use cases. We expect our approach to be widely generalizable across diverse clinical traits and accelerate system-scale analyses of protein isoform activities modulated by GWAS loci.

Session Title: Molecular Effects of Genetic Variation Poster Session I

PB1180 Loss of an oligodendrocyte specific silencer element drives the cell specific overexpression of lamin B1 in Autosomal Dominant Leukodystrophy

Authors:

B. Nmezi, Q. Padiath; Univ. of Pittsburgh, Pittsburgh, PA

The importance of non-coding regulatory elements in disease causation is only now being studied in depth. This is especially true of genomic silencers of which there have been almost no cases that have been directly shown to cause disease. We have identified a novel cell type specific silencer element that plays a role in the fatal neurological disorder, Autosomal Dominant Leukodystrophy (ADLD). ADLD is characterized by extensive CNS demyelination and most cases are caused by tandem genomic duplications involving the *LMNB1* gene, while a small subset are caused by genomic deletions upstream of the gene. How mutations involving a widely expressed gene such as *LMNB1* cause such a specific demyelinating disorder is unknown. Leveraging data from recently-identified families that carry *LMNB1* gene duplications without exhibiting demyelination, ADLD patient tissues, and CRISPR-modified cell lines and mouse models, we have identified a novel silencer element that is lost in ADLD patients that specifically targets *LMNB1* overexpression to oligodendrocytes. This element consists of CTCF binding sites that mediate three-dimensional chromatin folding involving the *LMNB1* promoter and the recruitment of the PRC2 repressor complex. The loss of *LMNB1*'s silencer element in ADLD identifies a novel role for silencer elements in tissue specificity and disease causation.

Session Title: Molecular Effects of Genetic Variation Poster Session II

PB1181 Loss-of-function variants in *CUL3* cause a syndromic neurodevelopmental disorder

Authors:

T. Wang[1], P. Blackburn[2], F. Ebstein[3,4], T-C. Hsieh[5], M. Motta[6], F. Radio[6], J. Herkert[7], T. Rinne[8], I. Thiffault[9], M. Rapp[10], M. Alders[11], S. Maas[11], B. Gerard[12], T. Smol[13], C. Vincent-Delorme[14], B. Cogné[15], B. Keren[16], A. Ziegler[17], L. Hasadsri[18], P. Pichurin[18], E. Klee[18], K. Grand[19], P. Sanchez-Lara[19], E. Krüger[3], S. Bézieau[4], H. klinkhammer[20], P. Krawitz[20], E. Eichler[21], M. Tartaglia[6], S. Küry[22]; [1]Peking Univ. Hlth.Sci. Ctr., Beijing, China, [2]St. Jude Children's Res. Hosp., Memphis, TN, [3]Univ. Med. Greifswald, Greifswald, Germany, [4]Nantes Université, Nantes, France, [5]Univ. Hosp. of Bonn, Bonn, Germany, [6]Ospedale Pediatrico Bambino Gesù, Rome, Italy, [7]Univ. of Groningen, Groningen, Netherlands, [8]Radboud Univ. Med. Ctr., Nijmegen, Netherlands, [9]Children s Mercy, Kansas City, MO, [10]Children's Hosp. Colorado, Aurora, CO, [11]Amsterdam Univ. Med. Ctr., Amsterdam, Netherlands, [12]Ctr. Hosp.ier Univ.ire de Strasbourg, Strasbourg, France, [13]CHU Lille, Lille, France, [14]CH Arras, Arras, Cedex, France, [15]CHU NANTES, Nantes, France, [16]Sorbonne Université, Paris, France, [17]Univ. Hosp. of Angers, Angers, France, [18]Mayo Clinic, Rochester, MN, [19]Cedars-Sinai Med. Ctr., Los Angeles, CA, [20]Univ. of Bonn, Bonn, Germany, [21]Univ of Washington, Seattle, WA, [22]CHU de Nantes, Nantes, France

Purpose: *De novo* variants in *CUL3* (Cullin-3 ubiquitin ligase) have been strongly associated with neurodevelopmental disorders (NDDs), but no large case series have been reported so far. Here we aimed to collect sporadic cases carrying rare variants in *CUL3*, describe the genotype-phenotype correlation, and investigate the underlying pathogenic mechanism. **Methods:** Genetic data and detailed clinical records were collected via multi-center collaboration. Dysmorphic facial features were analyzed using GestaltMatcher. Variant effects on CUL3 protein stability were assessed using patient-derived T-cells. **Results:** We assembled a cohort of 35 individuals with heterozygous *CUL3* variants presenting a syndromic NDD characterized by intellectual disability with or without autistic features. Of these, 33 have loss-of-function (LoF) and two have missense variants. *CUL3* LoF variants in patients may affect protein stability leading to perturbations in protein homeostasis, as evidenced by decreased ubiquitin-protein conjugates in vitro. Specifically, we show that cyclin E1 (CCNE1) and 4E-BP1 (EIF4EBP1), two prominent substrates of CUL3, fail to be targeted for proteasomal degradation in patient-derived cells. **Conclusion:** Our study further refines the clinical and mutational spectrum of *CUL3*- associated NDDs, expands the spectrum of cullin RING E3 ligase-associated neuropsychiatric disorders, and suggests haploinsufficiency via LoF variants is the predominant pathogenic mechanism. **Keywords** Cullin-3 ubiquitin ligase, ubiquitination, dystonia, neurodevelopmental disorder

Session Title: Molecular Effects of Genetic Variation Poster Session III

PB1182 Macula transcriptome and eQTL analyses identify regional specificity and susceptibility to age related macular degeneration

Authors:

J. Advani[1], J. M. Rouhana[2,3], A. R. Hamel[2,3], P. A. Mehta[2,3], M. A. English[1], M. Kwicklis[1], D. A. Ferrington[4], A. V. Segrè[2,3], A. Swaroop[1]; [1]Neurobiology, Neurodegeneration and Repair Lab., Natl. Eye Inst., NIH, Bethesda, MD, [2]Ocular Genomics Inst., Dept. of Ophthalmology, Massachusetts Eye and Ear, Harvard Med. Sch., Boston, MA, [3]The Broad Inst. of MIT and Harvard, Cambridge, MA, [4]Dept. of Ophthalmology and Visual NeuroSci.s, Univ. of Minnesota, Minneapolis, MN

Age-related macular degeneration (AMD) is a multifactorial progressive neurodegenerative disease and a major cause of blindness among the elderly. Previous genome-wide association studies (GWAS) have identified genetic variants at 63 loci linked to AMD susceptibility loci. Our previous eQTL and TWAS analyses of peripheral retina have uncovered a number of target genes in AMD GWAS loci. However, functional consequences of the disease are largely manifested in the macula, which is the central region of the retina associated with high visual acuity. We therefore carried out RNAseq analysis of macula punches from 185 postmortem donor eyes with different AMD (Minnesota Grading System or MGS) grades and generated a reference transcriptome from controls (MGS1 n=60). We identified 12,975 protein coding genes and 955 non-coding genes in the reference transcriptome and 52 (MGS 1 vs MGS4), 24 (MGS 2 vs MGS4) and 135 (MGS3 vs MGS4) differentially expressed genes across different MGS grades in macula with (FDR &lt 0.1). Gene ontology and pathway analysis of reference macula transcriptome showed enrichment of axonogenesis, proteasome, regulation of autophagy and mitochondria related genes. Applying FastQTL to expression and imputed genotype data, we identified 7,085 expression quantitative trait loci (eQTLs) in cis (&lt 1 Mb) (FDR &lt 0.05). Gene ontology analysis of eGenes identified genes involved in vacuolar lumen, secretory vesicle, and mitochondrial matrix. We also compared eQTLs and eGenes identified in peripheral retina and macula and found 897 eQTLs and 5,037 eGenes to be common. Colocalization analyses with eCAVIAR between eQTL and AMD GWAS loci identified target genes at 6 AMD loci. Our data describes relationship between genetic regulation and expression in the macula, and association of eQTLs with AMD. We highlight the importance of examining target tissues/cell types relevant to complex diseases such as AMD. Future work will examine mQTLs and eQTMs in the macula using DNA methylation in control and AMD patients and include colocalization analyses and other methods to help identify additional causal genes and pathways underlying AMD pathology.

Session Title: Molecular Effects of Genetic Variation Poster Session I

PB1183 Mapping genetic contributions to individual- and cell type-specific brain oxidative stress responses

Authors:

B. Umans, O. Allen, Y. Gilad; Univ. of Chicago, Chicago, IL

Oxidative stress is a key feature of several diseases that affect the human brain, including Parkinson's Disease and hypoxia arising from apneas and respiratory infection. Observations from other tissues suggest that inter-individual differences in response to changing oxygen availability are genetically regulated and, moreover, that oxygen stress responses differ across cell types. However, the cell type-specific factors governing the brain's response to oxidative stress remain unclear, and the extent to which these differences contribute to disease susceptibility is unknown.
To understand how human brain cells vary in their responses to oxidative stress, we collected single-cell transcriptome data from a panel of dorsal brain organoids derived from 21 human iPSC lines, which were acutely exposed to elevated or reduced oxygen for 24 hours. Organoids from all individuals gave rise to a developmental spectrum of cell types, including subtypes of radial glia, intermediate progenitors, excitatory and inhibitory neurons, and early glia. In our dataset of over 185,000 cells, we conservatively identified 12 principal cell types and fit a linear model in each to identify differentially expressed genes, finding a range of sensitivities to oxidative stress across cell types. To explore the basis of this sensitivity, we mapped the spatial positions of organoid cell types using a panel of antibody markers. While organoids show stereotyped architectures on a local scale, spatial position correlated poorly with oxidative stress sensitivity, suggesting instead that cell-intrinsic features govern the selective vulnerability of developing cortical neurons.
We also assessed the contribution of germline genetic variation to oxidative stress responses in each cell type. Using existing whole genome sequencing data from the individuals in our brain organoid panel, we mapped expression quantitative trait loci (eQTLs) in steady-state, low-oxygen, and high-oxygen conditions. Despite limited power, our preliminary analysis identified hundreds of eQTL genes across cell types and oxygen conditions, with substantial shared effects across cell types and conditions. Future studies will explore the impact of gene regulatory differences on brain development in the context of other environmental perturbations and aim to connect *in vitro* eQTLs to additional developmental and functional phenotypes.

Session Title: Molecular Effects of Genetic Variation Poster Session II

PB1184 Mapping the landscape of lineage-specific dynamic regulation of gene expression using single-cell transcriptomics and application to genetics of complex disease

Authors:

H. Abe[1], P. Lin[1], D. Zhou[2], D. Ruderfer[1,2], E. Gamazon[1,2]; [1]Vanderbilt Univ., Nashville, TN, [2]Vanderbilt Genetics Inst., Nashville, TN

Transcriptome wide association studies (TWAS) have become a widely-used approach to investigate the association between genetically regulated expression and complex traits. Though this approach has elucidated the functional interpretation for disease associations at tissue level, its application at the single-cell level remains to be investigated. Single-cell transcriptome data can provide valuable insights into the context specificity of the effect of genetic variation on various cellular processes involved in human biology and disease. Here, we present sc-TWAS, an approach for developing in silico cell type specific and cell-state adjusted prediction models of gene expression. By integrating individual level genotype data and single cell transcriptomic measurements from the HipSci consortium, we used a machine learning approach to train TWAS models in 8 dopaminergic neuronal cell types that are in the process of specializing from iPSCs across 3 time points. Using S-prediXcan, we then apply our gene expression prediction models to the latest GWAS summary data of neurodevelopmental disorders such as Schizophrenia (SCZ) obtained from Psychiatric Genomic Consortium (PGC). Our results show that sc-TWAS can detect known and novel genes associated with disease traits such as SCZ and enables insights into context-dependent disease mechanisms. We show that some SCZ associated genes have improved prediction performance R^2 for specific cell type and differentiation time point models. We further provide a resource from a phenome-wide application of our cell type specific and cell-state adjusted models to more than 4000 phenotypes from the UK Biobank. PheWAS on the cell type specific SCZ associated genes using this resource, showed the enrichment of blood cellular traits such as leukocyte and erythrocyte count ($-\log10p > 50$). Finally, using the longitudinal predicted expression data, we perform transcription factor motif enrichment analysis to identify genes that are co-regulated with each of significantly associated SCZ genes. The results from this analysis point to a wider network of disease relevant genes and transcription factor regulators. Overall, this work demonstrates the utility of incorporating cell type specificity and dynamics of genetic regulation in gene expression prediction and association to disease.

Session Title: Molecular Effects of Genetic Variation Poster Session III

PB1185 Massively Parallel Reporter Assay Reveals Promoter Position-Dependent and Tissue-Specific Effects in Islet TSSs

Authors:

M. Bose[1], A. Tovar[1], A. Varshney[2], Y. Kyono[1], J. Kitzman[1], S. Parker[1]; [1]Univ. of Michigan, Ann Arbor, MI, [2]Univ. of Michigan, Canonsburg, PA

Massively parallel reporter assays (MPRA) are a high-throughput method to assess activity of candidate cis-regulatory sequences, and can be used to detect allelic differences at disease-associated variants. Previously published MPRA studies have screened functional SNPs associated with various traits and conditions, including type 2 diabetes. Here, we used an MPRA library to quantify activity of 1,305 pancreatic islet-derived TSSs generated from CAP analysis of gene expression (CAGE) profiling. We cloned oligos upstream or downstream of a reporter gene along with either the human insulin (INS) promoter or a synthetic housekeeping promoter (SCP1). We used generalized linear models (GLM) to predict position-specific oligo activity from tissue-specific chromatin state regulatory annotations. Overlap with islet promoters ($\beta = 9.20$, $p < 0.01$) and positive strandedness ($\beta = 11.91$, $p < 0.001$) both predicted significantly higher activity for oligos cloned in the upstream position. We also used Lasso regression to predict position-specific oligo activity based on enrichment of transcription factor motifs (TFMs). Oligos that displayed preferential activity upstream of the reporter gene were enriched for TFMs such as FOXJ2, and BHLHE22, and E2F. These significant differences were not observed for oligos that had higher activity when cloned downstream of the promoter and reporter gene. Currently, we are experimentally validating observed strandedness effects with a follow-up MPRA library. We designed 20 oligos representing the extremes of strandedness (i.e., oligos with strongest effects of strandedness) which we cloned with the same promoters along with promoters for other metabolic cell types including liver (APOA2), adipose (ADIPOQ), and skeletal muscle (MYBPC2). Together, these results support use of MPRA strategies that account for positional and promoter context-dependent factors when assaying candidate regulatory elements in pursuit of understanding complex genetic diseases.

Session Title: Molecular Effects of Genetic Variation Poster Session I

PB1186 Massively parallel reporter assays identify functional enhancer variants at QT interval GWAS loci.

Authors:

A. Kapoor[1], D. Lee[2], A. Chakravarti[3]; [1]UTHlth.Houston, Houston, TX, [2]Harvard Med. Sch. & Boston Children s Hosp., Boston, MA, [3]NYU Sch. of Med., New York, NY

Following a *cis*-regulatory mechanistic hypothesis for GWAS signals in the noncoding genome, we utilized massively parallel reporter assays (MPRA) to identify enhancer variants among 1,018 candidate noncoding variants across the 35 QT interval GWAS loci from the QTIGC study. 200-base long oligos with a 20-base 5′ flank, variant-centered 129-base sequence, 18-base spacer with SbfI and EcoRI sites for subsequent cloning of a minimal promoter-driven *eGFP* reporter cassette, 13-base unique barcode and a 20-base 3′ flank were designed. Each allele/variant was linked to 50 different barcodes for a total of 101,800 unique oligos. We used two different vector backbones in our episomal MPRA: pLS-mP, a lentiviral MPRA vector, propagated in E. coli Stable cells (pLS-mP/S), and pGL4.23, a classical enhancer assay vector, modified for MPRA and propagated in E. coli Stable (pGL4.23/S) or in DH10β (pGL4.23/D) cells. Mouse HL1 cardiomyocyte cells were transfected with the final MPRA plasmid libraries in 10 replicates, and paired-end sequencing of barcodes was performed in input plasmid DNA (3 replicates) and HL1 derived cDNA samples (10 replicates). Read counts for each barcode from input plasmid DNA were aggregated across the three replicates and normalized to total DNA reads. Read counts for each barcode from cDNA in each replicate were normalized to total cDNA reads. Reporter activity for each barcode in each replicate was assessed as $\log_2$ of the ratio of normalized read counts from cDNA to input plasmid DNA. Test element-level reporter activity was estimated as average of reporter activities across all corresponding barcodes and replicates. Of the 1,018 variants, both alleles at 1,014 (99.6%), 729 (71.6%) and 896 (88.0%) variants were successfully propagated in pLS-mP/S, pGL4.23/S and pGL4.23/D experiments, respectively, with failed elements being enriched for short C-rich sequences. Overall, the dynamic range of reporter expression in pLS-mP backbone was relatively limited, a likely outcome of the 5’ LTR promoter interfering with the minimal promoter downstream of the test elements, and was not used to identify enhancers. Among the shared variants (n=729) across the two pGL4.23-based experiments, the correlation was high (r=0.98), highlighting robust reporter activity measurements. Using a >10% increase in cDNA reads count over input plasmid (for either ref. or alt. test element) to define enhancers, 177 enhancer elements were identified of which 106 showed significant allelic difference by t-test at false discovery rate of <0.01. These enhancer variants map to 22 GWAS loci, and are being evaluated for causality by explaining target gene expression variation.

Session Title: Molecular Effects of Genetic Variation Poster Session II

PB1187 Massively parallel reporter assays of enhancer variants at selected QT interval GWAS loci.

Authors:

L. Gunamalai[1], D. Lee[2], A. Kapoor[1]; [1]UTHlth.Houston, Houston, TX, [2]Harvard Med. Sch. & Boston Children s Hosp., Boston, MA

Electrocardiographic QT interval, a measure of ventricular de- and re-polarization, is a moderately heritable quantitative trait of clinical relevance as prolongation or shortening of the QT interval is associated with an increased risk of cardiac arrhythmias and sudden cardiac death. Genome-wide association studies (GWAS) have identified tens of trait associated loci, however, the identities of the causal variants and their molecular mechanisms underlying these common noncoding variants-based associations remain largely unknown. Following a *cis*-regulatory mechanistic hypothesis for GWAS signals in the noncoding genome, we used massively parallel reporter assays (MPRA) to identify enhancer variants at 12 selected QT interval GWAS loci. These 12 selected loci encompass genes implicated in Mendelian disorders of abnormal QT interval (long- and short-QT syndromes) and/or genes known to effect cardiomyocyte action potential duration or excitation-contraction coupling. Using a recently published QT interval GWAS meta-analysis, we identified all variants (n=15,455) that are within ± 500 kb of the lead variant and are either genome-wide significant ($P<5\times10^{-8}$) or in moderate-to-high linkage disequilibrium ($r^2>0.3$) with a genome-wide significant variant. Next, we used variant overlap with publicly available open chromatin regions identified in human cardiac left ventricle by DNase-seq or in human ventricular cardiomyocytes by single cell ATAC-seq (extended 300 bp on either side of peak summits) to filter 1,198 candidate variants. Of these, nine variants were filtered out as either being INDELs >20 bases or either allele along with its flanking sequence matching restriction enzyme sites used in cloning. MPRA oligos (300 base long) were designed to test both alleles at each of the remaining 1,189 variants in three different 229-base long flanking sequence contexts (left-shifted, center and right shifted) for a total of 7,134 unique test elements, with each element linked to 20 unique barcodes for a total of 142,680 oligos across two equally-sized pools. Once cloned in a MPRA-modified pGL4.23 backbone along with a minimal reporter-driven reporter gene, the two MPRA libraries will be transfected in mouse cardiomyocyte HL1 cells in 10 replicate transfections. Barcode sequencing from input plasmid DNA and HL1-derived RNA (cDNA) samples will be followed to assess enhancer activities and allelic differences. We expect our MPRA-based screen to identify multiple *cis*-regulatory enhancer variants that collectively regulate target gene expression, thereby underlying trait association.

Session Title: Molecular Effects of Genetic Variation Poster Session III

PB1188 Maternal genetic variation shapes human milk and the infant gut microbiome

Authors:

K. Johnson[1], T. Heisel[1], E. Isganaitis[2], M. Rudolph[3], D. Fields[3], C. Gale[1], F. Albert[1], E. Demerath[1], R. Blekhman[4]; [1]Univ. of Minnesota, Minneapolis, MN, [2]Joslin Diabetes Ctr., Harvard Med. Sch., Boston, MA, [3]Univ. of Oklahoma Hlth.Sci. Ctr., Oklahoma City, OK, [4]Univ. of Chicago, Chicago, IL

Human milk is a complex mix of nutritional and bioactive components that provide complete nutrition for the infant. However, we lack a systematic understanding of the genomic factors shaping milk composition, and how milk variation influences infant health. Here, we generated multi-omic profiles of 242 exclusively breastfeeding mother-infant pairs, including maternal genotypes, milk transcriptome, milk metabolome, and infant fecal metagenome, to identify connections between human milk variation and infant development.

We identified cis-eQTLs for 2,690 genes expressed in human milk (FDR<5%), 487 of which were specific to the lactating mammary gland (i.e. did not colocalize with GTEx eQTLs). Genes with milk-specific eQTLs were enriched for genes in dairy cattle milk trait QTL (odds ratio=2.1, $P=1.7x10^{-3}$). Motivated by the protective role of lactation duration against breast cancer, we identified 8 milk eQTLs colocalized with breast cancer GWAS loci, including a milk-specific eQTL for *LMX1B* with a GWAS locus previously lacking a functional annotation (H4 posterior probability=0.99).

Leveraging our multi-omic study design, we identified 67 milk eQTL variant-metabolite associations mediated by milk gene expression (FDR<5%), including known examples like *FUT2* and fucose ($B=0.88$, $P=5.9x10^{-21}$). The *SORT1* locus, which was a milk eQTL, was associated with multiple metabolites including gamma-glutamylglutamine ($B=0.63$, $P=3.3x10^{-7}$). Expression of creatine transporter *SLC16A12* was associated with decreased milk creatine ($B=-2.4$, $P=1.3x10^{-7}$). As milk metabolites can promote or inhibit bacteria in the infant gut, we tested if eQTLs that influenced milk metabolites were also associated with infant gut microbe abundances. Three eQTLs associated with an infant gut microbe (FDR<5%); e.g., the variant associated with increased *SLC16A12* expression and decreased milk creatine was correlated with higher levels of an unnamed *Streptococcus sp.* in infant feces ($B=0.46$, $P=3.5x10^{-5}$).In sum, our results reveal new insights into how human milk variation can impact maternal and infant health, highlighting a need for further work on this biological system heretofore neglected by the field of human genetics.

Session Title: Molecular Effects of Genetic Variation Poster Session I

PB1189 Measuring (almost) all possible *STK11/LKB1* missense variants in monogenic disease and cancer

Authors:

D. Zimmerman, A. Cote, W. Loggerenberg, M. Gebbia, J. Weile, N. Kishore, F. Roth; Univ. of Toronto, Toronto, ON, Canada

Germline loss of the Serine/Threonine Kinase 11 (*STK11/LKB1*) tumor suppressor leads to Peutz-Jeghers Syndrome (PJS), a rare monogenic disease conferring gastrointestinal complications and an increased lifetime cancer risk. Somatic loss of *STK11* variants are found in up to 30% of lung cancers, 20% of cervical, and several other tissue types at lower frequencies. Early genetic diagnosis of PJS helps establish life-long disease management, allowing for proactive screening and early interventions. Despite this and the American College of Medical Genetics (ACMG) inclusion of *STK11* as one of their 73 clinically actionable genes, over 85% of *STK11* germline clinical missense variants are currently classified as Variants of Uncertain Significance (VUS). To provide evidence to improve variant interpretation, we applied a Multiplexed Assay of Variant Effect (MAVE), based on *en masse* expression in mammalian cells and ensuing growth advantage of dysfunctional coding variants, to characterize variant functions for >70% of possible *STK11* missense variants. The resulting variant effect map shows strong agreement with known structural and biochemical properties, and performs well in separating pathogenic or benign annotations. Thus, we provide a resource for sequence-structure-function insights in STK11, and provide evidence for both known and yet-to-be-observed clinical STK11 variants.

Session Title: Molecular Effects of Genetic Variation Poster Session II

PB1190 Mediation of *APOE* Associations with Cognition through ATN

Authors:

J-W. Jang; Kangwon Natl. Univ. Hosp., Chuncheon, Korea, Republic of

Mediation of APOE Associations with Cognition through ATNBackground: The e+4 allele of the apolipoprotein E (APOE) gene is a significant genetic risk factor for age-related cognitive decline. However, the precise causal relationship between e+4 alleles and cognition remains poorly understood. This study aimed to investigate the involvement of Amyloid/Tau/Neurodegeneration (ATN) in the associations between APOE and cognition.Methods: Multiple linear regression analyses were performed on a sample of 1158 subjects (mean age of 73.3 years; 45.0% female; 46.2% APOE e+4 allele carriers). Causal mediation analyses, using 5,000 bootstrapped iterations, were conducted to examine the mediation effects.Results: The APOE e+4 allele showed a negative association with cognition ($P < 0.05$) and ATN (all $P < 0.05$). The impact of the APOE genotype on cognition was partially mediated by ATN (all $P < 0.05$).Conclusion: ATN plays a partial mediating role in the potential links between the APOE genotype and cognition. Overall, the APOE e+4 allele may contribute to dysregulation of the ATN cascade, which subsequently leads to cognitive impairment.

Session Title: Molecular Effects of Genetic Variation Poster Session III

PB1191 METTL23 mutation alters histone H3R17 methylation in inherited normal tension glaucoma

Authors:

T. Iwata[1], Y. Pan[1], A. Suga[2], I. Kimura[3], C. Kimura[4], Y. Minegishi[5], M. Nakayama[1], K. Yoshitake[1], D. Iejima[1], N. Minematsu[1], M. Yamamoto[6], F. Mabuchi[7], M. Takamoto[8], Y. Shiga[9], M. Araie[8], K. Kashiwagi[7], M. Aihara[8], T. Nakazawa[9]; [1]Nat Hosp. Organization Tokyo Med. Ctr., Meguro-ku, Japan, [2]Natl. Hosp. Organization Tokyo Med. Ctr., Tokyo, Japan, [3]Tokai Univ. Hachioji Hosp., Hachioji, Japan, [4]Kimura Eye Clinic, Iwaki, Japan, [5]Japanese Fndn. for Cancer Res., Toyoshima-ku, Japan, [6]JAC Ltd., Chiyodaku-ku, Japan, [7]Univ. of Yamanashi, Yamanashi, Japan, [8]Univ. of Tokyo, Bunkyo-ku, Japan, [9]Tohoku Univ. Graduate Sch. of Med., Sendai, Japan

Normal-tension glaucoma (NTG) is a heterogeneous disease characterized by retinal ganglion cell (RGC) death leading to cupping of the optic nerve head and visual field loss at normal intraocular pressure (IOP). The pathogenesis of NTG remains unclear. Here, we describe a single nucleotide mutation in exon 2 of the methyltransferase like 23 *(METTL23)* gene identified in a three-generation Japanese NTG family. We assessed its effects on expression, localization, and methylation by splicing using three transfected cell lines, patient-derived induced pluripotent stem cells (iPSCs), and *METTL23* c.A83G knock-in mice. We also determined whether *METTL23* defects cause RGC degeneration and optic nerve head fiber loss, the typical human NTG phenotype, in vivo*;* and explored the role of METTL23 in methylation in vitro and in vivo. This mutation caused *METTL23* mRNA aberrant splicing, which abolished normal protein production and altered subcellular localization. *Mettl23* knock-in (*Mettl23*$^{+/G}$ & *Mettl23*$^{G/G}$) and knockout (*Mettl23*$^{+/-}$ & *Mettl23*$^{-/-}$) mice developed a glaucoma phenotype without elevated IOP. METTL23 is a histone arginine methyltransferase expressed in murine and macaque RGCs. However, the novel mutation reduced Mettl23 expression in RGCs of *Mettl23*$^{G/G}$ mice, which recapitulated both clinical and biological phenotypes. Moreover, our findings demonstrated that Mettl23 catalyzed the dimethylation of H3R17 in the retina, and was required for the transcription of *pS2*, an estrogen receptor α target gene that was critical to RGC homeostasis through the negative regulation of NF-κB-mediated TNF-α/IL-1β feedback. This study showed for the first time that alterration of epigentics in the retina can devolp NTG. Our findings suggest an etiologic role of METTL23 in NTG with tissue-specific pathology.

Session Title: Molecular Effects of Genetic Variation Poster Session I

PB1192 Microbiome and host genetic effects on gene regulation in Inflammatory Bowel Disease.

Authors:

S. Nirmalan[1], R. Blekhman[2], R. Pique-Regi[1], F. Luca[1]; [1]Wayne State Univ., Detroit, MI, [2]Univ. of Chicago, Chicago, MN

Inflammatory Bowel Disease (IBD) leads to chronic inflammation and scarring of the digestive tract. The gut microbiome is altered in IBD and many genetic risk variants associated with IBD are in genes that function in host response to microbes. Both host genetic variants and the gut microbiome regulate host gene expression in the colon; however, it is still unknown whether interactions between host genetics and the microbiome (GxM) regulate host gene expression in the gut in IBD. Identifying these GxM that modify genetic risk for IBD is imperative for understanding its pathogenesis. Here, we used data from the Human Microbiome Project to perform interaction eQTL mapping to identify GxM in IBD. Our sample consisted of host genotype, and paired host gene expression and mucosal gut microbiome (16S rRNA) data from rectum biopsies for 86 individuals (64 patients with IBD and 22 controls). We identified 3298 genes with eQTLs (eGenes, FDR = 10%). These eGenes are enriched in both the GTEx sigmoid colon (OR = 2.5, p < 2.2 * 10^-16) and transverse colon (OR = 3.7, p < 2.2 * 10^-16) eQTL datasets. We integrated the microbiome data with the eQTLs to identify GxM. We tested for interaction with the relative abundance of each microbe and found 37 interactions for 28 genes and 20 microbial taxa (FDR = 10%). We also tested for interaction with the presence/absence of a microbe and found 15 additional GxM (12 genes and 10 taxa; FDR = 10%). Six genes harbor multiple GxM and 8 microbes interact with multiple genes. For example, the abundance of the Blautia genus, an anaerobic microbe with anti-inflammatory activity, interacts with 7 eQTLs in genes involved in cell signaling and inflammatory response. The expression of 13 GxM genes is associated with IBD risk in PhenomeXcan. For example, increased expression of LAMTOR is associated with increased risk of IBD. We found that increased abundance of Blautia is associated with decreased expression of LAMTOR in individuals with the alternative allele, potentially decreasing IBD risk. Our results show the first evidence of genetic effects on gene expression that are modulated by the microbiome composition in IBD, and provide insight into how IBD risk could be decreased by targeting specific microbial taxa depending on host genotype.

Session Title: Molecular Effects of Genetic Variation Poster Session II

PB1193 Mitochondrial DNA haplogroup K is protective of Autism Spectrum Disorder risk in populations of European ancestry.

Authors:

T. Wang, X. Chang, K. Ostberg, H. Qu, Y. Liu, J. Glessner, H. Hakonarson; Children's Hosp. of Philadelphia, Philadelphia, PA

Cumulative evidence indicates a critical role of mitochondrial function in Autism Spectrum Disorder (ASD) and implies that ASD risk may be linked to mitochondrial dysfunction due to variations in mitochondrial DNA (mtDNA). Although a few studies have explored the association between mtDNA variations and ASD, the role of mtDNA in ASD is still unclear. Here, we aimed to investigate whether mitochondrial DNA haplogroups are associated with the risk of ASD. Two European cohorts and an Ashkenazi Jewish (AJ) cohort were analyzed, with a total of 2,062 ASD patients compared with 4,632 healthy controls. DNA samples were genotyped using Illumina HumanHap550/610 and Illumina 1M arrays, inclusive of mitochondrial markers. Mitochondrial DNA (mtDNA) haplogroups were identified from genotyping data using HaploGrep2. A mitochondrial genome imputation pipeline was established to detect mtDNA variants. We conducted a case-control study to investigate potential associations of mtDNA haplogroups and variants with the susceptibility of ASD. We observed that the ancient adaptive mtDNA haplogroup K was significantly associated with decreased risk of ASD in the two European cohorts, which had a total of 2,006 cases and 4,435 controls (odds ratio 0.64, $P=1.79\times10^{-5}$), and we replicated this association in the Ashkenazi Jewish (AJ) cohort of 56 cases and 197 controls (odds ratio 0.35, $P=9.46\times10^{-3}$). Moreover, we demonstrated that mtDNA variants rs28358571, rs28358584 and rs28358280 are significantly associated with ASD risk. Additionally, expression quantitative trait loci (eQTL) analysis indicated that the rs28358584 and rs28358280 genotypes are associated with expression levels of nearby genes in brain tissues, suggesting that these mtDNA variants may confer risk to ASD via regulation of mitochondrial gene expression. In conclusion, this study sheds light on the role of mitochondria in ASD and provides new insights into the genetic mechanism underlying ASD via the potential involvement of mtDNA encoded proteins in the development of ASD.

Session Title: Molecular Effects of Genetic Variation Poster Session III

PB1194 Molecular and cellular characterization of a lymphatic anomaly-associated mutation in *TEK*

Authors:

M. March[1], D. Li[2], S. Sheppard[3], L. Matsuoka[4], M. E. Matos[1], C. Kao[1], H. Hakonarson[2]; [1]Children s Hosp. of Philadelphia, Philadelphia, PA, [2]Children's Hosp. of Philadelphia, Philadelphia, PA, [3]NICHD/NIH, Rockville, MD, [4]Children S Hosp. of Philadelphia, Philadelphia, PA

As part of our ongoing efforts to characterize putatively causative variants that we have uncovered in a growing number of patients with arterial, venous, or lymphatic malformations (collectively referred to as cardiovascular anomalies or CVAs), we have identified a variant in the *TEK* gene (p.R915C) in a patient with a venous malformation. In order to characterize the effects of this variant in a relevant cell system, we expressed the variant and wild-type *TEK* in primary human dermal lymphatic endothelial cells, a cell system we have used to great effect to determine pathogenicity of variants and to characterize the response of variant-expressing cells to potential treatments. Cells expressing p.R915C *TEK* displayed increased activity in a spheroid sprouting assay (used to measure lymphangiogenesis potential *in vitro*) and displayed increases in phosphorylation in both ERK and AKT, indicating activation of both MAP kinase and PI-3 kinase signaling pathways. Treatment of cells with inhibitors of MEK (for MAPK signaling) and PIK3CA (for PI-3 kinase signaling) partially reversed the observed phenotypes, while treatment with rebastinib, a broadly active compound that possesses inhibitor activity towards Tie2 (the protein encoded by the *TEK* gene) completely reversed the observed changes in cell behavior. Taken together, these results suggest that directly targeting the Tie2 protein itself would be most beneficial in patients carrying *TEK* variants, as activation of Tie2 stimulates multiple downstream pathways.

Session Title: Molecular Effects of Genetic Variation Poster Session I

PB1195 Molecular characterization of *PDGFRB* p.Asn666 variants

Authors:

T. Gladkauskas, O. Bruland, R. Mehrasa, E. Rødahl, C. Bredrup; Univ. of Bergen, BERGEN, Norway

Several gain-of-function substitutions at p.Asn666 platelet-derived growth factor receptor beta (PDGFRB) have been identified, namely p.Asn666Tyr, p.Asn666Ser, p.Asn666His, and p.Asn666Lys. The *PDGFRB* p.Asn666Lys variant is specifically associated with somatic infantile myofibromatosis, with the presence of this variant observed solely in a single case of gingival myofibroma. In contrast, the *PDGFRB* p.Asn666Ser variant is linked to Penttinen premature aging syndrome, characterized by premature aging, lipodystrophy, acro-osteolysis, chronic ulcers, and corneal vascularization. Additionally, the *PDGFRB* p.Asn666Tyr variant is associated with Ocular-pterygium digital keloid dysplasia, a condition marked by corneal vascularization and keloid formation on fingers/toes in otherwise healthy individuals. Lastly, the *PDGFRB* p.Asn666His variant has been identified in a single patient presenting with hemangioma, sagittal craniosynostosis, intracranial cyst, brachydactyly, and acro-osteolysis. The objective of this study was to investigate the underlying molecular mechanisms that contribute to the distinct phenotypes observed in these four p.Asn666 *PDGFRB* variants. Using ELISA and immunoblot analyses of transduced immortalized fibroblasts expressing wild-type PDGFRβ and these four variants, we observed constitutive activation of PDGFRβ across all substitutions. Importantly, we found that each variant displayed a unique pattern of phosphorylated tyrosine residues in PDGFRβ, leading to variant-specific upregulation of downstream signaling proteins including, p-AKT, p-STAT1, p-MAPK3/ERK1, and p-PLCγ1. In conclusion, our study provides evidence that the four described variants at the p.Asn666 position of PDGFRβ exhibit unique phosphorylation patterns in specific tyrosine residues of PDGFRβ and downstream signaling proteins. This could contribute to the distinctly different clinical conditions caused by these variants.

Session Title: Molecular Effects of Genetic Variation Poster Session II

PB1196 Molecular tools for determination of post-trauma injury complications

Authors:

D. McDaniel, S. B. Robertson, B. Williams, G. A. Timberlake, L. S. McDaniel; Univ Mississippi Med Ctr, Jackson, MS

Background: An assessment of applying transcript level of genes with influential variability and differential expression analysis are powerful tools for sepsis biomarker determination during post trauma injury complication. Clinical characteristics associated with sepsis in combination with molecular markers are study tools to derive informative gene subsets for early determination of disease. Sepsis is a life-threatening response of the immune system to trauma induced complications which can potentially lead to tissue damage, organ failure, and death. Numerous risk factors in the context of inflammatory parameters such as cytokines and Toll-like receptor genes contribute to an early inflammatory response to the pathogen or damage associated molecular patterns (DAMPs) associated with trauma injury. Methods: A panel of candidate risk factors associated with the innate immune response genes were evaluated using blood from trauma patients within the first 72 hours upon the admission to SICU. Leukocytes including monocytes, macrophages and neutrophils for mRNA transcript analysis of cytokines, and the TLR-signaling pathway genes were tested. The expression was confirmed by quantitative PCR and plasma specimen from patients by ELISA. Results: The study confirms an early upregulation of the innate immune response through monocytes/macrophages, extends to neutrophil functional dynamics. TLR-2 and TLR-4 transcripts were significantly increased comparing day 1 vs. day 3 in patients who later developed sepsis. The IRAK-1, a major mediator of the TLR-signaling pathway was reversibility increased (2.0- fold) in patients with sepsis vs. no sepsis. TOLIP gene, a Toll-like receptor inhibitory protein, also known as Cox-2 gene was increased 4-fold in patients with no sepsis. The CD86 an HLA-Class II receptor molecule on monocytes and macrophages significantly increased in patients who later developed sepsis. Conclusions: Trauma causes changes in macrophage and monocyte activation. This data support correlation between TLR-2 and TLR-4 signaling molecules with the expression of pro-inflammatory cytokines in trauma induced sepsis. The performance of gene selection model is likely to be improved by extending to a broader dataset.

Session Title: Molecular Effects of Genetic Variation Poster Session III

PB1197 *MTOR* variants increase the risk of developing thoracic aortic disease

Authors:

D. Guo[1], X. Duan[1], Z. Zhou[1], P. Gu[1], Y. Yu[1], I. Marin[1], M. Bamshad[2], D. Milewicz[1]; [1]UT McGovern Med. Sch. at Houston, Houston, TX, [2]UUniv. of Washington, Seattle, WA

Normal 0 false false false false EN-US ZH-CN X-NONE /* Style Definitions */ table.MsoNormalTable {mso-style-name:"Table Normal"; mso-tstyle-rowband-size:0; mso-tstyle-colband-size:0; mso-style-noshow:yes; mso-style-priority:99; mso-style-parent:""; mso-padding-alt:0in 5.4pt 0in 5.4pt; mso-para-margin-top:0in; mso-para-margin-right:0in; mso-para-margin-bottom:8.0pt; mso-para-margin-left:0in; line-height:107%; mso-pagination:widow-orphan; font-size:11.0pt; font-family:"Calibri",sans-serif; mso-ascii-font-family:Calibri; mso-ascii-theme-font:minor-latin; mso-hansi-font-family:Calibri; mso-hansi-theme-font:minor-latin; mso-bidi-font-family:"Times New Roman"; mso-bidi-theme-font:minor-bidi;} To identify new genes for heritable thoracic aortic disease (HTAD), we developed a machine learning method to prioritize human genes based on established HTAD genes, protein-protein interaction, and mutational constraints, and *MTOR* emerged as a candidate gene (missense z-score = 7.0). *MTOR* rare variants (minor allele frequency ≤ 0.0001) predicted to be damaging were identified in exome data from 392 affected HTAD probands, 58 probands from trios, and 524 individuals with early-onset sporadic aortic dissections (< 60 years old; ESTAD). Eleven *MTOR* rare variants were identified and half the variants were located at the FAT or kinase domain of mTOR, which is location of gain-of-function *MTOR* mutations in cancer. Kinase function of mTOR was assessed for *MTOR* p.Arg1616His, p.Asn2292Ser, and p.Asp2520Val (located in kinase or FAT domains) by co-transfecting Flag-tagged *MTOR* constructs, with wildtype, an activating variant, or three exome variants, and a downstream protein phosphorylated by mTOR (HA-tagged 4EBP1) into HEK293T cells. Immunoblot analyses revealed that the *MTOR* variants p.Asn2292Ser and p.Asp2520Val significantly increased the phosphorylation of 4EBP1 when compared to wild-type mTOR, and to a similar levels as gain-of-function variants in cancer. To confirm a role of mTOR signaling in TAD, mTOR signaling was assessed in a mouse model of thoracic aortic dissection (mice exposed to beta-aminopropionitrile (BAPN)). Immunoblot analyses revealed increased levels of mTOR ($p<0.001$) and p70S6K ($p<0.05$) phosphorylation in the proximal aortas of BAPN-treated mice compared to control mice. Importantly, an mTOR inhibitor, rapamycin, prevented dissection deaths in this mouse model ($p <0.0001$). These data support that activating rare variants in *MTOR* predispose to TAD, and increased mTOR signaling contributes to the molecular pathogenesis of thoracic aortic disease.

Session Title: Molecular Effects of Genetic Variation Poster Session I

PB1198 Multi-ancestry Whole-exome Sequencing Study of Alcohol Use Disorder in Two Cohorts

Authors:

L. Wang; Yale Univ., New Haven, CT

Background/Hypothesis: Alcohol use disorder (AUD) is a leading cause of death and disability worldwide. There has been substantial progress in genetic studies of AUD and problematic drinking. However, the SNP-based heritability attributable to common variants is low compared to expectations based on genetic epidemiology. Whole-exome sequencing (WES) is a powerful tool for identifying disease-associated variants, especially rare variants that may convey high odds risk ratios, but there is a lack of WES studies of AUD. Methods: WES datasets with 4,530 subjects from the Yale-Penn cohort and 413,476 samples from the UK Biobank were investigated to identify associated variants. DSM-IV alcohol dependence (AD) was diagnosed using the Semi-Structured Assessment for Drug Dependence and Alcoholism in the Yale-Penn cohort. AUD phenotype definition for the UK Biobank cohort was based on ICD-10 criteria. After quality control procedures, 1,668 European ancestry (EUR, 1,409 cases and 259 controls) and 1,547 African ancestry (AFR, 1,156 cases and 391 controls) participants from Yale-Penn were retained for subsequent analysis. WES data from 406,195 EUR samples (9,767 cases and 396,428 controls) and 7,281 AFR samples (111 cases and 7,170 controls) from the UK Biobank were also analyzed. Single-variant analyses and gene-based tests were conducted, followed by within-ancestry meta-analyses and cross-ancestry meta-analyses integrating both Yale-Penn and UK Biobank datasets. Enrichment of deleterious protein-truncating variants and other downstream analyses were also performed to investigate the contributions of coding variants. Results: Cross-ancestry meta-analysis integrating the Yale-Penn and the UK Biobank datasets identified the well-known functional variant rs1229984 in *ADH1B* associated with AUD (p=1.73e-31). Three other common variants—rs2298753, rs33988198 and rs17526590—in high linkage disequilibrium in the *ADH1C* gene were also significant after correcting for multiple testing. All these three noncoding variants are independent signals from rs1229984 and are also captured as significant signals in our previous problematic alcohol use genome-wide association study. Conclusions: Our study confirmed the association between rs1229984 and AUD and provided evidence for the exome-wide significant associations in the *ADH1C* gene with AUD. Recruitment of additional AUD subjects is needed to provide sufficient statistical power to identify rare coding variants that could account for missing heritability in the analysis of common variations.

Session Title: Molecular Effects of Genetic Variation Poster Session II

PB1199 Multimodal analysis of RNA sequencing data powers genomic discovery.

Authors:

D. Munro[1,2,3], N. Ehsan[2], R. Vogel[1,2,3], M. Esmaeili-Fard[3], A. Gusev[4,5], A. A. Palmer[1], P. Mohammadi[3,6,2]; [1]Univ. of California San Diego, La Jolla, CA, [2]Scripps Res., La Jolla, CA, [3]Seattle Children's Res. Inst., Seattle, WA, [4]Dana-Farber Cancer Inst., Boston, MA, [5]Harvard Med. Sch., Boston, MA, [6]Univ. of Washington, Seattle, WA

Transcriptome data has been shown to identify the biological mechanisms driving genome-wide association study (GWAS) signals in populations. While RNA sequencing (RNA-seq) can reveal many aspects of transcriptome regulation, due to the practical complexity of extracting a diverse array of RNA phenotypes and applying them to downstream analyses, transcriptomic studies are often limited to one or two forms of RNA phenotypes, primarily total gene expression.

Here, we present Pantry (Pan-transcriptome phenotyping), a framework to generate diverse RNA phenotypes from RNA-seq data and perform downstream integrative analyses with GWAS data. Pantry generates six categories of transcriptional phenotypes (gene expression, isoform ratios, splice junction usage, alternative TSS/polyA, and mRNA stability) and applies them to molecular quantitative trait locus (QTL) mapping and transcriptome-wide association study (TWAS).

Using GEUVADIS data we mapped between 4,098 (alternative TSS) and 14,145 (gene expression) independent cis-QTLs per RNA phenotype category. We also demonstrate Pantry's utility with generalized TWAS using six categories of transcriptional phenotypes from GEUVADIS data and 114 GWAS traits. We found 10,845 significant hits across 80 traits involving 4,796 unique RNA phenotypes for 2,064 genes. Notably, of the 4,794 unique trait-gene associations mapped, 49.8% involved only non-expression RNA phenotypes, which would not have been identified in standard TWAS using only gene expression. This approach also allows for more fine-grained identification of regulatory mechanisms underlying GWAS signal in a large fraction of previously associated gene-trait pairs. Finally, we provide statistical analysis of the informational overlap across RNA phenotypes to help identify biologically independent signals. We provide the Pantry code, RNA phenotypes from all GEUVADIS and GTEx samples, and downstream QTL and TWAS results on the web.

Session Title: Molecular Effects of Genetic Variation Poster Session III

PB1200 Multiome single-nuclear profiling of chromatin accessibility and gene expression in human liver samples: insights into cell-type-specific regulation.

Authors:

A. Alkhawaja[1], K. W. Currin[1], H. J. Perrin[1], S. Vadlamudi[1], A. S. Etheridge[1,2], G. H. Cannon[3], C. W. Anderson[3], A. H. Moxley[1], E. G. Schuetz[4], F. Innocenti[2,5], A. C. Iuga[6], T. S. Furey[1], K. L. Mohlke[1]; [1]Dept. of Genetics, Univ. of North Carolina, Chapel Hill, NC, [2]Eshelman Sch. of Pharmacy, Univ. of North Carolina, Chapel Hill, NC, [3]Advanced Analytics Core, Univ. of North Carolina, Chapel Hill, NC, [4]Dept. of Pharmaceutical Sci., St. Jude Children's Res. Hosp., Memphis, TN, [5]Lineberger Comprehensive Cancer Ctr., Univ. of North Carolina, Chapel Hill, NC, [6]Dept. of Pathology and Lab. Med., Univ. of North Carolina, Chapel Hill, NC

Disruption of metabolic regulation in liver can lead to impaired glucose and lipid accumulation, inflammation, and development of liver steatosis, fibrosis, and hepatitis. While previous studies using bulk tissue sequencing methods identified genetic variants associated with gene regulation in liver tissue, cell-type-specific effects remain understudied. Identifying genetic variants that affect gene regulation within distinct liver cell types may provide valuable insights into how the different liver functions are regulated at the molecular level. We jointly profiled single nucleus chromatin accessibility (snATAC-seq) and gene expression (snRNA-seq) in 40 human liver samples using the 10X Genomics multiome platform. These samples were obtained from 35 white and 5 black donors (25 males/15 females), aged 13 to 75 years old with body mass index values of 16 to 42 kg/m^2. Histopathological examination of 38 of the 40 liver samples revealed varying degrees of steatosis and fibrosis, including 7 samples that exhibited both conditions. After quality control, we identified 69,216 nuclei from 39 samples, representing 7 cell types based on canonical cell markers. Similar to published liver single cell studies, most nuclei were from hepatocytes (68%), while B-cells were the least abundant (1.4%). Unlike non-parenchymal cells, hepatocytes clustered by donor, suggesting a substantial influence of donor-specific genetic or environmental factors on hepatocytes. Of the 151,296 accessible chromatin regions (peaks) identified across all cell types, 97% overlapped peaks from bulk liver ATAC-seq from a superset of 138 individuals. We identified 50,641 peaks that were more accessible in one cell type relative to all others (DESeq2, FDR<5%), including 33,046 peaks more accessible in non-hepatocyte cell types. GWAS heritability enrichment in upregulated peaks identified contributions of distinct cell types to different traits, including enrichment of alanine aminotransferase and apolipoprotein B levels in hepatocytes, gamma-glutamyl transferase levels in cholangiocytes, and blood pressure in mesenchymal cells. We are currently identifying cell type expression quantitative trait loci (eQTL) and chromatin accessibility QTL (caQTL) and will colocalize QTL signals with GWAS loci for liver disease-relevant traits. In addition, cell type differential chromatin accessibility and gene expression in healthy vs steatosis/fibrosis samples will be investigated. These integrative analyses should provide insights into the role of liver cell type gene regulation in disease-relevant traits and identify potential targets for therapeutic interventions.

Session Title: Molecular Effects of Genetic Variation Poster Session I

PB1201 Multiple sequence modifications result in loss of chitinolytic activity in YKL-40.

Authors:

K. Suzuki[1], K. Okawa[1], M. Ohkura[1], T. Kanaizumi[1], E. Tabata[1,2], F. Oyama[1]; [1]Kogakuin Univ, Tokyo, Japan, [2]Res. Fellow of Japan Society for the Promotion of Sci. (PD), Tokyo, Japan

Human YKL-40 (CHI3L1, HC GP-39), a chitinase-like protein (CLP) family member, shares structural similarities with the active chitinase CHIT1 but lacks chitinolytic activity. Elevated expression of YKL-40 has been observed in asthma, chronic obstructive pulmonary disease (COPD), inflammatory bowel disease (IBD), alcoholic cirrhosis, Alzheimer's disease, and cancer, making it a potential diagnostic marker. Understanding the biochemical properties of YKL-40 is crucial for unraveling disease mechanisms and developing effective treatments. While amino acid substitutions in the catalytic motif have been proposed as the cause of YKL-40 inactivation, our attempts to activate YKL-40 through specific substitutions (A138D and L140E) in the catalytic motif did not yield the desired results. This result suggests that the loss of chitinolytic activity in YKL-40 cannot be solely attributed to these amino acids. To identify the regions responsible for YKL-40 inactivation, we generated chimeric proteins by combining segments of MT-YKL-40 and CHIT1. Incorporating certain exons of YKL-40 into CHIT1 led to enzyme inactivation while introducing corresponding CHIT1 exons into MT-YKL-40 activated the protein. Furthermore, WT- and MT-YKL-40, along with the chimeric proteins, exhibited chitin recognition and binding, indicating that YKL-40 retains its chitinase framework but has selectively lost chitinolytic activity. Our findings suggest that the loss of enzymatic activity in YKL-40 results from extensive sequence modifications throughout the molecule rather than isolated amino acid mutations.

Session Title: Molecular Effects of Genetic Variation Poster Session II

PB1202 Multi-signal eQTL power in the context of missing colocalization.

Authors:

J. Rosen[1], S. M. Brotman[1], K. A. Broadaway[1], K. L. Mohlke[1], M. I. Love[1,2]; [1]Dept. of Genetics, Univ. of North Carolina, Chapel Hill, NC, [2]Dept. of Biostatistics, Univ. of North Carolina, Chapel Hill, NC

One method to implicate genes at genome-wide association study (GWAS) signals is colocalization with gene expression quantitative trail loci (eQTL) signals; however, the modest number of GWAS-eQTL colocalizations observed to date has led to speculation that these two categories of signals are fundamentally different. To identify factors that may contribute to this colocalization gap, we investigated statistical power in eQTL studies, specifically considering that better powered eQTL studies exhibit more allelic heterogeneity, detected as multiple conditionally distinct signals per gene. We fit several parametric models to the empirical distribution of effect sizes of conditionally distinct signals observed in an eQTL study of ~2,200 individuals, demonstrating that the optimal fit is achieved using a log normal distribution. We utilize a model for estimating the power to detect eQTL signals as a function of effect size, sample size, and minor allele frequency (MAF) and propose a method to estimate the percentage of all eQTL signals detected as a function of sample size. We assessed how well this method could predict the relative number of primary eQTL signals identified in dorsolateral prefrontal cortex samples from CommonMind and BrainSeq as reported in the eQTL Catalogue, observing agreement between the predicted and observed number within 2%. We then determined the sample sizes needed to detect 80% of all eQTL signals in a hypothetical study across a range of minor allele frequencies. Identifying eQTL with MAF of 25, 10, 5, and 1% required sample sizes of ~5,000, 10,500, 19,800, and 95,000 individuals, respectively. Current sample sizes may be sufficient to saturate detection of genes with at least one eQTL but are insufficient to saturate detection of all eQTL signals. We then estimated the percentage of all eQTL signals detectable in studies of comparable size to the Genotype-Tissue Expression project (GTEx). At MAF of 25, 10, 5, and 1%, studies with a sample size of 700 would detect 29, 14, 6.5, and <1% of the theoretical total of all QTL signals, respectively. These results demonstrate that typical eQTL studies are only well-powered to detect large eQTL effect sizes and remain underpowered to detect most eQTL signals. Finally, we explored the relationships between signal distance to transcription start site (TSS) and signal strength and, in conjunction with the power analyses, interpreted these results in the context of eQTL colocalization with GWAS signals, ultimately concluding that insufficient power contributes substantially to missing colocalization.

Session Title: Molecular Effects of Genetic Variation Poster Session III

PB1203 Nanoparticle enrichment mass-spectrometry proteomics identifies protein-altering variants for precise pQTL mapping

Authors:

S. Batzoglou; Seer, Inc, Redwood City, CA

Genome-wide association studies (GWAS) with proteomics generate hypotheses on protein function and offer genetic evidence for drug target prioritization. Although most protein quantitative loci (pQTLs) have so far been identified by high-throughput affinity proteomics platforms, these methods have some limitations, such as uncertainty about target identity, non-specific binding of aptamers, and inability to handle epitope-modifying variants that affect affinity binding. Mass spectrometry (MS) proteomics has the potential to overcome these challenges and broaden the scope of pQTL studies. Here, we employ the recently developed MS-based Proteograph™ workflow (Seer, Inc.) to quantify over 18,000 unique peptides from almost 3,000 proteins in more than 320 blood samples from a multi-ethnic cohort. We implement a bottom-up MS-proteomics approach for the detection and quantification of blood-circulating proteins in the presence of protein altering variants (PAVs). We identify 184 PAVs located in 137 genes that are significantly associated with their corresponding variant peptides in MS data (MS-PAVs). Half of these MS-PAVs (94) overlap with cis-pQTLs previously identified by affinity proteomics pQTL studies, thus confirming the target specificity of the affinity binders. An additional 54 MS-PAVs overlap with trans-pQTLs (and not cis-pQTLs) in affinity proteomics studies, thus identifying the putatively causal cis-encoded protein and providing experimental evidence for its presence in blood. The remaining 36 MS-PAVs have not been previously reported and include proteins that may be inaccessible to affinity proteomics, such as a variant in the incretin pro-peptide (GIP) that associates with type 2 diabetes and cardiovascular disease. Overall, our study introduces a novel approach for analyzing MS-based proteomics data within the GWAS context, provides new insights relevant to genetics-based drug discovery, and highlights the potential of MS-proteomics technologies when applied at population scale.
Highlights (1) This is the first pQTL study that uses the Proteograph™ (Seer Inc.) mass spectrometry-based proteomics workflow. (2) We introduce a novel bottom-up proteomics approach that accounts for protein altering variants in the detection of pQTLs. (3) We confirm the target and potential epitope effects of affinity binders for cis-pQTLs from affinity proteomics studies. (4) We establish putatively causal proteins for known affinity proteomics trans-pQTLs and confirm their presence in blood. (5) We identify novel protein altering variants in proteins of clinical relevance that may not be accessible to affinity proteomics.

Session Title: Molecular Effects of Genetic Variation Poster Session I

PB1204 Natural variation in splicing is widespread across humans and can be used to interpret variants in complex traits

Authors:

H. Jacobs[1], B. Gorissen[1], B. Neale[2], H. Finucane[2], K. Karczewski[3], C. Burge[1]; [1]MIT, Cambridge, MA, [2]Massachusetts Gen. Hosp., Boston, MA, [3]Broad Inst., Medford, MA

A fundamental goal of human genetics is to characterize how genomic variants impact human phenotypes. Studies of quantitative trait loci, including those of expression (eQTLs) and splicing (sQTLs), have been successful in identifying the molecular mechanisms of some complex trait-associated variants. However, these studies explain a small fraction of known GWAS loci. In particular, sQTL studies are primarily powered to detect highly expressed known splicing events. A full understanding of splicing variation across humans might enhance GWAS interpretation by providing additional functional support for associated variants.

To explore natural variation in splicing, we use the GTEx project data to quantify population variation of percent spliced in (PSI) for thousands of exons in 49 different tissues. We discover ~70k "naturally-varying" splice events, 80% of which are not found in the GENCODE reference annotations. Non-reference splice sites have the canonical GT/AG structure and high splice site motif scores (median MaxEntScan = 7.85 versus ~8-9 for known splice sites). Many of these splice sites are found in multiple individuals, providing support for true splicing events missing in the reference dataset, and not just sequencing noise.

We implemented empirical bayes to quantify the fraction of individuals in a tissue in which an exon is included at a potentially biologically relevant level (PSI > 5%), calling this fraction exon frequency (EF). Interestingly, many splicing events have low exon frequency ("low EF"), meaning they exist only in a small subset of individuals, and are completely absent in others. Most low EF splicing events are missing from reference annotations, likely because they are only present in a subset of the population and in specific tissues. We highlight reasonably common (MAF 12-15%) genetic variants in genes ASGR1 and ECRG4 that are both sQTLs in GTEx and associated with traits in the UK BioBank actually increase splicing of an unannotated, natural-varying splicing event. Interestingly, splicing of those exons ranges between 0-20% across all individuals. This demonstrates the need to reliably detect splicing events with lower PSI values to aid in the interpretation of complex traits.

Besides genetic variants that explain low EF splicing, we also provide examples of trans-acting factors, such as age, that increase the number of low EF splicing events. Finally, we found that genetic variants associated with low EF splicing events are depleted in GTEx sQTL discovery, suggesting that these are identifying independent signals. We are currently performing a systematic detection of variants that impact low EF splicing events.

Session Title: Molecular Effects of Genetic Variation Poster Session II

PB1205 Naturally occurring compounds block an exonic splicing silencer and partially rescue normal gene expression in a genotype of choroideremia.

Authors:

I. MacDonald, A. Radziwon; Univ Alberta, Edmonton, AB, Canada

In two unrelated families, an unusual synonymous change in the CHM gene, c.1359CtoT, p.(S453S) causes choroideremia, an X-linked retinal dystrophy. This non-canonical splice variant results in skipping of exon 11 in roughly 99.5% of total transcripts. In-silico analysis predicted the involvement of the known repressor hnRNPA1.Methods: Bibliographic search identified several compounds that could potentially inhibit hnRNPA1's repressor activity to promote exon 11 inclusion, restoring a normal transcript. Patient cells were treated with these and full-length CHM expression quantified by qPCR.Results: In-vitro screening showed that quercetin and shikonin significantly increased the inclusion of exon 11. Quercetin is a flavonoid polyphenol commonly found in fruit, while shikonin is a natural product isolated from purple gromwell root used in traditional Chinese medicine.Conclusions: These pre-clinical results may provide a lead to a potential treatment for this particular CHM variant. Encouragingly, shikonin was able to produce a statistically significant difference even at 50 nM, suggesting that an effective concentration may be achievable through oral supplementation. While this is the first reported instance in CHM where the creation of a putative hnRNPA1 recognition motif is the molecular mechanism, this class of mutation has been identified in numerous other genetic disorders. Once identified, these cases may be amenable to a similar treatment.
Two brothers with this variant presented with disparate phenotypes: typical disease progression in one and the other virtually unaffected. The aberrant transcript predominated in both, and qPCR found no statistically significant difference between them in two assayed cell types. Retinal pigment epithelium has not been examined. These results likely indicate that the observed residual 0.5% CHM expression relative to normal is very near the sufficiency threshold. The differences in the phenotypes may relate to the critical amount of protein that must be produced to maintain normal retinal function. As such, a modest increase in normal transcript may have dramatic effect on disease progression.
A clinical trial of shikonin supplementation could be planned, comparing the natural history of disease before and after treatment.

Session Title: Molecular Effects of Genetic Variation Poster Session III

PB1206 Navigating the health landscape of sex chromosome trisomies using phenome-wide associations of medical records from three international cohorts

Authors:

D. Lapato[1], S. Davis[2], A. Liu[3], C. Teerlink[4], A. Gentry[1], E. Lancaster[1], M. Singh[1], C. Chris[5], G. Genovese[6], B. Gorman[7], T. Bigdeli[5], J. Lynch[8], J. Ross[9], A. Ganna[3], R. Hauger[10,11], R. E. Peterson[5]; [1]Virginia Commonwealth Univ., Richmond, VA, [2]Univ. of Colorado Sch. of Med., Aurora, CO, [3]Inst. for Molecular Med. Finland, Helsinki, Finland, [4]Veterans Admin. Informatics and Computing Infrastructure, Salt Lake City, UT, [5]SUNY Downstate Hlth.Sci. Univ., Brooklyn, NY, [6]Broad Inst. of Harvard and MIT, Boston, MA, [7]VA Boston Hlth.care System, Boston, MA, [8]Salt Lake City VA Med. Ctr., Salt Lake City, UT, [9]Nemours Children's Hlth., Wilmington, DE, [10]Univ. of California San Diego, San Diego, CA, [11]VA San Diego Hlth.care System, San Diego, CA

Sex chromosome trisomies (SCT) involve the gain of chromosome X or Y, resulting in Klinefelter syndrome (KS; 47, XXY), Jacob syndrome (JS; 47, XYY), or Triple X syndrome (TX; 47, XXX). SCTs are among the most prevalent chromosomal abnormalities observed in liveborn humans and affect ~1 in every 500-1000 males and ~1 in 1000 females born each year. SCTs have been associated with medical and neuropsychiatric comorbidities, including learning disorders, metabolic syndromes, and atypical sexual development and fertility. Because the majority of SCT research has focused on clinically ascertained cases, the full spectrum of health risks and comorbidities for individuals with a SCT is unknown due to biased sampling. This study sought to identify a broader array of health outcomes associated with SCT by applying a hypothesis-free reverse genetics approach to paired genotype measures and health records from three international cohorts.

A phenome-wide association study (PheWAS) meta-analysis was conducted using data from the Million Veteran Program (MVP), FinnGen Research Project, and UK Biobank (UKB). Individuals with SCTs were identified using standard genotype microarray quality control assessments that estimate chromosomal sex. ICD9/10 codes were mapped to phecodes, and the relative prevalence of phecodes was compared between SCT cases and controls matched 1:5 (cases:controls) for age, sex, and race within each dataset. Logistic regression was run with SCT as the predictor, age at enrollment, race, and the first ten genetic principal components as covariates, and phecodes as the binary outcome. P-values were corrected using FDR Benjamini Hochberg procedures. Random effects meta-analysis was applied given the expected heterogeneity across biobanks.

N=2758 individuals across the three biobanks were predicted to carry a SCT based on genotypic data. The SCTs identified were KS (n=1320), JS (n=1096), and TX (n=342). In total, over 90,000 ICD codes were mapped onto 1875 unique phecodes. Significant phecodes reflected a multi-system array of comorbidities, including dermatological (e.g., skin ulcers, cellulitis), circulatory (e.g., chronic venous insufficiency), metabolic (e.g., Type 2 diabetes mellitus), immune/autoimmune (e.g., arthritis), and musculoskeletal (e.g., osteoporosis) conditions.

These results align with published epidemiological studies and suggest that SCTs may have a wider range of comorbidities than previously reported. Additional work will be needed to determine the etiopathogenesis of those conditions. Robust, generalizable knowledge about the comorbidities of SCTs will be invaluable for providing informed health care.

Session Title: Molecular Effects of Genetic Variation Poster Session I

PB1207 Neuronal transcription of autism gene *PTCHD1* is regulated by a conserved downstream enhancer sequence

Authors:

S. Pastore, T. Muhammad, P. Frankland, P. Hamel, J. Vincent; Univ. of Toronto, Toronto, ON, Canada

Introduction: *Patched domain-containing 1* (*PTCHD1*) is a susceptibility gene for autism and intellectual disability (ID) encoded at Xp22.11. Research has suggested that alterations in the dosage of *PTCHD1* may contribute to the pathoetiologies of these disorders. Despite this, there is a paucity of data to illuminate the mechanisms that regulate *PTCHD1* expression. This study sought to characterize the *Ptchd1* promoter in a mouse neuronal model, and identify distal cis-regulatory elements that may affect its transcription. **Methods:** To characterize the *Ptchd1* promoter, mouse embryonal carcinoma (P19) lines stably-expressing reporter constructs with promoter truncations were generated. Stable lines were promoted towards the neuronal lineage, and luciferase activity was evaluated. Putative transcription factor binding sites (TFBSs) were inferred within the promoter using the SwissRegulon database. In parallel, analyses of publicly-available DNase I-seq. datasets were conducted to describe the chromatin landscape surrounding *Ptchd1*. Specifically, comparisons were performed between mouse neonatal forebrain and liver tissues, which display high and low relative levels of *Ptchd1* expression, respectively. Next, the Wellington algorithm from the pyDNase package was utilized to identify DNase footprints within accessible regions, followed by motif discovery using HOMER. Lastly, CRISPR-Cas9 was employed to delete candidate regulatory regions in P19 lines, followed by differentiation, RNA isolation, and RT-qPCR to assess *Ptchd1* expression. **Results:** The 276 bp genomic region spanning -422 to -147 relative to the transcription start site appears to be critical for the expression of *Ptchd1*, and also contains evolutionarily-conserved TFBSs for the transcription factor Sp1, as well as for the chromatin-remodeling protein CHD1. An enhancer-containing genomic region located downstream of the *Ptchd1* coding sequence was accessible to DNase I in the forebrain, but inaccessible in the liver. This region included a DNase footprint with a putative binding site for the transcriptional activator-repressor YY1. Deletion of this entire downstream open chromatin region attenuated *Ptchd1* expression in neurons by over 64%. **Conclusion:** This study described segments of the *Ptchd1* promoter, and detected conserved TFBSs that may mediate its expression. In addition, an enhancer-containing putative downstream regulatory region was ascertained, deletion of which diminished *Ptchd1* expression in neurons. Collectively, these data clarify the *PTCHD1* promoter and provide context regarding structural variations that may have neurodevelopmental consequences.

Session Title: Molecular Effects of Genetic Variation Poster Session II

PB1208 New insights into biotinidase deficiency in a large carrier screening cohort.

Authors:

E. Kramer Dugan[1], Y. Wang[1], M. Hajjar[1], B. Prigmore[1], V. Souter[1], P. Benn[2], B. Wolf[3]; [1]Natera, Inc., Austin, TX, [2]Univ. of Connecticut Hlth.Ctr., Farmington, CT, [3]Ann and Robert H. Lurie Children's Hosp. of Chicago, Chicago, IL

Introduction: Biotinidase deficiency (BD) is a treatable, autosomal recessively inherited disorder with a reported prevalence of approximately 1 in 60,000 for profound and partial enzyme deficiency. BD is a core disorder included in newborn screening in the US. While newborn screening will likely detect all cases of profound BD, it may fail to detect all cases of partial deficiency because of enzymatic cut-offs. The objective of this study was to determine the variants in the *BTD* gene, and the carrier frequency ascertained in a reproductive carrier screening cohort. Methods: This retrospective study included female individuals, aged 18-55 years who had reproductive carrier screening with a 274-gene panel between Jan 2020 and Sep 2022. Carrier frequencies for pathogenic/likely pathogenic (P/LP) variants were evaluated for the entire cohort and stratified by self-reported race/ethnicity. In addition, carrier frequencies were calculated for D444H with or without other variants, because D444H is usually associated with partial BD but can cause profound BD if in allelic combination with A171T (and a small number of other variants). Results: The overall frequency of P/LP *BTD* variants in the study population was 5,620/91,637 (6.1%) including 5,092/91,637 (5.6%), which were D444H variants. The frequency of D444H was higher in those self-identifying as White (7.9%) compared to other racial/ethnic groups, with the lowest frequency (1.5%) in the Black group. After excluding D444H (alone or in combination with another variant), the frequency of all other P/LP variants combined was 508/91,637 (0.6%) with a lower frequency in the Asian group (0.3%) compared to White (0.6%). The most common combination of variants was homozygosity for D444H in 88 individuals (0.10%), followed by D444H/A171T found in 53 individuals (0.06%). Excluding these combinations, we identified 19 individuals with two or more pathogenic variants (0.02%), 16 of whom had at least one copy of D444H. These could be indicative of an affected individual (variants in trans) or a complex carrier state (in cis), and where additional variant phase testing was indicated.Conclusion: Reproductive carrier screening that includes evaluation of *BTD* identifies a high frequency of variants, but only a small proportion have the potential of resulting in a child with profound BD. Accurate counseling about the D444H variant is important for patient understanding of the disease variability. Benefits of prenatal screening include identification of potentially affected pregnancies and subsequent neonatal surveillance.

Session Title: Molecular Effects of Genetic Variation Poster Session III

PB1209 New T2T CHM13 genome emphasizes the importance of proper genomic structure and indicates dramatic gene movements. Genomics: #ItsComplicated

Authors:

M. Wadsworth[1,2], M. Page[1,2], B. Aguzzoli Heberle[1,2], J. Miller[1,3,4,5], M. Ebbert[1,2,6]; [1]Sanders-Brown Ctr. on Aging, Univ. of Kentucky Coll. of Med., Lexington, KY, [2]Dept. of NeuroSci., Univ. of Kentucky Coll. of Med., Lexington, KY, [3]Dept. of Pathology and Lab. Med., Univ. of Kentucky, Lexington, KY, [4]Dept. of Microbiol., Immunology, and Molecular Genetics, Univ. of Kentucky, Lexington, KY, [5]Div. of BioMed. Informatics, Dept. of Internal Med., Univ. of Kentucky, Lexington, KY, [6]Div. of BioMed. Informatics, Internal Med., Univ. of Kentucky Coll. of Med., Lexington, KY

Background:
Having a correct genome structure and annotation is critical to truly understand human health and disease for many reasons, including establishing which genomic variations are “normal” or “healthy”. The first complete telomere-to-telomere genome assembly, CHM13, provides us a second de novo human reference genome to compare genome structure and annotations, and provides invaluable insight into the previously incomplete regions of the genome. Here, we perform a systematic comparison of the regions common to both the HG38 and T2T CHM13 reference genomes to assess structural and annotation differences between the two references and found significant variation. Our results highlight: (1) how difficult it is to generate an accurate genome annotation, and (2) how essential a high-quality pan-genome with proper annotations will be to fully understanding human health and disease.
Methods:
We performed systematic comparisons between the structures and annotations for both the HG38 and T2T CHM13 reference genomes using genomic arithmetic (bedtools), sequence identity comparisons (BLAT), and camouflage assessment based on our previous work (genes that cannot be adequately assembled or aligned with short reads).
Results:
Between the HG38 and CHM13 references, we identified four major classes of annotation differences, totaling 1193 genes that change chromosomes entirely and/or are duplicated in new regions of the genome; the four classes include: (1) 68 single-copy non-syntenic genes (change chromosomes, entirely); (2) 24 non-syntenic duplications (change chromosomes & duplicated); (3) 476 non-syntenic and syntenic duplications (remain on original chromosome & duplicated); and (4) 625 syntenic duplications (remain on original chromosome & duplicated on same chromosome). As expected, ~70% overlap camouflaged regions, leaving them unrepresented in short-read sequencing data. One region (~5852 nucleotides) includes five genes whose sequences, according to HG38, are duplicated on chr1 from the mitochondria, but are entirely absent from chr1 in CHM13 (confirmed in population data). We also identified 281 genes that are uniquely annotated in HG38.
Conclusion:
While the CHM13 reference genome finally “completes” the human genome, it also demonstrates how far we have to go. We highlight major structural changes between the two genomes, some of which we validated as legitimate population differences rather than being corrections to HG38. A high-quality pan-genome with accurate annotations will simply be the beginning of fully understanding genomic variations, as we ultimately need to construct each individual’s genome structure.

Session Title: Molecular Effects of Genetic Variation Poster Session I

PB1210 N-of-1 personalized medicine: Identification of a putative small fiber peripheral polyneuropathy-causing mutation in the glycolytic intermediate enzyme,*PHGDH*.

Authors:

M. Glanz[1], J. McCarthy[2], T. Bowers[3], S. Hofherr[2], E. Frise[2], M. Reese[2], M. E. Merritt[1], E. T. Wang[1]; [1]Univ. of Florida, Gainesville, FL, [2]Fabric Genomics, Oakland, CA, [3]Broad Inst., Kingston, MA

Millions of Americans are impacted by peripheral neuropathies, a heterogenous class of neurological diseases characterized by debilitating neuropathic pain and sensory loss. One-third of peripheral neuropathies are idiopathic; therefore, limited understanding of underlying mechanisms makes diagnosing and treating these diseases extraordinarily challenging. Here, we describe the characterization of a candidate gene, phosphoglycerate dehydrogenase (*PHGDH*), for a small fiber peripheral polyneuropathy in a presumed n-of-1 patient. Whole genome paired-end Illumina sequencing was performed using 10X Genomics library preparation. Genomic analysis was conducted with Fabric Genomics GEM, a proprietary, cutting-edge Artificial Intelligence driven algorithm that integrates multiple genomic and clinical datasets to produce a ranked list of candidate casual variants with >95% sensitivity, including candidate genes of uncertain significance. In this n-of-1 proband, we identified two allelic mutations in *PHGDH*: 1468G>A and 792+6T>G. Two children of the proband carry a single 1468G>A variant, while a third child and the proband's mother carry a single 792+6T>G variant. The 1468G>A variant reduces the catalytic activity of *PHGDH*, while the effect of 792+6T>G was unknown. Mini-gene splicing assays revealed that this latter mutation results in exon 7 skipping, potentially through weakening of the 5' splice site. Analysis of patient fibroblasts detected a reduction in *PHGDH* transcript and protein levels. To explore strategies for abrogating *PHGDH* exon 7 skipping and restore optimal enzyme levels, we performed a high-throughput point mutagenesis screen of *PHGDH*exon 7 and flanking introns. This approach identified putative exon 7 regulatory splice sites which may serve as therapeutic targets. Given the prominent sensory nerve clinical phenotype, we developed a new clinical diagnostic method characterizing specific nerve fiber subtype innervation patterns *in situ* from 3 mm patient skin punch biopsies. This technique enables accurate identification of distal nerve fiber degeneration patterns, thus providing an unprecedented phenotypic analysis of nerve fiber death. Our work reveals novel functional roles for *PHGDH* mutants and their causal candidacy for a rare form of neuropathy. Collectively, our approach lays groundwork for new phenotypic diagnostics and RNA-targeting therapeutic development for polyneuropathies.

Session Title: Molecular Effects of Genetic Variation Poster Session II

PB1211 Non-syndromic congenital sideroblastic anemia: Phenotype and genotype of nine Indian patients.

Authors:

R. DONGERDIYE[1], P. Kedar[2]; [1]ICMR-Natl. Inst. of Immunohaematology, Mumbai, Maharashtra, India, [2]Natl. Inst. of Immunohematology (ICMR), MUMBAI, India

Congenital Sideroblastic Anemia (CSA) are a group of inherited and acquired bone marrow failure disorder defined by pathological iron accumulation in the erythroid precursors or mitochondria. CSA is caused by mutations in genes involved in the heme biosynthesis, iron -sulfur [Fe-S] cluster biosynthesis, and mitochondrial protein synthesis. The most prevalent form of CSA is X-linked SA (XLSA), caused by mutation in erythroid-specific **δ**;- aminolevulinate synthase (*ALAS2*), which is the first enzyme of heme synthesis pathway in erythroid cells. Other lesser- known genes responsible for causing CSA are *SLC25A38, HSPA9, SLC19A2, TRNT1, ABCB7* and few more. Here, we describe nine cases of CSA confirmed by genomic sequencing. The detailed clinical examination revealed microcytic anemia (in 7 of 9 cases), presence of ring sideroblasts (in 7 of 9 cases), dyserythropoiesis (in 2 of 9 cases), and iron accumulation in bone marrow in all the cases. Presently, only two patients are responding well to pyridoxine while others are on blood transfusion support. Here, we report two patients with XLSA, inheriting two hemizygous varinats in gene *ALAS2*: c.844G>T, p. (Ala282Ser) and c.508C>T, p. (Arg170Cys). We have also identified two homozygous mutations in gene *SLC25A38*: c.569C>G, p. (Pro190Arg); c.400C>T, p. (Arg134Cys) and two heterozygous variants in *HSPA9* gene: c.976G>T p. (Asp326Tyr), c.1388C>T p. (Thr463Ile). The latter is co-inherited with a frameshift mutation (c.137_138delCA, p. (Pro46Glnfs*8)) caused by a 2 base deletion in the *ALAS2* gene.The functional consequences of these variants were assessed using bioinformatics tools and their effect on the protein structure was investigated using PyMoL. In view of the results obtained in this study, we suggest bone marrow examination combined with genetic sequencing offers reliable and confirmed diagnosis to the CSA patients. Also, congenital sideroblastic anemia due to mutation in *SLC25A38* is prevalent than *ALAS2* in India. In conclusion, this study demonstrates the seven mutations in three distinct genes that causes non-syndromic CSA in the Indian population.

Session Title: Molecular Effects of Genetic Variation Poster Session III

PB1212 Novel and known minor alleles of *CNTNAP2* gene variants are associated with comorbidity of intellectual disability and epilepsy phenotypes: A case-control association study.

Authors:

B. Mudassir, Z. Agha; COMSATS Univ., Islamabad, Pakistan

Neurodevelopmental disorders are heterogeneous due to underlying multiple shared genetic pathways and risk factors. Intellectual disability, epilepsy and autism spectrum disorder phenotypes overlap which indicates the diverse effects of common genes. Recent studies suggested the probable contribution of *CNTNAP2* gene polymorphisms to the comorbidity of these neurological conditions. This study was conducted to investigate the role of *CNTNAP2* polymorphisms rs147815978 (G>T) and rs2710102 (A>G) as a risk factor for comorbidity of intellectual disability and epilepsy in a group of 345 individuals including 170 patients and 175 healthy controls recruited from various ethnic groups of Pakistani population. Our case-control study group was genotyped by tetra primer ARMS-PCR technique and results were analysed to know the effects of *CNTNAP2* rs 147815978 (G>T) and rs2710102 (A>G) polymorphisms in the group. The frequency of risk allele T (rs147815978) and risk allele G (rs2710102) for homozygous recessive genotypes (TT/GG) in our study group was 36.47% while odds ratios for risk allele T (rs147815978) was 5.45 (3.90-7.61:95%CI, P=0.000) and that for risk allele G (rs2710102) was 2.39 (1.76-3.24: 95%CI, P=0.0001). Homozygous recessive genotypes (TT/GG) appeared only in cases and not in control group which indicated these as suspected risk genotypes and the significant association ($p<0.05\%$) of *CNTNAP2* gene polymorphisms rs147815978 (G>T) and rs2710102 (A>G) with co-occurrence of intellectual disability and epilepsy phenotypes in our study group which is in HWE ($\chi^2=174$, $P<0.0001$). Logistic regression analysis shows additive ($p<0.0001$) and multiplicative ($p<0.001$) models which confirms significant association of both the polymorphisms in our data, which are closely located on same haplotype (D'=-0.168). We propose that *CNTNAP2* rs 147815978 (G>T) and rs2710102 (A>G) polymorphisms are possible risk loci for overlapping neurodevelopmental disorders in Pakistani population. We propose the role of a previously reported SNP rs2710102 (A>G) with a novel SNP rs147815978 (G>T) for *CNTNAP2* gene association with neurodevelopmental disorders in our data. Our study has expanded the knowledge of *CNTNAP2* gene polymorphisms as probable biomarkers for susceptibility of co-occurrence of intellectual disability and epilepsy phenotypes in Pakistani population. We hope that our study will open new horizons of *CNTNAP2* gene variants research to cure the neurological conditions in Pakistani population where consanguinity is a tradition and prevalence of neurodevelopmental disorders has increased from 1 to 2 percent during last five years.

Session Title: Molecular Effects of Genetic Variation Poster Session I

PB1213 Novel disease associations identified in pediatric restrictive cardiomyopathy by genomic and transcriptomic analysis.

Authors:

M. Padua[1], A. Phatak[1], A. Haaning[1], C. C. Sucharov[2], S. D. Miyamoto[3], Y. Liu[1], S. M. Ware[4]; [1]Indiana Univ. Sch. of Med., Indianapolis, IN, [2]Univ. of Colorado Anschutz Med. Campus, Aurora, CO, [3]Dept. of Pediatrics, Colorado Fetal Care Ctr., Children's Hosp. Colorado, Sch. of Med., Aurora, CO, [4]Indiana Univ., Indianapolis, IN

Restrictive cardiomyopathy (RCM) in children is a rare myocardial disorder characterized by normal chamber size and thickness with impaired ventricular filling due to diastolic dysfunction. It is associated with high morbidity and mortality and there is no effective medical therapy. Pathogenic variants in sarcomeric genes can cause RCM, but the underlying etiology is frequently unknown. To investigate causation and pathogenesis of RCM, we performed exome, mRNA- and small RNA-sequencing (RNA-seq) analyses using explanted left ventricular heart tissue from pediatric RCM transplant subjects (n=14) and compared them to pediatric controls (n=4). From exome data, known pathogenic variants were identified in *BAG3* and *FLNC* as well as novel variants, which were absent from population databases, including a *FLNC* in-frame duplication variant (*c.6527_6547dup*; p.Arg2176_Ser2182dup), and a heterozygous splice acceptor variant in *TNNI3* (rs397516340) ascertained as likely pathogenic per the American College of Medical Genetics and Genomics guidelines. This splice acceptor variant generated an exon-fusion between *TNNI3* exon 7 and the neighbor gene *TNNT1* exon 10, causing a decrease in *TNNI3* expression as assessed by qPCR. Variants of uncertain significance in known cardiomyopathy genes were also identified by exome sequencing. From RNA-seq analysis, 1931 genes were differentially expressed (upregulated n=825; downregulated n=1106). Among them, *MYH6*, which has been previously implicated in other cardiomyopathies, was downregulated in RCM, as well as *HOPX*, a gene involved in cardiac development. In addition, numerous long non-coding RNA (lncRNAs) were differentially expressed between the RCM and control group. *BANCR*, a lncRNA involved in heart development that is highly expressed in other cardiomyopathies, was significantly upregulated in RCM cases versus controls. Other lncRNAs such as *TMC6* (*TNRC6C-AS1*), *CRMA* (*C20orf166-AS1*) and *HAND2-AS1* were also upregulated, suggesting an important role for epigenetic regulation of RCM. Finally, small RNA-seq showed a few microRNA differentially expressed between the two groups. Notably, microRNA-150, which is often downregulated in cardiovascular disease and is associated with heart failure prognosis, was significantly downregulated in RCM. These novel findings on the genetic landscape and its regulation in pediatric RCM provide an opportunity for important investigation and development of therapeutics.

Session Title: Molecular Effects of Genetic Variation Poster Session II

PB1214 Novel *MARVELD2* Frameshift Variant Associated with Post-lingual Autosomal Recessive Non-Syndromic Hearing Impairment

Authors:

E. ABOAGYE[1,2], S. Adadey[3], E. Wonkam-Tingang[1], K. K. Esoh[1], L. Xhanka[1], T. Kubheka[1], L. De Souza Rios[1], L. Amenga-Etego[2], I. Schrauwen[4], S. Mowla[1], G. Awandare[2], S. M. Leal[4], A. Wonkam[5,1]; [1]Univ. of Cape Town, Cape Town, South Africa, [2]Univ. of Ghana, Legon-Accra, Ghana, [3]NIH, Bethesda, MD, [4]Columbia Univ., New York, NY, [5]Dept. of Genetic Med., JHU, Baltimore, MD

Background: Studies of multiplex and simplex congenital hearing impairment (HI) cohorts in Ghana have implicated a *GJB2*-p.Arg143Trp founder variant as the predominant (16% - 42%) cause of the condition. Though most heritable HI cases have monogenic etiology, there is still high genetic and allelic heterogeneity, and many causative variants have been reported in only specific populations. In addition, analysis of known HI genes fails to identify the causative anomaly in 30% of new cases. **Methods**: We analyzed whole - exome sequenced data of a consanguineous multiplex family with variable late age of onset non-syndromic hearing impairment (NSHI), negative for the common *GJB2* pathogenic variant. Bi-directional Sanger validation was performed for genotype-phenotype segregation analysis. *In-silico* variant effect prediction tools were used to investigate the elucidated whole-exome variant pathogenicity. Homology modeling was used to predict the mutant protein structure. The variant was screened in 46 ethnic representative hearing controls and 151 unrelated simplex probands to validate pathogenicity. **Results**: Homozygous MARVEL domain containing 2 (*MARVELD2)*: c.1058dupT-p.Val354SerTer5 frameshift variant on exon 2 was present in all the affected kindreds. The bi-allelic novel *MARVELD2* (*MIM 610572*) variant co-segregated with varied age of onset and degree of post-lingual phenotype. The putative novel variant was absent in the 46 hearing controls and 151 unrelated simplex cohort. The variant was also absent in databases, including heredity hearing loss homepage (HHL), online Mendelian inheritance in man (OMIM), Genome Aggregation Database (gnomAD), dbSNP, ClinVar, and ClinGen. The elucidated *MARVELD2*: c.1058dupT-p.Val354SerfsTer5 thymine duplication was predicted to truncate the encoded protein (C-terminal occluding_ELL domain). We found strong evidence to suggest this variant is associated with the intra-family variability of bilateral autosomal recessive non-syndromic hearing impairment (ARNSHI). Review of the gene identified twelve isolated *MARVELD2* pathogenic variants reports in seven countries, and knockout/knockdown *MARVELD2* gene mice that showed progressive HI. **Conclusion**: We have shown that this novel variant segregate with bilateral ARNSHI, thus expanding pathogenic spectrum of *MARVELD2* and refining HI genotype - phenotype gene curation. Cell-based characterization of mutant MARVELD2 expression, cellular transport, morphology, and tight junction architecture investigations are needed to confirm our prediction on the pathobiology of occludin_ELL domain truncation.

Session Title: Molecular Effects of Genetic Variation Poster Session I

PB1216 Personal transcriptome variation is poorly explained by current genomic deep learning models.

Authors:

R. Shuai, C. Huang, P. Baokar, R. Chung, R. Rastogi, P. Kathail, N. Ioannidis; Univ. of California, Berkeley, Berkeley, CA

Genomic deep learning models can predict genome-wide epigenetic features and gene expression levels directly from DNA sequence. While current models perform well at predicting gene expression levels across genes in different cell types from the reference genome, their ability to explain expression variation between individuals due to cis-regulatory genetic variants remains largely unexplored. Here we evaluate four state-of-the-art models—Enformer, Basenji2, ExPecto, and Xpresso—on paired whole genome sequencing and RNA-sequencing data (n=421) measured on lymphoblastoid cell lines (LCLs) from the Geuvadis Consortium. We focus on the 3,259 genes for which the Geuvadis analysis of expression quantitative trait loci (eQTLs) identified at least one statistically significant genetic association. **First, we validate that these models explain a significant fraction of expression variation across genes using the reference genome** by comparing expression predictions on the reference genome to the median expression level measured for each gene over all individuals in the Geuvadis dataset. Across genes, we find Spearman rank correlations between predicted and observed expression levels of 0.57 for Enformer, 0.52 for Basenji2, 0.54 for ExPecto, and 0.33 for Xpresso. **Next, we show that these models struggle to explain variation in expression across individuals.** We construct personal input sequences for each individual by inserting their single nucleotide variants (SNVs) into the reference sequence around each gene transcription start site (TSS). Using each individual's personal input sequences, we predict expression levels of the aforementioned 3,259 genes and find that although the correlation between predicted and observed expression computed across genes (*cross-gene correlation*) for each individual is similar to the reference genome performance of the corresponding model, the distribution of *cross-individual* correlations for each gene is centered close to zero for all models. We find that although genes with strong eQTLs tend to have larger magnitude cross-individual correlations, the models often predict incorrect directions of genetic effect (negative correlation) for these genes. Furthermore, we find that the four tested models are more consistent with one another in the *magnitude* of their cross-individual correlation for a given gene than in the *direction* of that correlation. **These results suggest that current genomic deep learning models can recognize the presence of causal regulatory variation but struggle with understanding the direction of effect of such variation on gene expression.**

Session Title: Molecular Effects of Genetic Variation Poster Session II

PB1217 Pharmacogenomics in a dish: Lithium and valproic acid molecular response QTLs in human neural progenitor cells

Authors:

B. Le[1], J. M. Valone[2], N. Matoba[3], J. Mory[1], M. Love[4], J. L. Stein[5]; [1]UNC, Durham, NC, [2]UNC Chapel Hill, Raleigh, NC, [3]Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, [4]UNC-Chapel Hill, Chapel Hill, NC, [5]UNC Chapel Hill, Chapel Hill, NC

Lithium (Li) and valproic Acid (VPA) are widely used first-line treatments for BD, but clinical responses are highly variable and mechanisms driving their effects are largely unknown. Fetal VPA exposure can impair neurodevelopment and increase risk for ASD. Genetic influences on clinical responses in patient populations are beginning to be characterized, but are limited by low power, difficulty in isolating specific effects due to the prevalence of polypharmacy, and uncontrolled compliance, duration, and dose across participants. Here we present a "pharmacogenomics in a dish" approach to understand how genetic variation impacts response to these compounds. We describe the effects of Li and VPA on gene regulation, and how these changes are tuned by common genetic variation within primary human neural progenitor cells (hNPCs). Our experiment measured chromatin accessibility via ATAC-seq and gene expression via RNA-seq in 78 genotyped hNPC lines after 48 hour exposure to approximately clinically-relevant concentrations of either 1.5mM LiCl, 1mM VPA, or vehicle. Over 10,000 and 45,000 regulatory elements showed significant changes in chromatin accessibility in response to Li or VPA, respectively (FDR &lt 0.1, |logFoldChange| &gt 0.25). Many transcription factor binding site (TFBS) motifs, such as RFX, were enriched within regulatory regions opened by Li treatment. While Li is known to inhibit GSK3B, a critical component of the canonical Wnt pathway, TCF/Lef TFBS motifs were not enriched in peaks opening or closing due to Li stimulation. We detected 493 and 7,700 differentially expressed genes in response to Li or VPA treatments, respectively (FDR &lt 0.1, |logFoldChange| &gt 0.25). Cellular proliferation genes were upregulated by Li and downregulated by VPA, consistent with proliferation assays conducted in the same samples. We measured the effects of common genetic variants by mapping chromatin accessibility and gene expression quantitative trait loci (caQTLs and eQTLs, respectively). 20,762 and 4239 chromatin accessibility peaks were altered by caQTLs specifically during Li or VPA respectively - an up to 80% increase in discovery compared to the 25,710 caPeaks detected under vehicle conditions. 464 and 447 unique eGenes were regulated by eQTLs specifically during Li or VPA exposure respectively - a 50% increase in discovery compared to the 883 eGenes detected under vehicle conditions (FDR &lt 0.1). These context-specific QTLs reveal novel genetic influences on gene regulation previously undetected by QTL approaches in unstimulated bulk post-mortem tissue and facilitate pharmacogenomic interpretations of Li and VPA treatments.

Session Title: Molecular Effects of Genetic Variation Poster Session III

PB1218 Phenomic environment-wide association study to evaluate the complexity of the exposome

Authors:

N. Palmiero[1], T. Gonzalez Zarzar[2], K. Passero[3], J. McGuigan[4], J. Zhou[5], D. Crawford[6], M. Aguilar[3], A. Ludman[3], M. Hall[3]; [1]Penn State Univ., State College, PA, [2]Pennsylvania State Univ., State College, PA, [3]The Pennsylvania State Univ., State College, PA, [4]Pennsylvania State Univ., University Park, PA, [5]The Pennsylvania State Univ., University Park, PA, [6]Case Western Reserve Univ., Cleveland, OH

Phenome-wide association studies (PheWAS) that identify genomic-based interrelationships between phenotypes have not been applied to exposome research. Here, in a phenomic environment-wide association study (PheEWAS), we interrogated the relationship between the phenome and the exposome. Using National Health and Nutrition Examination Survey (NHANES) data with a discovery subset and a replication subset, we performed linear regression for 326 exposure predictors and 55 phenotypes, adjusting for age, sex, socioeconomic status, body mass index, self-reported race/ethnicity, and data release cycle year. We found associations between alpha-tocopherol (vitamin E) with triglycerides, retinol with triglycerides, gamma-tocopherol with triglycerides, retinol with uric acid, and retinol with blood urea nitrogen. A race/ethnicity-stratified PheEWAS to determine whether exposure-phenotype relationships varied by race/ethnicity identified 106 significant associations and 11 unique race/ethnicity-specific associations, indicating the importance of studying race/ethnicity effects across all groups. These results provide new information on the complexity of the exposome at the level of the phenome.

Session Title: Molecular Effects of Genetic Variation Poster Session I

PB1219 † Polymorphic short tandem repeats contribute to complex traits via multiple mechanisms.

Authors:

J. Margoliash, Y. Li, X. Zhang, A. Goren, M. Gymrek; Univ. of California San Diego, San Diego, CA

Short tandem repeats (STRs), genomic regions each consisting of a sequence of 1-6 base pairs repeated in succession, represent one of the largest sources of human genetic variation. Polymorphism in the number of repeats at thousands of STR loci has been associated with gene regulation. However, many STR effects are not captured well by standard genome-wide association studies (GWAS) or downstream analyses that are mostly based on single nucleotide polymorphisms (SNPs). To study the involvement of STRs in complex traits, we imputed genotypes for 445,720 autosomal STRs into SNP data from 408,153 UK Biobank participants and tested for association with 44 blood and serum biomarker phenotypes. We used two fine-mapping methods, SuSiE and FINEMAP, to identify GWAS signals putatively caused by STRs and estimate that STRs drive 5.2-9.7% of GWAS signals for these traits. By comparison, we demonstrate that STRs account for approximately 7.5% of allelic variation and 18.5% of base pair variation amongst the variants present in our fine-mapping analyses. Overall, we identify 118 high-confidence STR-trait associations that pass all of our fine-mapping tests. We hypothesize that STRs impact complex traits through a variety of mechanisms. The majority of STRs identified are non-coding, and we hypothesized they could play regulatory roles. Using data from GTEx, we find 17 of our set of high-confidence STRs are significantly associated with gene expression after multiple hypothesis correction, and 31 are significantly associated with changes in the DNA methylation level of nearby CpG sites. Additionally, we find two trait-associated repeats in protein-coding regions. Using AlphaFold, we suggest that one of these, a poly-serine repeat in E2F4, influences red blood cell counts by modulating the intermolecular distance of the binding between E2F4 and the protein RBL2. This binding protects E2F4 from ubiquitination and subsequent degradation. As a result, perturbing this binding can effect the stability of E2F4 and hinder E2F4s's role in cell proliferation. Together, our results suggest polymorphic tandem repeats make widespread contributions to complex traits through a variety of mechanisms. Our results provide impetus to further study these and other potential mechanisms to better understand the way STR polymorphism can contribute to complex traits and, in extreme cases, to diseases. We provide a set of stringently selected candidate causal STRs, and we demonstrate the need to routinely consider a more complete view of human genetic variation in GWAS.

Session Title: Molecular Effects of Genetic Variation Poster Session II

PB1220 Population-scale single-cell transcriptomics of clonal haematopoiesis.

Authors:

Z. Qiao[1,2], M. Geaghan[1], A. Hewitt[3], J. Powell[4,5], D. MacArthur[2,6], O. Siggs[1,2]; [1]Garvan Inst. of Med. Res., Sydney, Australia, [2]Ctr. for Population Genomics, Garvan Inst. of Med. Res., Sydney, Australia, [3]Menzies Inst. for Med. Res., Univ. of Tasmania, Hobart, Australia, [4]Garvan-Weizmann Ctr. for Cellular Genomics, Garvan Inst. of Med. Res., Sydney, Australia, [5]UNSW Cellular Genomics Futures Inst., Univ. of New South Wales, Sydney, Australia, [6]Ctr. for Population Genomics, Murdoch Children's Res. Inst., Melbourne, Australia

Clonal haematopoiesis (CH) is an age-related process characterised by the acquisition of somatic mutations in haematopoietic stem cells, leading to their preferential expansion. CH has been implicated in a wide variety of adverse clinical outcomes, including higher risk of haematological cancers and cardiovascular disease, but the mechanisms remain unclear. In order to better understand the biological impact of CH, we investigated the transcriptional consequences of CH in blood on transcriptomes utilising the OneK1K cohort: a population-scale genomic cohort of 977 individuals (mean age 64 years) with orthogonal whole genome sequencing (WGS), genotyping array, and single-cell RNA sequencing (scRNA-seq) datasets derived from peripheral blood. Using both WGS and genotyping array data we generated somatic variant callsets encompassing single-nucleotide, small indel, and large mosaic chromosomal alterations (mCAs), with a detectable somatic driver variant detected in over 11% of individuals. Orthogonal scRNA-seq data was then used to measure the impact of CH-associated somatic variants and genes on relative proportions of blood leukocytes, as well as their transcriptional impact within specific cell types. Our study offers new insight into the transcriptional impact of CH-associated genes and variants, with implications for their therapeutic manipulation.

Session Title: Molecular Effects of Genetic Variation Poster Session III

PB1221 Post-zygotic rescue of meiotic errors causes brain mosaicism and focal epilepsy

Authors:

T. Bedrosian[1], K. Miller[1], A. Rivaldi[1], N. Shinagawa[1], J. Navarro[1], S. Sran[1], J. Westfall[1], A. Miller[2], R. Roberts[2], Y. Akkari[1], R. Supinger[2], M. Hester[3], M. Marhabaie[1], M. Gade[4], J. Lu[5], O. Rodziyevska[6], M. Bhattacharjee[6], G. Von Allmen[6], E. Yang[7], H. Lidov[7], C. Harini[7], M. Shah[6], J. Leonard[1], J. Pindrik[1], A. Shaikhouni[1], J. Goldman[5], C. Pierson[2], D. Thomas[1], D. Boue[1], A. Ostendorf[2], E. Mardis[2], A. Poduri[7], D. Koboldt[3], E. Heinzen[4]; [1]Nationwide Children s Hosp., Columbus, OH, [2]Nationwide Children's Hosp., Columbus, OH, [3]Nationwide Childrens Hosp., Columbus, OH, [4]UNC Chapel Hill, Chapel Hill, NC, [5]Columbia Univ., Bronx, NY, [6]McGovern Med. Sch., Houston, TX, [7]Boston Children's Hosp., Boston, MA

Somatic mosaicism is a known cause of neurological disorders, including developmental brain malformations and epilepsy. Brain mosaicism is traditionally attributed to post-zygotic genetic alterations arising in fetal development. Here we describe post-zygotic rescue of meiotic errors as an alternate origin of brain mosaicism in patients with focal epilepsy who have mosaic chromosome 1q copy number gains. Genomic analysis showed evidence of an extra parentally-derived chromosome 1q allele in the resected brain tissue from 5 of 6 patients. This copy number gain is observed only in patient brain tissue, but not in blood or buccal cells, and is strongly enriched in astrocytes. Astrocytes carrying chromosome 1q gains exhibit distinct gene expression signatures and hyaline inclusions, supporting a novel genetic association for astrocytic inclusions in epilepsy. Further, these data demonstrate an alternate mechanism of brain chromosomal mosaicism, with parentally derived copy number gain isolated to brain, reflecting rescue in other tissues during development.

Session Title: Molecular Effects of Genetic Variation Poster Session I

PB1222 Predicting genetically regulated gene expression on the X chromosome

Authors:

X. Zhang[1], J. Below[2], A. Naj[3], B. Kunkle[4], E. Martin[5], W. Bush[1,6]; [1]Case Western Reserve Univ, Cleveland, OH, [2]Vanderbilt Univ Med Ctr., Nashville, TN, [3]Univ. of Pennsylvania, Philadelphia, PA, [4]Univ. of Miami, Miami, FL, [5]Univ. of Miami Miller Sch. of Med., Miami, FL, [6]Cleveland Inst. for Computational Biology, Cleveland, OH

Despite the potential importance of genetic variation on the X chromosome, it is often omitted in disease association studies. The exclusion of the X chromosome has also propagated into the post-GWAS era, as transcriptome-wide association studies (TWAS) also ignore the X due to the lack of adequate models of X chromosome gene expression. In this work, we trained elastic net penalized models in the brain cortex and whole blood using whole genome sequencing (WGS) and RNA-seq data. To make generalizable recommendations, we evaluated multiple modeling strategies in a homogeneous study population of 175 whole blood samples for 600 genes, and 126 brain cortex samples for 766 genes. SNPs (MAF>0.05) within the gene's two megabase flanking window were used to train the tissue-specific model of each gene. We tuned the shrinkage parameter and evaluated the model performance with nested cross-validation. Across different mixing parameters, sample sex, and tissue types, we trained 511 significant gene models in total, predicting the expression of 229 genes (98 genes in whole blood and 144 genes in brain cortex). The average model coefficient of determination (R^2) was 0.11 (range from 0.03 to 0.34). We tested a range of mixing parameters (0.05, 0.25, 0.5, 0.75, 0.95) for the elastic net regularization, and compared the sex-stratified and sex-combined modeling on the X chromosome. We further investigated genes escaping X chromosome inactivation to determine if their genetic regulation patterns were distinct. Based on our findings, sex-stratified elastic net models with a balanced penalty (50% LASSO and 50% ridge) are the most optimal approach to predict the expression levels of X chromosome genes, regardless of X chromosome inactivation status. The predictive capacity of the optimal models in whole blood and brain cortex was confirmed through validation using DGN and MayoRNAseq temporal cortex cohort data. The R^2 of tissue-specific prediction models ranges from 9.94×10^{-5} to 0.091. These models can be used in Transcriptome-wide Association Studies (TWAS) to identify putative causal X chromosome genes by integrating genotype, imputed gene expression, and phenotype information.

Session Title: Molecular Effects of Genetic Variation Poster Session II

PB1223 Predicting mode of action of missense variants by graph representation of protein structural context

Authors:

G. Zhong, Y. Shen; Columbia Univ., New York, NY

Accurate prediction of functional impact of missense variants is fundamentally important for disease gene discovery, clinical genetic analysis, and protein engineering. Previous efforts have focused on predicting a binary pathogenicity classification, but the functional impact of missense variants is multi-dimensional. Different pathogenic missense variants in the same gene may act through different modes of action, such as gain or loss of functions or partial loss on one of the functions, and ultimately increase the risk of different conditions or diseases. The mechanisms of the functional impact depend on the functions of the encoded protein, therefore, predicting mode of action requires components that are tuned to the functions specific to each protein or protein families. In this work, we developed a new method, PreMode, that learns universal representation of coding sequence variants and used transfer learning to predict modes of action in each protein or protein family. PreMode learns a general representation of sequence variation within protein structural context using SE(3)-equivariant graph neural networks on AlphaFold2 predicted protein structures. We curated the largest set of mode-of-action-labeled missense variants (1772 Gain-of-function, 7426 Loss-of-function) from human genetics data and deep mutational scanning experiments. We first pre-train PreMode on large labeled pathogenic/benign variant databases, then performed transfer learning for an array of mode of action prediction tasks. PreMode outperformed previously published methods in most of the tasks, especially in ion channel genes with an area under the curve of 0.92. We applied PreMode to published de novo mutation data from individuals with neurodevelopmental disorders and found several high confidence gain-of-function variants. Furthermore, we showed the PreMode framework is not limited to individual missense variants in human proteins but can also model the combined effects of multiple variants in engineered proteins.

Session Title: Molecular Effects of Genetic Variation Poster Session III

PB1224 Predicting the effects of genetic variants on chromatin accessibility with a deep learning approach

Authors:

J. Jaureguy; UCSD, Salk Institute for Biological Studies, CA

Genome-wide association studies (GWAS) have identified thousands of genetic variants associated with human traits and diseases. However, the majority of these variants lie within the non-coding portion of the genome and, we currently lack knowledge of the molecular function of the associated variants and of which of them are causal. Molecular quantitative trait locus (QTL) analysis can help determine variant function, by identifying variants associated with molecular traits such as chromatin accessibility (caQTLs) and gene expression (eQTLs). However, due to linkage disequilibrium, the identity of causal variants remains ambiguous. Furthermore both genome-wide association and QTL studies lack power to detect associations with rare genetic variants. Machine learning methods have recently emerged as an alternative method for determining variant function. These methods predict gene expression, chromatin accessibility, and transcription factor binding from DNA sequence alone. However, currently, machine learning models are trained on reference genome sequences and do not consider human genetic variation in training. We hypothesized that including genetic variation and allele-specific open chromatin (ASOC) in machine learning models could increase their ability to predict variant function. Here, we introduce a "variant aware" model that is able to predict the impact of both common and rare genetic variants on chromatin accessibility. We utilized transfer learning to incorporate genetic variation and allele-specific information into the training of a machine learning model. For training and testing my model, I selected a human macrophage dataset containing ATAC-seq data and genotype information for 23 unique donors under four treatment types. The variant aware model yielded a Spearman rank correlation (rho) of rho=0.29 for allelic imbalance predictions and rho=0.62 for ATAC-seq read depth predictions, which exceeded the performance of reference-trained models DeepSea (rho=0.26 for allelic imbalance; rho=0.43 for ATAC-seq read depth), and Basset (rho=0.11 for allelic imbalance; rho=0.34 ATAC-seq read depth). In summary, our machine learning model that leverages genetic variation and allele-specific information, more accurately predicts the effects of common genetic variants on chromatin accessibility, and we are in the process of predicting the effects of rare variants. These results indicate that variant aware machine learning models have great potential in improving predictions of common and rare genetic variants on molecular traits, especially rare genetic variants that cannot be assessed by QTL-based methods.

Session Title: Molecular Effects of Genetic Variation Poster Session I

PB1225 Predicting the functional impact of compound heterozygous genotypes from a large scale variant effect map

Authors:

G. Cromie[1], M. Xie[1], K. Owens[2], M. Timour[1], M. Tang[1], J. N. Kutz[2], A. El-Hattab[3], R. McLaughlin[1], A. Dudley[1]; [1]Pacific Northwest Res. Inst., Seattle, WA, [2]Univ. of Washington, Seattle, WA, [3]Coll. of Med., Univ. of Sharjah, Sharjah, United Arab Emirates

Reduced activity of the enzymes encoded by *PHGDH*, *PSAT1*, and *PSPH* causes a set of ultrarare, autosomal recessive diseases known as serine biosynthesis defects. These diseases present in a broad phenotypic spectrum, with Neu-Laxova syndrome at the severe end, infantile serine biosynthesis defects, with severe neurological manifestations and growth deficiency, in the intermediate range, and in the form of childhood disease with intellectual disability at the mild end. However, L-serine supplementation, especially if started early, can ameliorate and in some cases even prevent symptoms. Therefore, knowledge of pathogenic variants can improve clinical outcomes. On this basis, we utilized a functional growth-based assay in haploid yeast cells to measure the impact of 1,914 SNV-accessible amino acid substitutions in PSAT. Our results agree well with existing clinical variant interpretations, evolutionary data and protein structure-function relationships, supporting the use of our data as functional evidence under the ACMG interpretation guidelines. Using existing variants with clinical classifications enabled us to define deleterious, non-deleterious and uncertain ranges of our growth assay and to then place 1634 (85%) of the tested amino acid substitutions in the informative deleterious or non-deleterious categories. In addition to assaying the functional impact of individual variants in yeast haploid cells, we also experimentally assayed pairwise combinations of *PSAT1* alleles that recapitulate human genotypes, including compound heterozygotes, in yeast diploids. Results from our diploid assay successfully distinguish the genotypes of affected individuals from those of healthy carriers and agree well with disease severity. Finally, we present a linear model that uses our individual allele measurements (from haploid cells) to accurately predict the behavior of allele pairs from our diploid assay. This model then allowed us to leverage our 1,914 individual allele measurements to predict the biallelic function of ~1.8 million allele combinations, each corresponding to a potential human *PSAT1* genotype. Taken together, our work provides an example of how large-scale functional assays in model systems can be powerfully applied to the study of ultrarare diseases.

Session Title: Molecular Effects of Genetic Variation Poster Session II

PB1226 Predicting the pathogenicity of missense variants in AMPA receptors

Authors:

L. Montanucci[1], T. Brünger[2], N. Bhattarai[1], D. Lal[3]; [1]Cleveland Clinic, Cleveland, OH, [2]Cologne Ctr. for Genomics, Brühl, Germany, [3]Harvard Med. Sch. / Broad Inst., Boston, MA

Ionotropic glutamate receptors (iGluRs) are ligand-gated ion channels activated by the neurotransmitter glutamate. They mediate the majority of excitatory synaptic transmission throughout the central nervous system and are key players in synaptic plasticity, which underlie critical brain functions like learning and memory. iGLluRs comprise four sub-classes of receptors: NMDA receptors (encoded by the GRIN genes), AMPA receptors (encoded by the GRIA genes), kainate receptors (encoded by the GRIK genes) and delta receptors (encoded by the GRID genes).Genetic variants in the genes encoding for iGluRs are associated with severe neurological and neuropsychiatric disorders. Most clinically characterized genetic variants in iGluRs belong to the class of NMDA receptors, with about 400 clinically characterized variants from patients and 1,500 control variants from general population. Variants in AMPA receptors, instead, are characterized to a much lesser extent with only ~30 variants for which the clinical significance has been annotated and ~100 known control variants for general population. Therefore, computational methods able to predict the pathogenicity in AMPA receptors are crucial for genetic interpretation in clinical practice and inform precision therapies for patients with AMPA receptor variants.Taking advantage of the big number of well-characterized variants in NMDA receptors, we previously developed a pathogenicity predictor, specifically trained on NMDA receptors variants and based on variant proximity to ligands. We showed that proximity of the variant to ligands is key feature for predicting pathogenicity, outperforming existent predictors. Here, our goal is to transfer the predictive knowledge we gathered from well-characterized GRIN genes, to GRIA genes, by exploiting the high structural conservation between NMDA and AMPA receptors. We therefore adapted the pathogenicity predictor previously developed for NMDA receptor missense variants, to predict the pathogenicity of AMPA receptors missense variants.

Session Title: Molecular Effects of Genetic Variation Poster Session III

PB1227 Prioritizing coronary artery disease risk variants in atherosclerosis using deep learning models of chromatin accessibility in mouse

Authors:

S. Kundu[1], R. Wirka[2], D. Li[1], J. Monteiro[1], D. Sharma[1], L. Sundaram[1], T. Quertermous[1], A. Kundaje[1]; [1]Stanford Univ., Stanford, CA, [2]Univ. of North Carolina, Chapel Hill, NC

Arterial smooth muscle cells (SMCs) respond to atherosclerotic lesions in coronary artery disease (CAD) by de-differentiating, proliferating, and migrating in a process known as phenotypic modulation. Recent work has revealed that these SMCs transform into fibroblast-like cells, known as fibromyocytes, which play a protective role by forming fibrous caps that cover these lesions and prevent adverse events, such as arterial thrombosis and myocardial infarction. Furthermore, genes at loci implicated in CAD risk through genome-wide association studies (GWAS), such as TCF21, ZEB2, and SMAD3, play a crucial role in the formation of these protective SMC-derived fibromyocytes. Although the transcriptomic changes underlying this phenotypic modulation of SMCs have recently been studied, the gene regulatory programs driving these transcriptomic changes remain poorly understood. Similarly, the causal variants modulating disease risk at hundreds of CAD GWAS loci remain unidentified.
In order to characterize the changes in the regulatory landscape during phenotypic modulation, we profiled the gene expression and chromatin accessibility of lineage-traced SMCs in aortic atherosclerotic tissue from mice at single cell resolution. We found a trajectory of cell states from quiescent SMCs to fibromyocytes and chondromyocytes, which are similar to endochondral bone-forming cells and are found at the end of the phenotypic modulation trajectory. To identify the transcription factors (TFs) regulating the genes expressed along this trajectory, we trained convolutional neural network models that can map regulatory DNA sequence to base resolution chromatin accessibility profiles in each cell state. Using model interpretation tools, we annotated the putative binding sites for each of these TFs along the trajectory and linked them to their target genes.
Finally, we used these cell-state specific models trained on mouse data to score the functional impact of all non-coding human variants associated with CAD risk. Strikingly, we found more than 500 variants with significant effects on predicted accessibility, spanning the entire phenotypic modulation trajectory. These variants disrupt motifs of key transcription factors, such as SRF, AP-1, MEF2, and ZEB, and have concordant scores from models trained on human cells with matching phenotype, highlighting the power of this approach to identify functional variants across species. Experimental validation of these prioritized variant effects in human coronary artery smooth muscle cells is underway.

Session Title: Molecular Effects of Genetic Variation Poster Session I

PB1228 Promoter region functional polymorphism (30bpuVNTR) of monoamine oxidase A gene association with serum protein levels in psychiatric disorders

Authors:

A. Hashmi[1], R. Qamar[2], M. Azam[3]; [1]Comsats Univ. islamabad, Islamabad, Pakistan, [2]Sci. and Technology Sector, ICESCO, Rabat, Morocco, [3]COMSATS Univ. Islamabad, Islamabad, Pakistan

Background: Monoamine oxidase A (MAOA) is a mitochondrial outer membrane enzyme that has a role in the metabolism of different biogenic amines, including dopamine, serotonin and norepinephrine. The functional studies of the *MAOA* have revealed the effect of the variable number of tandem repeat sequences (VNTR) in the promoter region on the transcriptional activity of the *MAOA* that consequently affects the homeostasis of the biogenic amines which may be implicated in the aetiology of multiple psychiatric conditions and mood disorders. **Objective:** The current study was aimed to determine the effect of the promoter region 30 base pair (bp) VNTR of the *MAOA* (*MAOA*-30bp µVNTR) on its serum levels and association with major depressive disorder (MDD), bipolar disorder (BD) and schizophrenic (SHZ) patients of the Pakistani population. **Methodology:** A total of 1062 subjects [MDD *n*=416; BD *n*=200; SHZ *n*=97; and controls *n*=349], were genotyped for *MAOA*-30bp µVNTR through standard polymerase chain reaction (PCR) technique. Univariant followed by multivariant logistic regression analysis was applied to determine the genetic association. Serum MAOA levels were determined through enzyme-linked immunosorbent assay (ELISA) and the Mann-Whitney U test was applied. **Results:** Eight different repeats (R) alleles; 2R, 3R, 3.5R, 4R, 4.5R, 5R, 5.5R and 6R of 30bp *MAOA*-µVNTR were observed where 4.5R, 5.5R and 6R were the rare repeat alleles, found in the current Pakistani cohort with no significant association with the studied psychiatric conditions. However, compared to controls the total MAOA serum levels were found significantly elevated in SHZ (P=0.004). In sub-group analysis cases and controls were grouped with respect to alleles; low-activity alleles (MAOA-LA=2R, 3R, 5R), high-activity alleles (MAOA-HA=3.5R, 4R) and rare alleles (MAOA-R=4.5R, 5.5R, 6R). Significant elevated MAOA serum levels were observed in higher activity (CON-HA) and lower activity (CON-LA) alleles compared to rare alleles (CON-R), (P<0.03), within the control group. Whereas, in case-control analysis, significantly higher serum levels of MAOA were observed only in the rare allele groups of BD-R (P=0.008) and SHZ-R (P=0.05). **Conclusion:** This is the first report from the Pakistani population, giving us further insights into the complex nature of *MAOA* regulation and its association with different psychiatric conditions. Though the study focused only on the 30bp µVNTR element, the potential interactive contribution of other functional polymorphisms and the gene/environment cannot be ruled out to influence the expression of *MAOA* and its serum-based levels in different ethnicities.

Session Title: Molecular Effects of Genetic Variation Poster Session II

PB1229 Promoter Usage Quantitative Trait Loci in Chronic Obstructive Pulmonary Disease

Authors:

A. Saferali, Z. Xu, R. Chase, M. Cho, P. Castaldi, C. Hersh; Brigham & Women's Hosp., Boston, MA

Rationale: Transcriptional quantitative trait locus study results can be integrated with genome wide association study (GWAS) data to provide insight into biological mechanisms involved in disease. We have previously identified multiple COPD GWAS loci which can be explained by splice QTLS (sQTLs), 3'UTR length QTLs (apaQTLs), and eQTLs, however, there remain several loci with unexplained functions. Here, we investigate SNPs that are associated with promoter activity (proQTLs) to identify novel functions for GWAS variants in chronic obstructive pulmonary disease (COPD). **Methods:** RNA sequencing data and whole genome sequencing data from lung tissue of 1241 subjects from the Lung Tissue Research Consortium (LTRC) was obtained through the NHLBI TOPMed project. ProActiv was used to quantify promoter activity (the amount of transcription initiated at each promoter), by quantifying the number of junctional reads aligning to the first set of introns belonging to a transcript corresponding to a given promoter. COPD and lung function associated SNPs for colocalization analysis were ascertained from a published GWAS (Sakornsakolpat et al., Nat Genet 2019;51:3). **Results:** We identified 23,007 active promoters from 16,337 unique genes. Activity of 7,328 promoters from 6,365 genes was associated with at least one SNP (proQTLs). ProQTLs were most commonly associated with the first promoter located in the gene (63.7%), while 17.8% were associated with the second. For the 7,086 unique lead proQTL-variants, the median distance from the transcript promoter was 19 bases upstream. In relation to the gene body, 53.7% of lead variants were intronic, followed by intergenic (27.1%), exonic (6.3%), and 5'UTR variants (3.8%). We found that 6,114 proQTL lead variants (86%) were also eQTL-variants at FDR<0.05 in LTRC lung tissue. Colocalization analysis revealed that out of the 82 loci associated with COPD, thirteen colocalized with proQTL data with a posterior probability > 0.8. Of these, the proQTL with the highest effect size was the association between rs11580609 with promoter activity of *Peroxiredoxin 1 (PRDCX1*) ($p=1.1x10^{-157,}$, $\beta=-1.7$). Specifically, rs11580609-G, which is associated with decreased risk of COPD ($p=1.24x10^{-7}$, $\beta=-0.05$) and increased lung function (FEV1/FVC; $p=6.1x10^{-11}$, Z=-6.5), is also associated with increased activity of a promoter located at GRCh38 chr1:45523047 (mean activity CC=3.6, GG=1.5). **Conclusions:** We identified 7,086 unique variants associated with promoter activity. A total of 13 COPD GWAS loci contain evidence of proQTLs, suggesting that analysis of promoter activity in lung tissue can provide novel insights into disease mechanisms.

Session Title: Molecular Effects of Genetic Variation Poster Session III

PB1230 Quantitative evaluation of DNA damage and repair dynamics as predictors of clinical phenotypes in individuals with germline *PTEN* variants.

Authors:

R. Wei[1,2], M. Hitomi[1,3], T. Sadler[1], L. Yehia[1], J. Scott[3,2,4,5], C. Eng[1,5,6,7,8]; [1]Genomic Med. Inst., Lerner Res. Inst., Cleveland Clinic, Cleveland, OH, [2]Dept. of Nutrition and Systems Biology and Bioinformatics Program, Case Western Reserve Univ. Sch. of Med., Cleveland, OH, [3]Translational Hematology & Oncology Res., Cleveland Clinic, Cleveland, OH, [4]Dept. of Radiation Oncology, Cleveland Clinic, Cleveland, OH, [5]Taussig Cancer Inst., Cleveland Clinic, Cleveland, OH, [6]Ctr. for Personalized Genetic Hlth.care, Med. Specialties Inst., Cleveland Clinic, Cleveland, OH, [7]Dept. of Genetics and Genome Sci., Case Western Reserve Univ. Sch. of Med., Cleveland, OH, [8]Germline High Risk Cancer Focus Group, Case Comprehensive Cancer Ctr., Case Western Reserve Univ., Cleveland, OH

Persons with germline variants in the tumor suppressor gene phosphatase and tensin homolog, *PTEN*, have PTEN hamartoma tumor syndrome (PHTS). PHTS confers a high risk of malignancies, but up to 23% are diagnosed with autism spectrum disorder (ASD) and/or developmental delay (DD). Importantly, the accurate prediction of these two disparate phenotypes (cancer vs. ASD/DD) for PHTS at the individual versus cohort level remains elusive. Notably, PTEN is a guardian of genome integrity. However, integrative studies that quantify the DNA damage response (DDR) in individuals with PHTS in relation to phenotypes and genotypes have not been explored. In this study, we used 43 lymphoblastoid cell lines (LCLs) derived from individuals with PHTS to investigate the associations between DDR and *PTEN* genotypes and/or clinical phenotypes ASD/DD vs. cancer. We used the exponential decay model, which describes signal decrease at a rate proportional to its current value, to fit our quantified time-course γH2AX immunofluorescence signals induced by γ-irradiation of the LCLs. We found that *PTEN* nonsense variants (n=18) are associated with less DNA damage repairing ability/higher DNA damage levels after repair than *PTEN* missense variants (n=17, p=0.01). More specifically, PTEN p.Arg335Ter (n=5) showed higher DNA damage levels after repair than PTEN p.Arg173Cys (n=8, p=0.01). As related to PHTS phenotypes, those with neither ASD/DD nor cancer (n=8) showed higher DNA damage repairing half-life than those diagnosed with ASD/DD alone (n=17, p=0.049). We also applied the reaction-diffusion partial differential equation (PDE) mathematical model, a tumor cell growth model with a DNA damage term, to accurately describe the DDR process in the LCLs. The numerical results of the PDE reflecting DDR were visualized as heatmaps for each LCL, including DNA damage repair and cell growth capacity. By calculating the peak signal-to-noise ratio (PSNR) and structural similarity index measure (SSIM) for each heatmap, we found that LCLs with different *PTEN* genotypes have different DDR signatures. In order to better understand *PTEN* genotype-associated cellular phenotype and organismal phenotype-associated cellular phenotype, we used an integrated experimental and theoretical modeling approach to derive a signature of growth and repair. This integrative analysis of DDR dynamics will enable more precise prediction of clinical phenotypes (ASD/DD vs. cancer) at the individual level, thus enabling more precise medical management of individuals with PHTS.

Session Title: Molecular Effects of Genetic Variation Poster Session I

PB1231 Quantitative trait loci mapping of circulating metabolites in cerebrospinal fluid reveals insights into biological mechanisms

Authors:

T. Boltz[1], L. Reus[2], M. Francia[1], M. Bot[1], N. Ramesh[1], W. M van der Flier[3], P. Visser[2], S. van der Lee[3], B. Tijms[2], L. Olde Loohuis[1], C. Teunissen[3], R. Ophoff[1]; [1]UCLA, Los Angeles, CA, [2]Alzheimer Ctr. Amsterdam, Amsterdam, Netherlands, [3]VU Univ. Med. Ctr., Amsterdam, Netherlands

The use of functional phenotypes within genome-wide association contexts has provided mechanistic insights into complex disease architectures, though these insights lag behind particularly for neuropsychiatric diseases. In this study, we use metabolomics to measure the levels of 5,543 CSF metabolite levels, both targeted and untargeted, in 977 Dutch individuals with cerebrospinal fluid (CSF) and genetic data. Individuals originated from two separate cohorts including the Utrecht cohort (n=490 cognitively healthy subjects) and the Amsterdam Dementia Cohort (n=487 subjects from a well-characterized memory clinic cohort). We performed genome-wide metabolite quantitative trait loci (mQTL) mapping on CSF metabolomics and found 126 Bonferroni-significant mQTLs, representing 65 unique CSF metabolite levels across 52 independent loci. By imputing the CSF levels of the significant metabolites into the GWAS of various brain-related diseases, we identified 40 significant metabolite-trait associations. Similarly, by imputing brain-derived gene expression into our CSF metabolomics data, we find colocalized gene-metabolite associations for over 90% of our genome-wide significant mQTL. These findings highlight new metabolic pathways that may be involved in the dysregulation of neurodegenerative and psychiatric disorders.

Session Title: Molecular Effects of Genetic Variation Poster Session II

PB1232 Rare variant annotation tools: a practical guide for researchers.

Authors:

A. Ziegler, C. Riccio; MedizinCampus Davos, Davos, Switzerland

Due to the declining costs of sequencing, the feasibility of whole-genome sequencing on a large scale has increased, enabling rare variant association studies (RVAS) involving thousands of subjects. In contrast to common variant genome-wide association studies (GWAS), RVAS hold the promise of identifying high-impact variants. Nonetheless, the interpretation of RVAS findings poses a complex challenge because of the many different types of variations and the abundance of annotation tools. This presentation provides a practical guide to selecting the most appropriate tools for rare variant annotation. A systematic search was done in PubMed to identify tools that predict the effects of genetic variants. Tools that are disease-specific, cannot be installed locally, or were not recently updated were excluded from further consideration. The tools were evaluated based on criteria such as types of variants covered, predicted consequences, documentation, and redundancy with other databases. A total of 72 tools were identified and included in this review. These tools varied substantially in terms of annotation coverage, ranging from tools focusing on one functional element, e.g., enhancers to general variant annotation tools. The types of annotations provided include functional consequences in coding and non-coding regions, population frequencies, conservation scores, and pathogenicity scores. We assigned each tool with Sequence Ontology (SO) terms corresponding to the variants and effects covered by the tool. We provide a file that can be easily queried by researchers to select tools covering specific types of variants. Some tools offer focused annotation for specific variant types or molecular mechanisms, and general variant annotation tools provide broader coverage across variant types and consequences. Integration of multiple tools can enhance the accuracy of variant effect prediction, thus aiding the understanding of the biological basis of rare variant/disease associations. Researchers can use this guide to efficiently select appropriate tools for rare variant interpretation, which allows them to create experimentally testable hypotheses.

Session Title: Molecular Effects of Genetic Variation Poster Session III

PB1233 Rare variant association analyses of plasma protein levels in a multi-ancestry cohort.

Authors:

C. Sarnowski[1], D. Dutta[2], J. Zhang[3], J. Ma[1], E. Boerwinkle[1], J. Coresh[3], B. Yu[1], N. Chatterjee[3]; [1]Univ. of Texas Hlth.Sci. Ctr. at Houston, Houston, TX, [2]Natl. Cancer Inst., Bethesda, MD, [3]Johns Hopkins Bloomberg Sch. of Publ. Hlth., Baltimore, MD

Background: Majority of protein quantitative trait loci (pQTL) studies to date have focused on local associations with common genetic variants (minor allele frequency [MAF] > 1%) primarily in participants of European descent. The effect of rare variants on both proximal (cis) and distal (trans) proteins could have notable impact on the genetic architecture of the proteome. Aim: We analyzed whole exome sequencing data from 5,281 European American (EA) and 1,797 African American (AA) participants in the Atherosclerosis Risk in Communities (ARIC) Study to evaluate the contribution of rare variants on 4,955 plasma proteins levels measured by SomaScan at visit 2. Methods: We evaluated gene-level association of rare (MAF ≤ 1%) loss of function (lof) variants with protein levels using variant-set mixed model association tests (SMMAT) or likelihood-based tests for quantitative traits (LRT-q), that model the effect of rare variants on either the mean or the overall distribution of protein levels respectively. We conducted analyses separately in EA and AA participants. Cis versus trans associations were defined based on a 5MB region around each protein transcription starting site. Gene results were filtered based on a cumulative minor allele count of at least 25. Threshold to define significance accounted for multiple testing ($P_{trans} \leq 6.5\times10^{-10}$; $P_{cis} \leq 9.9\times10^{-8}$). Results: The mean heritability of a protein attributed to cis-lof variants was slightly higher in AA (mean = 1.9% [0.01% - 7.9%]) compared to EA (mean = 1.3% [0.01% - 4.8%]). SMMAT detected 307 (EA) and 93 (AA) significant protein-gene associations. LRT-q identified large majority of the SMMAT signals (92.8% in EA & 79.5% in AA) along with large number of additional signals (1,076 in EA & 277 in AA). Across both methods, majority of the associations detected were in trans (76.1% in EA & 77.2% in AA). We detected with both methods a cis association of rare variants annotated to *C6* with C6 protein in both ancestries and trans-associations in EA participants of SPOCK1 protein (chr5) with rare variants annotated to *CRIPAK*-chr4 and *SLC17A9*-chr20. We also identified trans-associations in both ancestries of four proteins (PTGR2-chr14 [SMMAT], FARS2-chr6 [LRT-q], and VBP1-chrX & SUMF2-chr7 [LRT-q & SMMAT]) with rare variants annotated to *C6*-chr5. Cross-ancestry replication (P < 5% with same direction of effects) was observed for 32% (EA) and 25% (AA) of the significant genes, driven by MAF differences across ancestries. Conclusion: Our ancestrally diverse study can provide important insights into the overall contribution of rare variants in the genetic regulation of plasma proteins. Funding: R01HL148218, R01HG010480

Session Title: Molecular Effects of Genetic Variation Poster Session I

PB1234 Rare variant dosage in sudden cardiac death-related genes increases risk of sudden cardiac death among hypertrophic cardiomyopathy patients.

Authors:

Q. Huang[1], C. Zhang[1], M. Lu[1], X. Bo[2], J. Wang[1], L. Song[1]; [1]State Key Lab. of Cardiovascular Disease, Fuwai Hosp., Natl. Ctr. for Cardiovascular D, Beijing, China, [2]Inst. of Hlth.Service and Transfusion Med., Beijing, China

Introduction: Incorporating genetic testing to refine the risk stratification of hypertrophic cardiomyopathy (HCM) patients for sudden cardiac death (SCD) remains essential and challenging. We evaluated the dosage effect of rare deleterious variants on the SCD outcome in HCM patients.
Methods and results: A total of 836 unrelated HCM patients were prospectively recruited with a median follow-up time of 5.20 years and 25 patients reaching SCD. Participants with available and quality control qualified whole exome sequencing were screened for rare deleterious variants in 43 curated SCD-related genes based on the disease they caused (inherited cardiomyopathies and hereditary arrythmias) and their encoded proteins' function (sarcomere, desmosome, ion channels). Rare variant dosage in SCD-related genes was a significant predictor of SCD (HR, 3.5; 95% CI, 1.2-10.1, P = 0.02). After adjusting for age, gender and clinical SCD risk factors recommended by 2020 American Heart Association (full model), rare variant dosage showed independent association with SCD (adjusted HR, 4.4; 95% CI, 1.1-18.6, P = 0.03). The C-index for full model is 0.80 (95% CI 0.670-0.890, P < 0.001).
Conclusion: Our study defined an SCD-related gene set and demonstrated in a large HCM cohort that rare variant dosage in this gene set is an independent predictor of SCD and may be integrated into future SCD prediction model for HCM patients.

Session Title: Molecular Effects of Genetic Variation Poster Session II

PB1235 Rare variants affecting telomere length and disease identified through multi-omic modeling

Authors:

R. Keener[1], T. Li[2], F. Aguet[3], K. Ardlie[4], J. Rotter[5], S. Rich[6], A. Battle[1]; [1]Johns Hopkins Univ., Baltimore, MD, [2]Johns Hopkins Sch. of Med., Baltimore, MD, [3]Illumina, Inc., Foster City, CA, [4]Broad Inst., Cambridge, MA, [5]Lundquist Inst., Harbor-UCLA Med Ctr, Torrance, CA, [6]Univ. of Virginia, Scottsville, VA

Telomeres protect the ends of linear chromosomes and as humans age, telomere length (TL) decreases. When telomeres become critically short, a senescence or apoptosis signal prevents further telomere loss. Individuals with extremely short TL present with Short Telomere Syndromes (STS), including bone marrow failure and immunodeficiency, while individuals with extremely long TL are predisposed to cancer. To gain insight into TL genetic regulation, prior work from our group and others used genome-wide association studies to examine the role of common genetic variation in TL. This strategy identified novel genes involved in TL regulation, some of which we experimentally validated. However, this approach ignores the effects of rare variation, which can have larger effect sizes and uniquely impact genes under strong constraint.

Studies of rare variant effects on TL have improved our understanding of TL biology, but have largely required laborious STS patient pedigree studies. We leveraged TL estimates and rare variant data from the Trans-Omics for Precision Medicine (TOPMed) Program to broadly examine the impact of rare variation on TL. Previously we developed Watershed, a Bayesian hierarchical model, which uses whole genome sequencing with paired multi-omic data (expression, splicing, methylation, and/or protein levels) to prioritize rare variants causing significant disruption of molecular phenotypes. This multi-omic signature generates interpretable hypotheses for coding and non-coding rare variants, providing a posterior probability that the variant causes outlier status for each molecular signal, for example that splicing is disrupted but expression is not. We used data from 5,310 MESA individuals to train Watershed and observed that in 40/86 individuals with extremely short TL (<1% in TOPMed), Watershed prioritized rare variants in at least one gene from a panel of 16 STS genes. The variant with the largest posterior probability (0.984) was predicted to affect expression of *TPP1*, which encodes a protein critical for TL regulation. We will expand our analysis to another 103,812 TOPMed individuals and incorporate multi-omic data where available. Examination of highly weighted variants in individuals with extreme TL relative to average TL will potentially identify novel genes involved in TL regulation. In addition, we will examine the interplay between TL regulation and multi-omic signals over age (0-98 years old). Finally, we will apply our model to data from STS patients to improve their genetic diagnosis. Together this work has utility in improving STS patient diagnosis and furthering our understanding of the molecular mechanisms governing TL.

Session Title: Molecular Effects of Genetic Variation Poster Session III

PB1236 Rare*Plasmodium falciparum*coronin gene mutations following ACT treatment of malaria in South WesternNigeria.

Authors:

O. Ajibaye[1,2], Y. A. Olukosi[1], E. Oriero[2], M. Oboh[3], B. Iwalokun[4], I. Nwankwo[5], O. V. Adaramoye[6], S. Chukwuemeka[1], A. Ngwa[7]; [1]Nigerian Inst. of Med. Res., Lagos, Nigeria, [2]Med. Res. Council, The Gambia, Banjul, Gambia, [3]Med. Res. Council, The Gambia, ?Banjul, Gambia, [4]Nigerian Insttute oof. Med. Res., Lagos, Nigeria, [5]Ctr. for Molecular Parasitology, Dept. of Microbiol. and Immunology, Drexel Univ. Coll. of Med.,, Philadelphia, PA, [6]Coll. of Med., Univ. of Lagos, Lagos, Nigeria, [7]Nigerian Inst. of Med. Res., Lagos, Gambia

Non-Pfkelch13 parasite protein variants have been implicated in artemisinin (ART) resistance, but not in Africa. Genetic markers underlying in vivo reduced ART efficacy among African P. falciparum populations are currently unclear. We investigated SNPs in Plasmodium falciparum actin-binding protein (Pfcoronin) and host *cyp2b6* associated with in vivo ART tolerance in Nigeria.Seven isolates showing parasitaemia after Day 3 in a 28-day therapeutic efficacy study of artemether-lumefantrine among 51 volunteers in Lagos, Nigeria were investigated. Molecular diagnosis was done by microscopy, conventional and real-time PCR amplification of Pf18S rRNA gene, var acidic terminal sequence, telomere-associated repetitive elements-2 and coupled conventional and real-time Pf18S rRNA PCR. Twelve neutral P. falciparum microsatellite loci genotyping were analyzed to confirm recrudescence in comparison with msp2 genotyping. We genotyped drug resistance targets (DHFR_51, DHFR_59, DHFR_108, DHFR_164, MDR1_86, MDR1_184, DHPS_581 and DHPS_613), *cyp2b6* and sequenced PfCoronin bi-directionally for presence and association of mutations with ART tolerance.Molecular techniques employed detected P. falciparum infections. One infecttion was recrudescent by microsatellites analysis out of the four identified as recrudescent infections by msp2 genotyping. Presence of the drug resistance-associated haplotypes, pfdhfr/pfdhps/pfmdr1 ($108_{T/N}/51_{I}/164_{L}/59_{R}/581_{G}/86_{Y}/184_{F}$) was observed in two samples. H_e, allelic diversity, for each microsatellite locus from pre- and post-drug administration, revealed no significant difference in the mean H_e ($P = 0.19$, Mann-Whitney test). Significant LD ($I_A^S = 0.2865$, $P=0.02$, Monte Carlo simulation) around the neutral microsatellite loci was observed. Three new Pfcoronin SNPs (V55L, L77I and D154Y) were found. and *cyp2b6* SNP implicated SNPs here reported may guide investigations on mechanisms of emerging African ART resistance.

Session Title: Molecular Effects of Genetic Variation Poster Session I

PB1237 Recurrent and non-recurrent copy number variants in American Indian tribes with substance use disorders

Authors:

S. Majid, C. Hodgkinson, P-H. SHEN, C. Marietta, D. Sambo, D. GOLDMAN; NIAAA/NIH, ROCKVILLE, MD

Copy Number Variants (CNVs) are large structural variants that can alter function by several mechanisms, including gene deletion and duplication, and thereby may alter susceptibility to Substance Use Disorders (SUDs). CNVs have been implicated in several neuropsychiatric diseases including autism, schizophrenia, and Alzheimer's disease. The rarity or de novo nature of most CNVs impedes linkage to phenotypes; however even rare CNVs may be abundant in families and founder populations. Therefore, in this study, we identified recurrent CNVs (rCNVs) in American Indian tribes that are relatively low in genetic admixture but have high prevalence of SUD. CNVs were identified via Illumina SNP arrays in 400 Plains Indians and 320 Southwest Indians. All had been psychiatrically diagnosed via structured interviews (SADS-LA). Uniquely, we identified large (>200 kb) rCNVs that were abundant in both of two geographically separated American Indian tribes. A CNV at chromosome 6p21.33 was present in more than one in eight individuals in one of the populations and in more than one in four in the other, and multiple homozygotes were observed for that deletion. Haplotype and deletion breakpoint analyses showed that the recurrent CNVs derived from a common ancestor. Most individuals carried at least one rCNV, and CNV load ranged from 0 to 5. These rCNVs, including deletions and duplications encompassing various genes that may contribute to development of SUDs, are being evaluated for effect on transcriptome and linkage to behavior.

Session Title: Molecular Effects of Genetic Variation Poster Session II

PB1238 Regulatory interactions of autism-associated neuronal enhancers implicate novel genes in autism etiology.

Authors:

B. Mifsud[1], D. Aljogol[1], S. Hussain[1], M. Kohailan[2], S. Salim[1], K. Fakhro[3], F. Mohammad[1]; [1]Hamad Bin Khalifa Univ., Doha, Qatar, [2]Sidra Med., Doha, Qatar, [3]Sidra Med. and Res. Ctr., Doha, Qatar, Qatar

Autism spectrum disorders (ASD) are highly heterogeneous neurodevelopmental disorders that are usually associated with secondary phenotypes including intellectual disability, posttraumatic stress disorder (PTSD), anxiety and epilepsy. Over 100 risk genes have been identified, yet they only explain a fraction of the cases. The majority of sporadic ASD are attributed to the presence of *de novo* variants, most of which are in the non-coding portion of the genome. Here, we analyze non-coding single-nucleotide autosomal *de novo* variants in 4,864 ASD cases and 3,865 controls to pinpoint regulatory elements contributing to the etiology of autism. We identified 18 ASD-associated *cis*-regulatory elements that interact with 95 genes, four of which (*EXOC3, EEF1A2, PXDN and CHRNB3)* have been previously reported. Thirty-eight known transcription factor binding sites (TFBSs) were altered at these *cis*-regulatory elements in at least three cases, including motifs for the ASD-associated THRA, EGR2, NR1D1 and VEZF1. Lost TFBSs were enriched for transcription factors (TFs) involved in nervous system development and immunity, whereas gained TFBSs were enriched for those involved in response to sensory cues, including sensory organ development. Pan-neuronal knockdown of two genes during development, one novel (*AHRR*) and one ASD-associated gene (*SLIT3*), resulted in abnormal response to stimulus, learning, social interactions and anxiety in *Drosophila*. In conclusion, *cis*-regulatory elements contribute to autism etiology. Incorporating chromosome conformation capture technologies is a more accurate approach to define target genes, as they recognize those that are in spatial proximity irrespective of their linear genomic position.

Session Title: Molecular Effects of Genetic Variation Poster Session III

PB1239 † Response eQTLs in primary human chondrocytes identify novel putative osteoarthritis risk genes

Authors:

N. Kramer[1], E. Thulson[1], P. Coryell[1], S. D'Costa[1], K. L. Mohlke[1], B. O. Diekman[1,2], R. F. Loeser[1], D. H. Phanstiel[1]; [1]UNC, Chapel Hill, NC, [2]North Carolina State Univ., Raleigh, NC

Osteoarthritis (OA) is a complex degenerative disease of the joint, and genome-wide association studies (GWAS) have already identified over one hundred genetic loci associated with OA risk, but the putative genes and genetic mechanisms underlying OA are still unknown. Here we generate and use a well powered chondrocyte genomic dataset derived from the cartilage of 102 healthy donors in both a resting state and stimulated with fibronectin fragment (FN-f)-a known OA trigger-to characterize condition and sex-associated gene expression changes and to identify novel response expression QTLs (eQTLs). Condition-separated modeling of standard cis eQTLs identified 1517 significant control eGenes (FDR $p<0.05$, containing at least 1 eQTL) and 1482 significant FN-f eGenes, with 667 eGenes specific to control samples, 632 eGenes specific to FN-f samples, and 850 eGenes found in both conditions. To capture OA stimulus-specific eQTLs that may reflect conditionally regulated mechanisms, we specifically identified response eQTLs by testing for the interaction effect of donor genotype with condition either before or after FN-f treatment. Through this analysis, we identified 270 eGenes specific to resting chondrocytes and 353 eGenes specific to chondrocytes treated with FN-f. Notable FN-f response SNP-eGene pairs include rs12901081-G associated with higher SMAD3 expression (log2 allelic fold change 1.24), and rs8011143-C associated with lower ABHD4 expression (log2 allelic fold change -1.79). 177 genes differentially regulated by FN-f were also significant FN-f response eGenes, including the differentially downregulated response eGene DIO2, which has been cited as a potential OA susceptibility gene. Initial colocalization of FN-f response signals with various OA phenotype GWAS with coloc ($PP4>0.7$) after LD filtering ($r2 > 0.5$) revealed a robust colocalization of the response eQTL signal for SMAD3 with 7 OA GWAS phenotypes, prioritizing SMAD3 as a potential functional gene associated with OA progression. Sex-specific differential expression identified 259 genes in untreated samples and 84 genes after FN-f treatment, 48 of which were significant in both conditions (FDR $p<0.01$). One notable X-chromosome gene was KDM6A, which showed increased expression in females in both conditions and has previous associations with OA development. Taken together, the FN-f model system provides a controlled context to probe the stimulus-specific mechanisms of OA progression and discover novel molecular trait loci that are not detected as standard eQTLs, which will further prioritize candidate genes for future study and translation into OA treatments.

Session Title: Molecular Effects of Genetic Variation Poster Session I

PB1240 Revealing the Extensive Allelic Heterogeneity and Impact of Mobile Elements Across 64 Diverse Human Haplotypes

Authors:

W. Zhou[1], B. Rodriguez-Martin[2], P. Balachandran[3], M. Loftus[4], P. A. Audano[5], Y. Gan[6], The Human Structural Variation Consortium, S. E. Devine[7], J. O. Korbel[8], R. E. Mills[9], M. K. Konkel[10], C. R. Beck[11]; [1]Univ. of Michigan Med. Sch., Ann Arbor, MI, [2]European Molecular Biology Lab., Heidelberg, Germany, [3]The Jackson Lab., Bar Harbor, ME, [4]Clemson Univ., Clemson, SC, [5]The Jackson Lab., Farmington, CT, [6]Harvard Univ., Cambridge, MA, [7]Univ. of Maryland Sch. of Med., Baltimore, MD, [8]European Molec Biology Lab, Heidelberg, Germany, [9]Univ Michigan Med Sch, Ann Arbor, MI, [10]Clemson Univ, Clemson, SC, [11]The Jackson Lab. and UConn Hlth., Farmington, CT

Mobile element insertions (MEIs) diversify human genomes through retrotransposition, polymorphism, and recombination. Polymorphic MEIs contribute to both inter- and intra-individual genetic variation, leading to ongoing mutagenesis. In this study, we conducted a thorough analysis of MEIs in 64 phased haplotypes from sequence-resolved assemblies representing individuals from diverse populations, enabling an exploration into highly repetitive regions of the genomes that pose challenges for traditional sequencing methodologies. We identified 9,600 non-redundant polymorphic MEIs, including 1,183 LINE-1s, 7,838 *Alu* elements, 534 SVAs, 24 HERV-Ks, and 21 processed pseudogenes comprising 7.4 Mbp of insertion sequences. 64% of these insertions fell within existing repetitive regions, where we observed nested 'repeats in repeats', i.e. *Alu* in *Alu* and LINE-1 in LINE-1, with these insertions concentrated within A-rich reference MEI regions. Additionally, we observed 2,639 alleles lacking a reference MEI.

Our 64 haplotype assemblies enabled the interrogation of allelic heterogeneity associated with each sequence-resolved MEI locus. For full-length LINE-1 sequences, we identified 198 active loci where the extent of allelic heterogeneity positively correlated with insertion allele frequency, age, and subfamily status. Notably, allelic differences frequently impact the coding and mobilization potential for LINE-1s, which includes hot (i.e. highly active) elements in cancer. For instance, we observed a subset of haplotypes with intact ORFs for a hot LINE-1 at 9q32, while other haplotypes, including the reference genome, harbored a common protein-truncating single nucleotide polymorphism at ORF2p. We further observed that the greatest variability in SVA length was due to allelic variation within the hexameric and variable number of tandem repeat regions.

Two homologous MEIs can act as substrates of ectopic DNA repair leading to different types of structural variants. We identified 1,833 deletions, 223 duplications, and 63 inversions mediated by two distinct mobile elements that affect nearly 12 Mbp; *Alu* elements drive 80% of these rearrangements. Breakpoint analysis of the mobile element-mediated rearrangements revealed that almost 75% were driven by homologous recombination.

In summary, our results provide a comprehensive understanding of mobile element-derived variation between genetically diverse individuals, highlighting extensive differences in transposon content, allelic heterogeneity, mobile element-driven rearrangements, and begin to elucidate the effects of these variants on human genomes.

Session Title: Molecular Effects of Genetic Variation Poster Session II

PB1241 Role of genetic and environmental variation on gene expression in whole blood from ethnically diverse African populations

Authors:

C. Zhang[1], M. Hansen[1], Y. Feng[1], W. Beggs[1], S. Mpoloka[2], G. Mokone[3], T. Nyambo[4], G. Belay[5], C. Fokunang[6], A. K. Njamnshi[7], S. A. Tishkoff[1,8,9]; [1]Dept. of Genetics, Perelman Sch. of Med., Univ. of Pennsylvania, Philadelphia, PA, [2]Biological Sci., Univ. of Botswana, Gaborone, Botswana, [3]Faculty of Med., Univ. of Botswana, Gaborone, Botswana, [4]Dept. of Biochemistry, Kampala Intl. Univ. in Tanzania, Dar es Salaam, Tanzania, United Republic of, [5]Dept. of Microbial Cellular and Molecular Biology, Addis Ababa Univ., Addis Ababa, Ethiopia, [6]Dept. of Pharmacotoxicology and Pharmacokinetics, Faculty of Med. and BioMed. Sci., The Univ. of Yaoundé I, Yaoundé, Cameroon, [7]Dept. of Neurology, Central Hosp. Yaoundé, Yaoundé, Cameroon, [8]Dept. of Biology, Univ. of Pennsylvania, Philadelphia, PA, [9]Ctr. for Global Genomics and Hlth.Equity, Univ. of Pennsylvania, Philadelphia, PA

We generated whole blood transcriptomes using RNA-seq from 517 individuals from 19 ethnically and geographically diverse African populations from Ethiopia, Tanzania, Botswana, and Cameroon. These populations have diverse ancestries, speak different languages (Niger-Congo, Afroasiatic, Nilosaharan, and Khoesan), and have diverse subsistence practices (e.g., hunting-gathering, pastoralism and farming). Furthermore, they inhabit extreme environments such as rainforests and high altitudes, and some experience a high burden of communicable or non-communicable diseases. We compare expression profiles and identify numerous differentially expressed genes (DEGs) among African ancestries, some of which are associated with ancestry specific traits. For instance, in rainforest hunter-gatherers from Cameroon who have shorter stature, DEGs are enriched for intrauterine growth retardation and bone development loci, and in the Fulani, who have high prevalence of diabetes in adults, DEGs are enriched for metabolic processes (e.g., insulin signaling pathway). We show that the rainforest environment, independent of genetic ancestry, is significantly associated with differential expression of approximately 700 genes, many of which play a role in immune function (e.g. genes related to antigen presentation and inflammation response). We find a substantial number of RNA-Seq reads in rainforest populations that are mapped to parasite genomes, suggesting a high burden of pathogens may shape the differential expression profiles within these populations. We identify several key transcription factors mediating expression changes in response to varying parasite loads. By integrating transcriptomic and WGS data, we fine-map cis-regulatory variants, many of which are novel. We further identify ancestry-biased eQTLs, some of which are common only in certain populations. We colocalize regulatory variants with GWAS associations for 34 human traits/diseases identified in both African and European populations. We identify novel eQTL-disease-colocalizations in African GWAS (e.g. an eQTL at *P4HA1*, which influences collagen synthesis, colocalized with an African height GWAS) that could not be colocalized with GTEx data. In summary, we generated an atlas of African transcriptomic diversity and identified eQTLs which may influence differences in phenotypes or disease susceptibility among ethnically diverse populations. Our study sheds light on the understanding of molecular links of genetic variants to human traits/diseases and advances precision medicine in African populations and globally. Funded by ADA 1-19-VSN-02, NIH 1R35GM134957, R01AR076241.

Session Title: Molecular Effects of Genetic Variation Poster Session III

PB1242 Saturation mutagenesis data facilitate the interpretation of non-coding variants in an *IRF6* enhancer associated with nonsyndromic cleft-lip w/o cleft palate.

Authors:

R. Hollstein[1], F. Thieme[1], F. Lauber[1], A. Biedermann[1], J. Fazaal[1], M. Schubach[2], M. Kircher[3,2], K. Ludwig[1]; [1]Inst. of Human Genetics, Univ. of Bonn, Bonn, Germany, [2]Berlin Inst. of Hlth.at Charité – Univ.smedizin Berlin, Berlin, Germany, [3]Inst. of Human Genetics, Univ. Hosp. Schleswig Holstein, Lübeck, Germany

Nonsyndromic cleft lip with/without cleft palate (nsCL/P) is one of the most common birth defects. Its etiology is multifactorial with a strong genetic component. Several genes have been implicated in nsCL/P, e.g. because of their location near GWAS loci or by rare variants found in monogenic forms. However, these only partially explain its high heritability. The contribution of rare, noncoding variants has been rarely investigated, largely due to limitations in attributing pathogenic effects.

One of the best studied genes involved in nsCL/P is *IRF6*. Common and rare variants within the gene itself and an upstream enhancer (*MCS9.7*) are associated with different forms of orofacial clefting, including nsCL/P. Here, we combined two sequencing-based high-throughput methods to screen the *IRF6* enhancer *MCS9.7* for candidate variants across the entire allelic spectrum, including rare variants. We used molecular inversion probes (MIPs) to sequence 699 nsCL/P cases and 511 population controls of European descent. In total, we found 11 variants in cases: three common single nucleotide variants (SNV), two 4- and 9-bp indels, and six rare SNVs.

In parallel, we performed a massively parallel reporter assay (MPRA) to facilitate variant interpretation. We transfected HaCaT, GMSM-K and HEKT293T cells with a previously established saturation mutagenesis library of *MCS9.7* to analyse the effects of all possible single nucleotide exchanges within the *MCS9.7* sequence (362 bp). Fold-changes and p-values were calculated for each substitution, and these results were then integrated with the variants observed in nsCL/P cases.

One common SNV (rs76145088) and one rare variant (rs189972759) showed a significant decrease of enhancer activity that was specific to HaCaT cells ($0.05 > p$). Interestingly, the variant with the highest effect size (nominally significant, $0.05 > p > 1e\text{-}5$) in HaCaT and GMSM-K cells was recently associated with cleft palate only in a finnish cohort. Overall, five of six rare variants affected the enhancer activity at least in either HaCat or GMSM-K at nominal significance threshold. The two common variants had no effect ($p > 0.05$). In conclusion, our data suggest that the combination of MPRA data together with large-scale identification of individual variants might be a promising tool for the prioritization of causal non-coding variants. We have demonstrated this for orofacial clefting, but this approach might also be applicable to other human disorders with a suspected contribution of rare non-coding variants.

Session Title: Molecular Effects of Genetic Variation Poster Session I

PB1243 Scalable count-based models improve power and robustness to detect eQTLs in large-scale single-cell data

Authors:

Z. Zhang[1], E. Kawaguchi[1], S. Gazal[2], N. Mancuso[3]; [1]Univ. of Southern California, Los Angeles, CA, [2]USC, Los Angeles, CA, [3]Univ. of Southern California, South Pasadena, CA

Expression quantitative trait (eQTL) analyses have identified numerous variants associated with gene expression levels from bulk mRNA; however, integrative efforts with genome-wide association studies (GWAS) explain little observed risk. This limited overlap can be partially explained by bulk eQTL reflecting a mixture of cell type-specific eQTLs. To this end, single-cell RNA-seq data when combined with genetics enables identifying cell type-specific eQTLs and improves our understanding of GWAS mechanisms. Current eQTL mapping approaches rely heavily on continuous assumptions after data normalization, which is not optimal for single-cell gene expression count data. While generalized linear models (GLM) provide a path towards modeling count data, they are computationally demanding, and do not scale to the millions of tests required for cell type-specific cis-eQTL testing.

To address this limitation, we propose jaxQTL, a flexible framework for single-cell eQTL mapping using highly scalable GLMs. To scale inference, jaxQTL leverages five optimizations. First, we use conjugate gradient at each step, which leverages efficient sparse matrix-vector numerical routines. Second, we fit separately a covariate-only "null model" to initialize variant-specific models and improve convergence. Third, we "short-circuit" inference if the variant effect-size does not deviate sufficiently from the null after a small number of iterations. Fourth, we employ a Beta-approximation of the null distribution to reduce the number of permutations necessary to compute empirical p-values. Lastly, we implemented our approach using the Python framework JAX, which provides an end-to-end differentiable framework, makes use of "just-in-time" compilation to leverage highly optimized Fortran/C++ numerical libraries, and enables GPU or TPU execution without any additional efforts from the user.

To benchmark jaxQTL, we performed cis-eQTL mapping for 135 genes with known cis-eQTLs in natural killer cells from N=981 individuals in Onek1k single-cell RNA-seq data. Using a Poisson model, jaxQTL has improved statistical power compared with the linear model for identifying 135 previously reported cis-eQTL (P=1.54E-4), while maintaining calibrated P-values under permutations. Importantly, jaxQTL is ~10x faster compared with a GLM implemented in off-the-shelf software, thus substantially reducing overhead required for computing cis-eQTL scans in single-cell data.

Taken together, jaxQTL provides a robust and scalable approach to identifying cis-eQTLs in single-cell data.

Session Title: Molecular Effects of Genetic Variation Poster Session II

PB1244 sc-eQTGen: large scale federated cell-type specific eQTL mapping in PBMCs

Authors:

M. Bonder[1], D. Kaptijn[1], M. Korshevniuk[1], R. Oelen[2], sc-eQTLGen consortium, L. Franke[3], M. van der wijst[1]; [1]Univ. Med. Ctr. Groningen, Groningen, Netherlands, [2]UMCG, Groningen, Netherlands, [3]Univ. Med. Ctr. Groningen, Groningen, Groningen, Netherlands

Background/Objectives: Over the last decade, single-cell RNA sequencing (scRNA-seq) has grown in popularity as a potent alternative to bulk RNA-seq, as scale, cost and sensitivity have significantly improved. These developments have now paved the way for the generation of large numbers of population-based scRNA-seq datasets that are a valuable resource for studying the effect of genetic variation (through expression quantitative trait locus (eQTL) mapping) on expression of individual cell types. To fully leverage these emerging data resources, we have founded the single-cell eQTLGen consortium (sc-eQTLGen).
Methods: We have developed a pipeline that harmonizes the preprocessing, QC, cell type annotation and eQTL mapping of peripheral blood mononuclear cell (PBMC) scRNA-seq data. Every cohort in the consortium has run these pipelines, and shared the non-privacy sensitive summary statistics, which we combine in a federated manner.
Results: Here, we present the results of the first data freeze in which we have meta-analyzed scRNA-seq-derived eQTL summary statistics of over 2,000 individuals for each of the major cell types in PBMCs. These results reveal the cell types in which eQTLs are to be identified. We show the cell type specificity of cis-eQTLs in PBMCs and the colocalization of cell type specific eQTL signals with disease.
Conclusion: Upcoming data freezes will further increase the sample size, cell type resolution, and number of donor characteristics under study, and will expand our analyses to other data modalities. Thereby, we expect that the sc-eQTLGen framework enables us and the community at large to gain a more complete understanding of gene expression and its regulation in health and disease.

Session Title: Molecular Effects of Genetic Variation Poster Session III

PB1245 ScMetaBrain: Federated single-cell consortium for cell-type specific eQTL analysis of neurological disease variants.

Authors:

H-J. Westra[1], M. Vochteloo[1], R. Oelen[1], D. Neavin[2], R. Warmerdam[1], U. Vosa[3], M. Korshevniuk[1], D. Kaptijn[1], M. G. P. van der Wijst[1], M. Bonder[1], scEQTLgen Consortium, T. Esko[3], Y. Huang[4], H. Runz[4], D. Malhotra[4], J. Bryois[5], E. Tsai[4], L. Franke[1]; [1]Univ. Med. Ctr. Groningen, Groningen, Netherlands, [2]Garvan-Weizmann Ctr. for Cellular Genomics, Darlinghurst, Australia, [3]Univ. of Tartu, Tartu, Estonia, [4]Biogen Inc., Cambridge, MA, [5]Roche Innovation Ctr., Basel, Switzerland

Recently we have completed large scale eQTL meta-analyses in blood (eQTLGen; Vosa *et al*. 2021) and brain (MetaBrain; de Klein *et al*. 2023), providing insight into the downstream effects of disease-associated genetic risk factors. Although having substantial sample sizes, these studies lack the cell type context that single-cell RNA sequencing (scRNA-seq) can provide. To pinpoint these cellular contexts and better interpret neurological diseases, we have initiated the scMetaBrain study, which aims to enable a federated scRNA-seq eQTL analysis in human brain.

To harmonize and compare with our single cell eQTL efforts in blood, we have adapted our recently developed pipeline (scEQTLGen, van der Wijst *et al*. 2020) to brain. This includes robust containerized pipelines that perform quality control, demultiplexing, cell type classification, and eQTL analysis per cell-type and dataset. By applying a meta-analysis of the summary statistics, we enable easy inclusion of new datasets without the need of sharing person-identifiable data.

As proof of concept, we have applied our pipeline to automatically analyze 35 samples from Mathys *et al*. 2019. Using the Bakken *et al*. 2021 dataset as reference, our approach automatically annotates five major brain cell types with high accuracy (97% true positive compared to manual assignment). Moreover, we are able to replicate the *cis*-eQTLs observed in a large brain single-cell eQTL study (n=192; Bryois *et al*. 2022) with high agreement (allelic concordance >0.89, R_b > 0.86 for excitatory neuron and oligodendrocytes), highlighting the validity of our methodology.

Here we introduce scMetaBrain, in which we aim to setup a single-cell brain consortium for the identification of downstream consequences of trait-related genetic variants in specific brain cell types. We envision that this consortium will enable a unique opportunity to disentangle tissue and cell type specific regulatory effects in the brain.

Session Title: Molecular Effects of Genetic Variation Poster Session I

PB1246 Severe respiratory phenotype in *KCNH1*-related Temple Baraitser Syndrome

Authors:

E. Conner[1,2], I. Glass[1,2], C. Thompson[3], A. George[3], M. Chen[1,2], N. Natarajan[1,2], A. Bouldin[4], A. Scott[2], T. Wenger[1,2]; [1]Univ. of Washington, Seattle, WA, [2]Seattle Children's, Seattle, WA, [3]Northwestern Univ., Chicago, IL, [4]Swedish Med. Ctr., Seattle, WA

We describe two children with severe neurologic and respiratory presentations who were heterozygous for the same novel pathogenic activating variant in *KCNH1*. A 2-month old female infant was born by vaginal delivery (face presentation) at term following an uncomplicated pregnancy with normal birth indices. At delivery she had respiratory distress, central hypotonia and abnormal dystonic limb movements. Examination noted anonychia and hypoplasia of bilateral thumbs and halluces with the remaining digits having hypoplastic nails. Hyperekplexia was present and was partially responsive to clonazepam. Neurologic evaluation consisted of normal MRI, and EEGs without seizure activity. Nasoduodenal feeding was required due to aspiration. Rapid exome sequencing revealed a pathogenic variant in *KCNH1* (c.1478 C>T, p.T493I), not been previously reported. Though initially she had rare, self-resolving desaturations, clinical observation of her abnormal breathing pattern including tachypnea and apneic pauses prompted a polysomnogram. Her sleep study revealed severe sleep disordered breathing with apnea-hypopnea index of 70/hour (normal <15/hr) due to mixed (but predominantly central) apneas. Due to subsequent aspiration event and subsequently required frequent suctioning as well as high flow nasal cannula to treat desaturations and hypercapnia. She remains hospitalized awaiting repeat sleep study to determine the need for tracheostomy/ventilatory support. While pursuing potential targeted therapies, we learned that functional studies were in process for the T493I variant because of an additional previously unreported child who also had severe neurologic features and died at age 9 of respiratory failure. Functional studies of the variant supported gain of function. Specifically, the variant causes a large hyperpolarizing shift in the voltage-dependence of channel activation. Respiratory failure is an uncommon feature in *KCNH1*-related disorders. The combination of abnormal neurologic features and sleep study findings suggest the abnormal respiratory function is neurologically mediated. We hypothesize the T493I gain-of-function variant may be associated with more severe neurologic outcomes including neurologically-mediated respiratory failure. Moreover, it highlights the importance of pursuing polysomnography for clinically observed respiratory abnormalities rather than relying on oxygen desaturations alone in infants with *KCHN1*-related disorders.

Session Title: Molecular Effects of Genetic Variation Poster Session II

PB1247 Significant enhancement of eQTL and disease trait findings achieved by analysis of 2,344 adipose samples.

Authors:

S. Brotman[1], J. S. El-Sayed Moustafa[2], L. Guan[3], K. A. Broadaway[1], D. Wang[2], A. U. Jackson[4], R. Welch[4], K. W. Currin[1], M. Tomlinson[2], S. Vadlamudi[1], H. M. Stringham[4], T. A. Lakka[5,6,7], A. Oravilahti[8], L. Fernandes Silva[8], N. Narisu[9], M. R. Erdos[9], T. Yan[9], C. K. Raulerson[1], S. C. Parker[3,10], J. Kuusisto[8,11], P. Pajukanta[12], J. Tuomilehto[13,14,15], F. S. Collins[9], M. Boehnke[4], M. I. Love[1,16], H. A. Koistinen[13,17,18], M. Laakso[8,11], K. L. Mohlke[1], K. S. Small[2], L. J. Scott[4]; [1]Dept. of Genetics, Univ. of North Carolina, Chapel Hill, NC, [2]Dept. of Twin Res. and Genetic Epidemiology, King's Coll. London, London, United Kingdom, [3]Dept. of Computational Med. & Bioinformatics, Univ. of Michigan, Ann Arbor, MI, [4]Dept. of Biostatistics and Ctr. for Statistical Genetics, Sch. of Publ. Hlth., Univ. of Michigan, Ann Arbor, MI, [5]Inst. of Biomedicine, Sch. of Med., Univ. of Eastern Finland, Kuopio, Finland, [6]Dept. of Clinical Physiology and Nuclear Med., Kuopio Univ. Hosp., Kuopio, Finland, [7]Fndn. for Res. in Hlth.Exercise and Nutrition, Kuopio Res. Inst. of Exercise Med., Kuopio, Finland, [8]Inst. of Clinical Med., Kuopio Univ. Hosp., Univ. of Eastern Finland, Kuopio, Finland, [9]Natl. Human Genome Res. Inst., NIH, Bethesda, MD, [10]Dept. of Human Genetics, Univ. of Michigan, Ann Arbor, MI, [11]Dept. of Med., Kuopio Univ. Hosp., Kuopio, Finland, [12]Dept. of Human Genetics and Inst. for Precision Hlth., David Geffen Sch. of Med. at UCLA, Los Angeles, CA, [13]Univ. of Helsinki and Dept. of Med., Helsinki Univ. Hosp., Helsinki, Finland, [14]Dept. of Publ. Hlth., Univ. of Helsinki, Helsinki, Finland, [15]Diabetes Res. Group, King Abdulaziz Univ., Jeddah, Saudi Arabia, [16]Dept. of Biostatistics, Univ. of North Carolina, Chapel Hill, NC, [17]Dept. of Publ. Hlth.and Welfare, Finnish Inst. for Hlth.and Welfare, Helsinki, Finland, [18]Minerva Fndn. Inst. for Med. Res., Helsinki, Finland

Most eQTL studies have relatively modest sample sizes and focus on identifying one primary eQTL per gene, both of which limit power to detect the full range of regulatory variants. We expanded eQTL discovery for subcutaneous adipose tissue by analyzing 2,344 samples from five studies of primarily European ancestry. eQTL meta-analysis detected 34K conditionally distinct eQTL signals in 18K genes, which is 2.3-fold more signals and >1.6-fold more eQTL genes than detected by any of the five studies alone. Over half of eQTL genes had at least two eQTL signals. Compared to primary signals, non-primary signals had lower effect sizes and minor allele frequencies and corresponded to genes with higher heritability and lower tolerance for loss of function. Adipose eQTL variants for 1st-4th signals were enriched in adipose promoters and enhancers (P<3e-3), and 2nd, 3rd, and 4th signals showed successively less enrichment, particularly for promoters, indicating signals accounting for more gene expression variance are enriched in promoters. We compared adipose primary eQTL lead variants to blood eQTL lead variants from the eQTLGen study (n=31,684) and found that 45% of adipose primary eQTL were not detected as blood eQTL (P≤1e-6; r^2<0.2), highlighting the value of using diverse tissues. Next, we colocalized conditionally distinct eQTL and GWAS signals for 28 cardiometabolic disease (CMD) traits. We identified 3,605 colocalized GWAS-eQTL signal pairs for 1,861 eQTL genes (*coloc* PP4≥0.5; GWAS-eQTL r^2≥0.5). Compared to primary eQTL signals, non-primary signals identified 46% more eQTL-GWAS colocalizations and 488 (36%) additional candidate genes, suggesting that non-primary eQTL signals may explain some of the missing regulation of GWAS loci. Non-primary eQTL signals were slightly less likely than primary signals to colocalize with GWAS signals, even when accounting for eQTL strength. To determine if gene expression mediates variant effects on CMD traits, we used MRLocus to examine 26 genes with at least 2 nearly independent (r^2≤0.05; D'≤0.5) colocalized GWAS-eQTL signal pairs; 11 showed a consistent gene dosage effect on the GWAS trait (adjusted P≤0.1). Finally, 70% of GWAS-eQTL colocalized signals had a lead or proxy variant (r^2 ≥ 0.8) overlapping open chromatin in adipose tissue or adipocyte cells, indicating putative functional variants. In summary, this adipose eQTL analysis tripled the size of previous studies, furthered understanding of allelic heterogeneity in gene regulation, greatly expanded discovery of eQTL colocalized with CMD GWAS signals, and identified thousands of candidate genes that may point to new drug targets.

Session Title: Molecular Effects of Genetic Variation Poster Session III

PB1248 Single Cell Multiomics Profiling of Visceral Adipose Tissue Reveals New Insights into the Genetic Regulation of Cardiometabolic Disease

Authors:

U. Arasu[1], I. Selvarajan[1], T. ÖRd[1], M. Taipale[1], A. Toropainen[1], M. Sukhatme[2], A. Kar[3], S. Lee[4], M. Alvarez[4], D. Kaminska[1,4], V. Männistö[1], J. Pihlajamäki[1,5], P. Pajukanta[6,7], M. Kaikkonen[1]; [1]Univ. of Eastern Finland, Kuopio, Finland, [2]UCLA, Pasadena, CA, [3]UCLA, Saratoga, CA, [4]UCLA, Los Angeles, CA, [5]Kuopio Univ. Hosp., Kuopio, Finland, [6]Univ California Los Angeles, Pacific Palisades, CA, [7]Inst. for Precision Hlth., Los Angeles, CA

We conducted the first single cell multiomics (RNA and ATAC) profiling of human visceral adipose tissue (VAT) obtained from seven individuals with obesity in the Kuopio Obesity Surgery (KOBS) cohort. After extensive quality control, 3200 nuclei showed clustering into five main cell types: adipocytes, adipose stem/progenitor cells (ASPC), endothelial cells, mesothelial cells, and macrophages. We performed a comprehensive analysis of the identified cell types to investigate their associated transcription factors using SCENIC. This analysis not only confirmed known drivers of cell identity but also revealed novel transcription factors that play a crucial role in shaping the identity of these cells.Building upon the established knowledge that common genetic variants associated with cardiometabolic diseases often reside in non-coding regions of the genome, we employed Linkage Disequilibrium Score Regression (LDSC) to examine the enrichment of cardiometabolic trait -associated SNPs within 1 kb of the peaks in adipocytes (50k peaks), ASPCs (63k), mesothelial cells (75k) and macrophages (42k). As an example of our preliminary findings, we demonstrate that adipocyte, ASPC, and macrophage peaks account for a significant proportion of per-SNP heritability for Non-Alcoholic Fatty Liver Disease (non-permuted p-values=0.004, 0.017, 0.048) and triglyceride levels (TG; non-permuted p-values=0.025, 0.031, 0.045). The significant enrichment of TG heritability within the peaks of these three cell types was further validated using publicly available single-nucleus ATAC-Seq data from 30 distinct adult human tissues and 222 unique cell types (catlas.org).In order to identify the potential susceptibility variants responsible for the enriched heritability observed in regulatory elements for cardiometabolic traits, we performed experimental fine-mapping using a massively parallel reporter assay in preadipocytes, adipocytes (control and TGFb conditions), and macrophages (control, IL1b and TGFb conditions). In total, we identified variants with allele-specific activity for hundreds of loci associated with cardiometabolic traits across different cell types and conditions. Notably, this analysis unveiled previously unrecognized common regulatory variants near *COBLL1* and *IRS1* genes.In conclusion, our work provides a comprehensive resource for the functional characterization of cell types, variants, and genes associated with cardiometabolic traits in VAT. Additionally, this advanced multiomics approach opens new avenues for the exploration of disease mechanisms in other complex disorders.

Session Title: Molecular Effects of Genetic Variation Poster Session I

PB1249 Single-cell eQTL mapping identifies cell type- and disease-specific genetic control of COVID-19 severity.

Authors:

J. Astley[1], COMBAT Consortium, J. Knight[2], R. Edahiro[3], Y. Okada[4], Y. Luo[1], L. Jostins[5]; [1]Kennedy Inst. of Rheumatology, Univ. of Oxford, Oxford, United Kingdom, [2]Univ. of Oxford, Oxford, United Kingdom, [3]Osaka Univ., Ashiya, Japan, [4]Osaka Univ. / The Univ. Tokyo / RIKEN, Osaka, Japan, [5]Wellcome Trust Ctr. Human Genetics, Oxford, Oxfordshire, United Kingdom

The underlying mechanisms regulating immune response to Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) are not yet sufficiently characterized. Expression quantitative trait loci (eQTL) studies aid discovery of variants regulating gene expression, and differences in signal between cell types or disease severity contexts can be explored to gain insight into this process. We analysed transcriptomic and genetic data from 76 donors and ~$4.6x10^6$ cells of European ancestry: 8 healthy individuals, 16 diagnosed with sepsis at the time of sample collection, and 52 individuals diagnosed with Coronavirus disease (COVID-19) at 4 WHO-defined stages of severity. TensorQTL mapping software was used to perform a genome-wide search for conditionally independent cis-QTLs across 16 major peripheral blood mononuclear cell types, including T cells and monocytes. In total, 2163 cis-independent eQTL signals were found across 16 cell types; most frequently in CD4 cells (792) and classical monocytes (576), followed by CD8 (245) and natural killer (NK) cells (205). The strongest eQTL signal was found in classical monocytes (cMono) at the *RPS26* locus, regulated by variant rs10876864 ($P=1.46x10^{-25}$, $\beta=-1.34$). This relationship was observed in 8 other key cell types involved in immune response: increased dosage of the minor allele (A) was associated with decreased expression of *RPS26* in 3 sub types of mononuclear phagocytes (MNPs) - classical and non-classical monocytes (ncMono) and dendritic cells (DC) - as well as CD4, CD8 and other T cell groups. We report a context specific effect of this eQTL in 5 cell types including NK ($P=2.85x10^{-4}$, $\beta=46.1$), CD4, CD8 and DC. The signal we observed in cMono ($P=8.34x10^{-3}$, $\beta=9.65$) was also detected in monocytes by Edahiro et al (Nature Genetics 2022) in an independent study in a Japanese cohort ($P=3.52x10^{-2}$, $\beta=0.158$). Expression of *RPS26* decreases with increased COVID-19 severity in GG and heterozygous genotype samples, whereas the relationship is reversed with AA genotypes. This variant is in high linkage disequilibrium ($R^2=0.92$) with a variant associated with COVID-19 hospitalization (rs7297175) and is itself associated with several clinical phenotypes including Type 1 Diabetes. Our results emphasise the genetic contribution to individual immune response, in particular to SARS-CoV-2, and demonstrate the importance of expanding these studies to further our understanding of host defence mechanisms in the context of infectious disease.

Session Title: Molecular Effects of Genetic Variation Poster Session II

PB1250 Single-cell eQTL mapping in healthy and Crohn's-afflicted terminal ileum offers novel insights into inflammatory bowel diseases

Authors:

T. Alegbe[1,2], M. Krzak[1], L. Taylor[3], M. Ghouraba[1], M. Strickland[1], M. Tutert[1], M. Ozols[1], J. Ostermayer[1], J. Skelton[4], G. Jones[5], C. Cotobal Martin[1], V. Iyer[1], M. Parkes[6,2], R. E. McIntyre[1], T. Raine[6,2], C. A. Anderson[1]; [1]Wellcome Sanger Inst., Hinxton, United Kingdom, [2]Univ. of Cambridge, Cambridge, United Kingdom, [3]NIH, Bethesda, MD, [4]Cancer Res. UK Cambridge Inst., Cambridge, United Kingdom, [5]Univ. of Edinburgh, Edinburgh, United Kingdom, [6]Addenbrooke's Hosp., Cambridge, United Kingdom

Inflammatory bowel diseases (IBD) are complex and highly polygenic diseases of the gastrointestinal tract. Genome-wide association studies (GWAS) have associated over 320 loci with IBD or the two most common subtypes: Crohn's disease (CD) and ulcerative colitis (UC). As with many other polygenic diseases, for the vast majority of loci it remains unclear which genes and cell types are impacted by disease-associated variants. Expression quantitative trait loci (eQTL) have been used to identify genes whose expression is perturbed by disease associated variants. However, the limited resolution afforded by bulk RNA sequencing, and a lack of context specificity, has hampered these efforts.

To tackle these issues and better understand IBD we generated single-cell RNA-sequencing data from terminal ileal biopsies of 158 healthy and 88 CD-afflicted individuals with the aim of mapping single-cell eQTLS (sc-eQTLs). Our expression dataset consisted of over 1 million single-cell transcriptomes, characterising 49 cell types across immune, epithelial, and mesenchymal populations. We performed pseudobulk eQTL mapping at a high and low cell type resolution finding over 7500 eGenes (FDR < 0.05), just under half of which were undetectable without cell type annotation.

We reproduced eQTL effects from GTEX terminal ileum, GTEX colon, and the database of immune cell eQTLs (DICE) whilst also finding novel eQTLs not present in the bulk datasets. Utilising our cohort's heterogeneity, we were able to map sc-eQTL interactions with disease status and inflammation degree finding 26 interaction sc-eQTL effects. Colocalisation analysis of our TI sc-eQTLs with CD, UC and IBD GWAS resulted in 22%, 20% and 17% of GWAS loci, respectively, colocalising (PP4 > 0.75), implicating the identified eGenes and gut cell types as likely drivers of inflammatory bowel diseases.

Our findings demonstrate the value of mapping eQTLs with tissue-derived disease state single-cell data. We are able to assign IBD-relevant sc-eQTLs known from prior bulk work to intestinal cell types but beyond that, by sampling from one of the most commonly inflamed sites in Crohn's disease, with a disease cohort and at single-cell resolution, we have uncovered novel sc-eQTLs. These sc-eQTLs may have been previously masked by other cell types, undetectable in healthy, uninflamed samples or not found outside of disease site cell types. By combining our sc-eQTLs with existing GWAS, we have discovered genes and cell types that play a causal role in complex disease which brings us a step closer to potential novel, genetically-motivated therapeutics. As our sample size grows we expect to derive more insight into the biology of IBD.

Session Title: Molecular Effects of Genetic Variation Poster Session III

PB1251 Single-cell sQTL links cell type-specific regulation of splicing to complex diseases

Authors:

B. Liu[1], T. Chi[1], Y. Zhang[1], Y. Tong[1], W. Wang[1], K. Kock[2], Y. Okada[3], J. Shin[4], S. Prabhaker[4], Asian Immune Diversity Network; [1]Natl. Univ. of Singapore, Singapore, Singapore, [2]Brigham and Women s Hosp., Boston, MA, [3]Osaka Univ. / The Univ. Tokyo / RIKEN, Osaka, Japan, [4]Genome Inst. of Singapore, Singapore, Singapore

Alternative splicing is a fundamental mechanism that connects genetic variations to complex diseases. To identify genetic effects underlying complex diseases, it has become increasingly important to investigate the proper cell types and contexts. Leveraging the Asian Immune Diversity Atlas dataset, we provide a detailed dissection of the cell-type and context-specific genetic regulation of alternative splicing by single-cell RNA sequencing of ~1M peripheral blood mononuclear cells (PBMCs) from ~500 healthy donors of diverse Asian ancestries. We categorized PBMCs into 22 known cell types in the hematopoietic system and pinpointed two novel subtypes of $CD16^{+}$ monocytes based on *IFITM2* isoforms. We identified 13,460 independent cis-sQTLs with up to 16.1% of sGenes harboring more than one regulatory variant. Multivariate adaptive shrinkage modeling across all cell types revealed that 21.2% of sQTLs are shared across all cell types, while 7.5% are specific to a single cell type. We identified 865 dynamic introns as well as 107 dynamic sQTLs whose genetic effects change along the B cell developmental trajectory. Notably, we identified a dynamic intron usage for *PAX5*, a B-cell-specific transcription factor whose isoform-specific TF activity is in debate. Our data suggested that *PAX5A* upregulation and *PAX5B* downregulation promotes the differentiation of naïve B cells. Our analysis also identified 608 trans-sGenes, 393 (64.6%) of which only appears in a single cell type. In particular, we identified a trans-sQTL for *PTPRC*, a protein tyrosine kinase critical for T-cell development, that colocalizes with a cis-eQTL for *hnRNPLL*, a ribonucleoprotein known to bind to exon four to six of *PTPRC*. Notably, the genetic effects on *hnRNPLL* and *PTPRC* are dynamic across T-cell development, with two distinct lead sSNPs spaced more than 36kb apart and in low LD ($r^2 < 0.3$). Finally, we observed strong enrichment of sQTL effects in autoimmune and inflammatory GWAS loci. Up to 25% of GWAS loci colocalized with sQTLs in at least one cell type. Specifically, we observed a lead sSNP that disrupts the splice donor of exon 1B of *IRF5*, which leads to intron retention and non-sense mediated decay. Moreover, we observed a female-specific colocalization between *CLEC2D*, a natural killer surface receptor, with lymphocyte count. Whether *CLEC2D* is involved in sex biases in autoimmune disease risk requires further investigation. This work reveals several novel regulatory mechanisms through cell-type-specific and dynamic QTL mapping and uses them to explain mechanisms underlying autoimmune diseases.

Session Title: Molecular Effects of Genetic Variation Poster Session I

PB1252 Skipping of COL4A4 exon 27 is associated with hematuria

Authors:

F. Lona Durazo[1,2], K. Omachi[3,4], D. Fermin[5], F. Eichinger[5], J. P. Troost[5], J. H. Miner[3], A. Paterson[6,7,8], M. Barua[8,9,10], S. Gagliano Taliun[11]; [1]Montréal Heart Inst., Montreal, QC, Canada, [2]Université de Montréal, Montréal, QC, Canada, [3]Washington Univ. Sch. of Med., St. Louis, MO, [4]Kumamoto Univ., Kumamoto, Japan, [5]Univ. of Michigan, Ann Arbor, MI, [6]Hosp for Sick Children, Toronto, ON, Canada, [7]Dalla Lana Sch. of Publ. Hlth., Toronto, ON, Canada, [8]Univ. of Toronto, Toronto, ON, Canada, [9]Univ. Hlth.Network, Toronto, ON, Canada, [10]Toronto Gen. Hosp. Res. Inst., Toronto, ON, Canada, [11]Montréal Heart Inst./Université de Montréal, Montreal, QC, Canada

Hematuria is a complex trait that is a result of many genetic and environmental factors. It is an important, yet understudied sign of chronic kidney disease and can be associated with kidney failure. A better understanding of the biological pathways involved in hematuria can ultimately facilitate novel treatment options to improve kidney outcomes of patients. Here, we combined genetic variant-trait association statistics from the UK Biobank with predicted gene expression and splicing from GTEx kidney cortex samples from individuals of European genetic ancestry (n = 65) in a transcriptome-wide association study (TWAS), to identify additional biological mechanisms influencing hematuria. Our TWAS identified significant associations for 5 genes in terms of expression and 3 significant splicing events. Notably, we identified an association between hematuria and the skipping of *COL4A4* exon 27, which is genetically predicted by intronic variant rs11898094 (minor allele frequency 13%). We found independent evidence supporting the existence of this skipping event in glomeruli-derived mRNA transcriptomics data (n = 245) from the NEPTUNE dataset. We then performed functional assays, which revealed that the transcript with exon 27 spliced out affects collagen IV trimer assembly and secretion (Student's *t*-test p-value < 0.005, compared to wild type), highlighting the potential impact of this splicing event on kidney biology. Altogether, our results highlight the value of investigating the role of non-coding sequence, an underexplored region, by integrating multiple data types to shed light on kidney traits and their underlying disease mechanisms.

Session Title: Molecular Effects of Genetic Variation Poster Session II

PB1253 SNPred outperforms other ensemble-based SNV pathogenicity predictors and elucidates the challenges in performance evaluation of predictive models using ClinVar.

Authors:

I. Molotkov[1], M. Artomov[1,2]; [1]Nationwide Children's Hosp., Columbus, OH, [2]The Ohio State Univ. Coll. of Med., Columbus, OH

Accurate identification of pathogenic variants is crucial for the diagnosis, treatment, and prevention of genetic diseases. In this study, we introduce SNPred, an ensemble gradient boosting model for predicting the pathogenicity of nonsynonymous single nucleotide variants (SNVs). SNPred uses 40 variant pathogenicity predictors from dbNSFP database to create an aggregate pathogenicity score. The performance of SNPred was evaluated on seven distinct validation datasets derived from ClinVar and *BRCA1* Saturation Genome Editing (SGE) data. Across all validation scenarios, SNPred consistently outperformed other state-of-the-art tools, such as BayesDel, MetaRNN, and REVEL, particularly in the case of rare and cancer-related variants (ROC AUC of 0.99 for SNPred compared to 0.94-0.97 for the next four leading methods). We employed a novel approach to validation on a curated set of ClinVar variants that are inaccurately predicted by most pathogenicity prediction tools. In this challenging scenario, SNPred exhibited a remarkable ROC AUC of 0.99, surpassing the next four leading predictors by a margin of 0.05 to 0.08.

During the validation process, we discovered several challenges which are inherent to validation of SNP pathogenicity predictors on ClinVar and may lead to inaccurate performance estimations. Firstly, we sought to empirically examine the hypothesis that variants which are accurately classified using computational tools appear in ClinVar more frequently, making them overrepresented in the database. To that end, we considered a set of *BRCA1* variants from ClinVar and a non-intersecting set of *BRCA1* variants from an SGE study. We tested how 20 top-performing predictors would perform on these two datasets. Confirming our hypothesis, all 20 of the predictors classified ClinVar variants more accurately, with the mean difference in AUC ROC of 0.07. Secondly, we assumed that, as literature suggests, the misclassification rate in ClinVar is estimated to be at least 2%. As a result, we prove that the exceptionally high accuracy scores achieved by certain models on ClinVar are only attainable if the models overtrain by learning to misclassify variants that are already misclassified in ClinVar. These findings indicate that validation of any pathogenicity prediction algorithm using ClinVar data may result in inflated performance estimates.

In conclusion, we developed a new tool for SNV pathogenicity prediction that outperforms other ensemble models in various validation setting. Additionally, we highlighted potential sources of inflation in performance estimation when using ClinVar.

Session Title: Molecular Effects of Genetic Variation Poster Session III

PB1254 Statistical Mechanics of coding variants quantitatively predicts gene influence in select populations, including Parkinson's Disease cohorts.

Authors:

O. Lichtarge[1], Y. Lagisetty[2], M. Samieinasab[1], C. Wang[1], F. Nazem[1], D. Marciano[1], P. Katsonis[1], J. Asmussen[1], J. Botas[1], M. Raveendran[3], A. Harris[1], J. Rogers[1], I. Al-Ramahi[1], K. Lee[4], J. Botas[1]; [1]Baylor Coll. of Med., Houston, TX, [2]Univ. of Texas McGovern Med. school, Houston, TX, [3]BCM, Houston, TX, [4]Baylor Coll. Med., Houston, TX

Life is complex and seemingly difficult to describe with simple equations. Yet, this study shows that genetic information obeys statistical mechanics, leading to the powerful imputation of genes in complex diseases, such as Parkinson's.

A first step computes the energy of mutations in fitness landscapes with a modified evolutionary action equation. We may then calculate the energy distribution of coding variants in sequenced genomes and find it follows Boltzmann statistics, from genes to chromosomes and from individuals to populations of humans, macaques, and *E. coli*. These data suggest that statistical mechanics is a fundamental description of natural coding variants, across scales.

To rule out mere coincidence, we asked if statistical mechanics could predict a new genetic feature. We focused on the equipartion of energy, a central tenet of modern physics intrinsic to statistical mechanics, stating that systems in equilibrium distribute energy equally among their degrees of freedom. We found that for equipartition to hold, we must precisely define a new variable, w (normalized by gene length), corresponding to a gene's mutational influence (or *weight*) in the fitness landscape.

We found that the w value of each gene correlated with degree centrality in protein interaction networks (Spearman $\rho = 0.97$, $p < 10^{-60}$). Genes with top w values also interacted mutually ($p < 10^{-16}$); mediated essential biological processes; and were enriched for haplotype insufficiency ($p < 10^{-21}$) and disease associations ($p < 10^{-49}$) in humans. Similar findings held in macaques and *E. coli*, in which growth after gene knockdown correlated inversely with w ($R^2 = -0.72$, $p < 10^{-27}$). Thus, equipartition predicts gene weight and this weight agrees with observations on gene influence.

Strikingly, while most genes had equal w values upon comparing 1978 healthy subjects and 1826 Parkinson's Disease (PD, from AMP-PD) patients, a subset of genes had greater weight w either in patients or controls. Multiple tests link these genes to PD, including their ability to separate cases from controls in independent, held-out subjects ($AUC = 0.63$, $z = 6.2$). Thus, deviations in w between populations inform on disease associations.

To our knowledge, this work is the first to unify physics, genetics, and evolution into a computable, quantitative framework defining the mutational energy of coding variants and the weight (w) of genes. While future extension to non-coding variants is desirable, in practice, this paradigm shift to the physics of fitness landscapes already sheds light on influential genes in select sequenced populations, including cohorts harboring a complex disease, such as PD, or free from it.

Session Title: Molecular Effects of Genetic Variation Poster Session I

PB1255 † Statistically and functionally fine-mapped blood mRNA and protein expression QTLs from 1,405 humans reveals their distinct regulation patterns and disease relevance

Authors:

Q. Wang[1,2], T. Hasegawa[3], H. Namkoong[4], S. Ryunosuke[5], R. Edahiro[2], K. Sonehara[1,2], H. Tanaka[4], S. Azekawa[4], S. Chubachi[4], Y. Takahashi[2], S. Sakaue[6], S. Namba[2], K. Yamamoto[2], Y. Shiraishi[7], K. Chiba[7], H. Tanaka[3], H. Makishima[5], Y. Nannya[5], Z. Zhang[5], R. Tsujikawa[5], R. Koike[3], T. Takano[8], M. Ishii[4], A. Kimura[3], F. Inoue[5], T. Kanai[4], K. Fukunaga[4], S. Ogawa[5], S. Imoto[1], S. Miyano[3], Y. Okada[1,2,9], Japan COVID-19 Task Force; [1]The Univ. of Tokyo, Tokyo, Japan, [2]Osaka Univ., Osaka, Japan, [3]Tokyo Med. and Dental Univ., Tokyo, Japan, [4]Keio Univ., Tokyo, Japan, [5]Kyoto Univ., Kyoto, Japan, [6]Harvard Med. Sch., Boston, MA, [7]Natl. Cancer Ctr. Res. Inst., Tokyo, Japan, [8]Kitasato Univ., Tokyo, Japan, [9]RIKEN, Yokohama, Japan

Studying genetic regulations of protein expression (protein expression QTL = pQTLs) offers deeper understanding of regulatory variants uncharacterized by studies of mRNA expression regulation (expression QTL = eQTLs) alone. Here we report on a comprehensive cis-e/pQTL statistical fine-mapping from 1,405 genotyped samples with blood mRNA and plasma protein expression (n=2,932 proteins from Olink explore 3072) information, as part of the Japan COVID-19 Task Force (JCTF).

Fine-mapped eQTLs (defined as posterior inclusion probability = PIP>0.9, n=3,464, nearly tripling the size of our previous release) were enriched for 932 variants validated with in-house massive parallel reporter assay (MPRA) using K562 and HepG2 (>2 fold enrichment compared to those with PIP<0.1; p=2.5e-6). Fine-mapped pQTLs (n=582) were enriched for missense variations on structured and extracellular domains (>2,000 fold enrichment compared to random, 2.5 and 1.4 fold within missense variations; p<1e-100). When focusing on trans-e and pQTL associations, our analysis revealed associations between class I HLA and KIR genes (KIRDL1 at mRNA and KIRDL3 at protein level; p=7.2e-9 and p=1.2e-25).

We found a limited level of colocalization between e/pQTLs (only 12% of the genes presented colocalization posterior probability = CLPP>0.1). Genes with mRNA specific-QTLs are relatively universally expressed (average 46 GTEx tissues versus 41 for protein specific-QTLs; p=6.3e-4), suggesting buffering of mRNA expression variation at post-transcriptional, multi-tissue levels. Protein specific-QTLs were more relevant to complex traits in biobank studies (7.6 versus 5.6 fold enrichment of trait-causal variants in Biobank Japan; p=2.7e-3). As examples of disease relevant genes with protein-specific regulation, we highlight *TNFRSF11A*, *APOE*, and *EFHD1*.

Utilizing COVID-19 severity information, we also show that condition-specific interaction eQTL (ieQTL) effects have limited penetrance to plasma protein levels, due to expression leakage from non-blood tissues: even for *CLEC4C*, where the ieQTL effect is the most significant (p=4.1e-32 for rs7302014), ipQTL effect does not reach significance (p=2.4e-7). Finally, highlighting the ABO gene as an example, we show that negative correlation between mRNA and protein expression can arise due to linkage disequilibrium (LD) between distinct nearby eQTLs and pQTLs.

Our results highlight the value of pQTL analysis for a finer understanding of gene expression regulations and identification of disease-associated loci. We also developed an interactive browser, planned for public release by the time of the conference.

Session Title: Molecular Effects of Genetic Variation Poster Session II

PB1256 Strain-specific structural variants in mice impact CTCF-mediated transcriptional control.

Authors:

N. Moskwa[1], S. Syed[2], C. Lee[2]; [1]The Jackson Lab., Farmington, CT, [2]The Jackson Lab. for Genomic Med., Farmington, CT

Genomic variants are a major contributor to phenotypic diversity within a species. However, linking the functional relevance of noncoding genetic variants to specific phenotypes has been challenging due to the incomplete annotation of noncoding DNAs. Among the different type of genetic variants, structural variants (SVs) are widespread and affect more bases in the typical genome than single nucleotide variants. When SVs accumulate in noncoding regulatory regions, they have the potential to create population-enriched transcriptional profiles within a species.

In the cell, chromosomes are partitioned into topologically associating domains (TADs). TAD boundaries are enriched in CCCTC-binding factor (CTCF) sites. Interestingly, there are thousands of CTCF chromatin binding sites (CCBSs) riddling the noncoding genome that are not involved in TAD formation. In fact, approximately 40% of CCBSs are intronic. Intronic CCBSs are capable of mediating chromatin interactions that are implicated in gene regulation.

To better understand which SVs disrupt specific CTCF-mediated chromatin interactions and lead to phenotypic diversity, we have conducted an investigation using two phenotypically diverse inbred mouse strains: C57BL/6J (B6) and 129S1/SvImJ (129S). We have developed a strategy to identify noncoding genetic variants responsible for inter-strain phenotypic differences, with a particular focus on intronic SVs that disrupt CCBSs and chromatin contacts. We performed a comprehensive association study between B6 and 129S mouse embryonic stem cells (ESCs), integrating RNA-seq, chromatin organization, and long read whole genome sequencing data. We have identified a total of 4,168 B6-specific and 1,654 129S-specific CTCF-bound chromatin interactions overlapping SVs, which appear to result in the dysregulation of 1,053 genes (129S vs B6 ESCs comparison; padj<0.1 and |log2(fold change)| > 0.5). To test for the causality of intronic SVs that disrupt CCBSs in a strain-specific manner, we used CRISPR/Cas9 to introduce prioritized SVs into B6 ESCs and assessed gene expression, CTCF chromatin binding and genome interaction differences between control and SV-engineered ESCs. Our preliminary findings have identified SVs overlapping genes *Ireb2*, *Mtf1*, *Ccdc171*, *Mnd1*, and *Numbl* as possible contributors to phenotypic differences observed between the B6 and 129S ESCs. Our goal is to ultimately annotate all noncoding SVs in the mouse genome and understand the mechanisms by which noncoding SVs change cell type-specific transcriptomes, resulting in strain-specific phenotypes.

Session Title: Molecular Effects of Genetic Variation Poster Session III

PB1257 Structural bioinformatics for deep variant phenotyping of SMARCA4.

Authors:

M. Zimmermann[1], N. Haque[1], B. Ratnasinghe[1], S. Jorge[1], Y-I. Chi[2], J. Wagenknecht[1], R. Urrutia[1]; [1]Med. Coll. of Wisconsin, Milwaukee, WI, [2]Med. Coll. of Wisconsin, Wauwatosa, WI

Protein-coding genetic alterations are frequently observed in Clinical Genetics, but the high yield of variants of uncertain significance remains a limitation in diagnostics and decision-making. Interpreting genetic variation is now a more pressing challenge than sequencing individuals. For genes with established roles in defining congenital syndromes, we have yet to uncover the underlying mechanisms for why some genetic variation is pathogenic, and others are not. For example, SMARCA4 mutations cause Coffin-Siris Syndrome, yet 338 missense variants have already been observed, 311 (92%) of which lack interpretation primarily due to their observation in single cases, plus 27 (likely)pathogenic. Thus, new approaches are needed to enhance the information available to each patient. We take a proteogenomic approach and use computational tools to calculate how genetic variants alter their encoded protein structures, dynamics, and effects on protein complexation. These tools are established in basic sciences but have yet to achieve systematic use in genomics. Further, the revolution in protein structure prediction driven by artificial intelligence algorithms empowers our approach and enables scalability of structural bioinformatics. Herein, we characterize a broad mutational landscape of SMARCA4 variations using not only genomic scores, but calculations derived from cryo-EM and protein structure, such as rigidity and flexibility analyses and active site rearrangements. Our goal is to interpret how mutations affect the enzyme functionally. Then, how SMARCA4 fits within its cognate chromatin remodeling complex, and how genetic variation in SMARCA4 may also change the complex. This scale in the number of mutations characterized and structural context is key for improving the interpretation of novel and ultra-rare genomic variants. The mechanistic readouts our approach generates will indicate specific testable hypotheses for functional genomics validation, including changes to protein stability, histone binding capacity, and helicase processivity. Our study is trailblazing and practical for enhancing the field. Similar approaches can be applied to help interpret human genetic variation underlying many congenital and malignant conditions, such as we demonstrate for Coffin-Siris Syndrome.

Session Title: Molecular Effects of Genetic Variation Poster Session I

PB1258 Structural variants and heterochromatin link chromatin risk genes with neurodevelopment in autism spectrum disorder.

Authors:

M. Garvin[1], D. Kainer[2]; [1]Oak Ridge Natl. Lab., Oak Ridge, TN, [2]Univ. of Queensland, Brisbane, Australia

Autism spectrum disorder (ASD) is a highly heritable and highly heterogeneous neuropsychiatric condition whose cause is still unknown because there are no recurrent genes found among diagnosed individuals. One of the most common functional properties of the many reported risk-genes for autism is “chromatin modification” but it is not known how this biological process relates to neurodevelopment and autism. We recently reported frequent, recurrent genomic structural variants (SVs) in two cohorts of individuals with autism that were identified using Non-Mendelian inheritance (NMI) patterns in family trios. The genes harboring the SVs participate in neurodevelopment, glutamate signaling, and chromatin modification, confirming previous reports and providing greater detail for these processes in ASD. The majority of these ASD-associated SVs (ASD-SV) were found in non-coding regions of the genome and were enriched for expression quantitative trait loci (eQTL) suggesting that gene dysregulation results from these genomic disruptions rather than alteration of proteins. Here, we intersect the ASD-SV from our earlier work with different gene regulatory and epigenetic multi-omic layers to understand how they may function to produce autism. Our results indicate that the core of ASD resides in the dysregulation of a process called RNA-induced Initiation of Transcriptional gene Silencing (RITS) that is meant to maintain heterochromatin and produces SVs in the genes within these chromosomal regions, resulting in alterations in brain development. This finally links reported ASD-risk genes involved in chromatin remodeling with neurodevelopment. In addition, it may explain the role of de novo mutations in ASD and provide a framework for more accurate diagnostics and endophenotypes.

Session Title: Molecular Effects of Genetic Variation Poster Session II

PB1259 Systematic analysis of mitochondrial carrier functional redundancy using complex genetic interaction analysis.

Authors:

B. Greco, K. Dadelahi, E. Kuzmin; Concordia Univ., Montreal, QC, Canada

Complex genetic interactions occur when a combination of mutations in more than two genes leads to an unexpected phenotype that deviates from a combined effect of individual mutations and lower order combinations. Due to the prevalence of complex genetic interactions, they are thought to have the potential to affect the biology of inheritance. In this work, we use the budding yeast model organism to investigate the complex genetic interactions involving a conserved eukaryotic gene family of mitochondrial carriers to understand how they buffer each other's loss and contribute to genomic robustness. Mitochondrial carriers regulate the transfer of metabolites across cytosol and mitochondrial matrix and thus play an important role in cellular metabolism and human disease. We used Synthetic Genetic Array (SGA) haploid selection method to construct 240 unique double mutant with 23 single mutant control query strains involving 23 yeast mitochondrial carrier orthologous genes that have been directly implicated in human diseases. We will interrogate them using trigenic-SGA to construct 1330 triple mutants and measure their trigenic interactions to understand how mutations in three genes combine to affect fitness. Trigenic-SGA involves crossing double mutant query strains to an array of single mutants and through an automated mating and meiotic recombination to select haploid triple mutants, measure their colony size and quantify their trigenic interactions. By integrating these data with genome-wide association studies (GWAS) and validations in human cells should enable us to gain insight into disease heritability related to mitochondrial carrier gene family. Ultimately, this work will shed light on the genotype-to-phenotype relationship and complex disease heritability.

Session Title: Molecular Effects of Genetic Variation Poster Session III

PB1260 Systematic curation and validation of pathogenic and likely pathogenic 3’ and 5’ untranslated region variants using deep learning artificial intelligence models.

Authors:

E. Bohn, T. Lau, D. Merico, A. Celaj, A. Deshwar, S. Gandhi, B. Kanuparthi, T. Masud; Deep Genomics, Toronto, ON, Canada

Background: Variants in 5’ and 3’ untranslated regions (UTR) are implicated in disease. While predictive algorithms have been developed to assist in interpreting their pathogenicity, their utility is contingent on the quality of data used for training and validation. **Purpose**: We curated a high-confidence set of pathogenic and likely pathogenic (P/LP) variants in the 3’ and 5’ UTR and used publicly available deep learning artificial intelligence (DL-AI) models to validate these classifications. **Methods**: 3’ and 5’ UTR variants documented as P/LP were obtained from ClinVar. Evidence was curated for each variant to inform classification following American College of Medical Genetics and Genomics (ACMG) and Association for Molecular Pathology (AMP) guidelines, and recommendations for adaptation to non-coding contexts. DL-AI models (FramePool, Saluki, Enformer) trained on UTR-related mechanisms were applied for validation. Prediction scores were compared between P/LP variants acting through the mechanism for which the model was designed (model-matched), P/LP variants operating through a different or undetermined mechanism (model-mismatched), and putative benign variants. PhyloP was used to compare conservation scores between P/LP and putative benign variants. **Results**: Of 295 3’ and 188 5’ UTR variants identified through ClinVar, 26 and 68 were interpreted as P/LP, respectively. Mechanistic information was available for 22 3’UTR variants, the most frequent being modulating a microRNA binding site (n = 7), impacting the polyadenylation signal (n = 10), and impacting mRNA stability (n = 3). 5’ UTR variants operated at both the level of transcription and translation. Common mechanisms included altering promoter activity (n = 14), impacting an upstream open reading frame (n = 10), and introducing a novel upstream start codon (n = 13). A significant difference in prediction scores was obtained with all DL-AI models when comparing model-matched P/LP variants, to both model-mismatched P/LP and putative benign variants, which was also supported by our internally developed proprietary predictors. PhyloP conservation scores were significantly higher among P/LP compared to putative benign variants for both the 3’ and 5’ UTR. **Conclusions**: We present a high-confidence set of P/LP 3’ and 5’ UTR variants spanning a range of mechanisms. DL-AI models substantiate these interpretations, with a distinction in scores supporting the relevance of mechanism-informed use. These datasets will support further development of AI algorithms designed to predict the impact of 3’ and 5' UTR variants which may be implicated in disease.

Session Title: Molecular Effects of Genetic Variation Poster Session I

PB1261 Targeted CRISPRa/CRISPRi screen confirms functional SNPs and identifies novel target genes at multiple RCC susceptibility loci.

Authors:

T. Winter[1], T. Myers[1], L. Colli[2], L. Jessop[1], J. Choi[1], M. J. Machiela[1], M. P. Purdue[1], K. Brown[1], S. Chanock[1]; [1]Natl. Cancer Inst., Rockville, MD, [2]Ribeirao Preto Med. Sch. - Univ. of Sao Paulo, Sao Paulo, Brazil

Genome wide association studies (GWAS) have identified 13 renal cell carcinoma (RCC) risk regions as well as additional loci with nominal significance. Detailed investigation of how each susceptibility region functions can reveal the underlying biological bases of RCC susceptibility. To date, most characterized GWAS regions function through altered regulation of one or more genes in a critical pathway in RCC, as shown with regions 8q24, 11q13, 12p12.1 and 14q24. Disease risk alleles have been mapped to enhancers which loop to the promoters of target genes and regulate their expression. To identify such interactions between functional SNPs and target genes, we evaluated a set of variants based on a Massively Parallel Reporter Assay (MPRA), eQTL analysis, capture Hi-C, and an arrayed CRISPRa/CRISPRi screen in the same two cell lines as the MPRA. The MPRA identified 196 SNPs across 19 regions with significant allele-specific effects, indicating cis-regulatory activity. Of these, 39 SNPs displayed chromosome loops allowing physical interaction with the promoter of a nearby gene, as well as a significant cis-eQTL with the same gene. In this way, 24 putative target genes of 39 SNPs across 10 RCC risk loci were nominated. To confirm the presence of a cis-regulatory relationship between these SNPs and putative target genes, we performed an arrayed CRISPRa/CRISPRi screen in ACHN (RCC) and HEK293T (embryonic renal) cells covering each of the 10 regions, 39 SNPs, and 24 candidate target genes. Cis-regulatory relationships were assessed by TaqMan qPCR for the 24 putative target genes. For CRISPR inhibition, ACHN and HEK293T cells were stably transduced with inactive cas9 fused to the ZIM3 repressor domain and for CRISPR activation, fused to the VP64 activator domain. Initial results of the screen have identified multiple novel cancer relevant target genes of renal cell carcinoma predisposition, including *ABL2* at 1q25 and *MAP2K1* at 15q22. Further functional investigation of these regions is underway to characterize these SNP-gene relationships and elucidate their role in RCC risk.

Session Title: Molecular Effects of Genetic Variation Poster Session II

PB1262 Telomere-to-telomere, pangenomic references improve eQTL power and enable the detection of sex-specific eQTLs in the developing human brain.

Authors:

J. Lalli[1], J-Y. An[2], N. Sestan[3], S. Sanders[4], D. Werling[1]; [1]Univ. of Wisconsin-Madison, Madison, WI, [2]Korea Univ., Seoul, Korea, Republic of, [3]Yale Univ., New Haven, CT, [4]Oxford Univ., Oxford, United Kingdom

Autism spectrum disorder (ASD) is 4X more prevalent in males, and sex-specific differences in gene regulation likely contribute to this imbalance. Expression QTL (eQTL) studies are a well established method of determining the effect of naturally occurring genetic variants on nearby gene regulation. However, past studies have found sex-specific eQTLs difficult to detect using standard association methods.
The subtle effects of common variants on gene expression can be hidden by technical noise. One source of noise, especially in ancestrally diverse cohorts, is reference error. In regions of high homology (eg., pseudogenes, or tandem repeats), small differences between a reference sequence and a patient genome can result in variant calling errors and gene expression variability that is poorly captured by current methods.
Working with our previously published BrainVar cohort of bulk RNA and DNA from the pre-frontal cortex of 87 human fetuses, we measured the role of reference error in the detection of sex-specific eQTLs by extending the Nextflow RNAseq and Sarek WGS pipelines to align reads to either the GRCh38, T2T, or human pangenome references. These references have been shown to reduce mismapping during alignment, especially in regions of high homology. Implementing this pipeline required producing a lifted over GATK resource bundle and phased 1000 Genomes variant panel in T2T coordinates.
Aligning WGS reads to the T2T genome reference resulted in increased autosomal depth (30.3X -> 32.1X) and coverage (94.1% -> 99.4%). Pangenomic alignment also improved autosomal depth (31.6X), with reduced variability of coverage. More RNAseq reads aligned to the Refseq v110-T2T annotation than the GRCh38 annotation. Analysis of the resulting data showed that the T2T annotation in an over 20% increase in the number of sex-differential eGenes discovered, while the number of pseudogenic eGenes declined. Gene ontology analysis of the T2T alignment data revealed that genes regulated in a sex-specific manner were associated with impaired language skills and neurodevelopmental delay, and were associated with the transcription factor SP1, a transcription factor known to complex with estrogen receptor. These results were not observed in GRCh38 aligned data.
To our knowledge, this work is the first comparative study to assess how the T2T and/or pangenome reference affects eQTL association studies. We believe this work demonstrates the need for geneticists to make use of these new resources to address reference error. All resources and pipelines are available for general use on Github.

Session Title: Molecular Effects of Genetic Variation Poster Session I

PB1264 The challenges of variant classification - a snapshot of reality through proficiency testing

Authors:

S. Deans[1], D. Cregeen[1], F. Khawaja[1], M. Sales[1], M. Tabiner[2], R. Treacy[1]; [1]GenQA/UK NEQAS, Edinburgh, United Kingdom, [2]GenQA, Oxford, United Kingdom

Background/Objectives: Clinical genomic testing requires the classification of the pathogenicity of sequence variants, and the assessment of their clinical impact. As more guidance is generated to aid this work, there is a challenge to standardise approaches across variant types and in different clinical settings. GenQA has globally delivered laboratory proficiency testing (PT)/external quality assessments (EQAs) for variant interpretation since 2013 and has adapted this approach to provide a mechanism for individuals to demonstrate their competency, and to support the education of the scientific and clinical workforce.
Methods: Modules for classification of single nucleotide variants (SNVs) and copy number variants (CNVs) were provided online through the Genomics Individual Education (GENie) platform. Individual participants were provided with details of the variants and the clinical setting and were expected to classify them along with documenting the evidence used to obtain the classification. Real time assessment was applied through the platform and participants received a summary of their performance and the expected classification detailing the evidence applied.
Results: As a pilot, individuals completed the classification of three SNVs and/or three CNVs during a weeklong exercise. Following the successful delivery of this individual educational module, ongoing accessible assessments for either germline SNVs or CNVs were delivered whereby a random set of six variants were provided for classification each time a participant accessed GENie. These were generated from a vast bank of variants and included prenatal, postnatal, diagnostic, and predictive clinical scenarios. The range of variant classifications from the pilot and GENie modules will be presented.
Conclusion: Accuracy of variant classification is essential to ensure patients receive the correct results and clinical management. This assessment demonstrated variability in the use and application of the guidelines, and the continued need for PT/EQA to educate and promote standardisation.

Session Title: Molecular Effects of Genetic Variation Poster Session II

PB1265 The contribution of deleterious variants affecting PAM amidating activity on insulin response, incretin levels and diabetes risk.

Authors:

A. Giontella[1], M. Åkerlund[1], L. Lotta[2], A. Baras[2], J. Overton[2], P. Kaufmann[3], Y. Ilina[3], O. Melander[4]; [1]Lund Univ., Malmö, Sweden, [2]Regeneron Genetics Ctr., Tarrytown, NY, [3]PAM Theragnostics GmbH, Hennigsdorf, Germany, [4]Malmo Univ Hosp, Malmo, Sweden

Background: Peptidylglycine-α-amidating monooxygenase (PAM) is the only known enzyme responsible for the amidation process of bioactive peptides, an essential step for their biological activation.PAM is a key enzyme for the endocrine system, and perturbation in its activity has been reported to be associated with insulin secretion, diabetes, and cardiovascular diseases, with a mechanism that remains unclear. **Aim:** This work aimed to investigate at the population level the effect of coding variants that affect PAM amidating activity (PAM-AMA) on hormonal regulation of insulin secretion, using data from the Malmö Diet and Cancer (MDC, n∼27,000), the Malmö Preventive Project (MPP, n∼3,600), and the UK Biobank (UKB, n∼500,000). **Methods:** Exome-wide association analysis was used to identify deleterious coding variants associated with PAM-AMA, which was assessed with a validated assay in ∼ 8,000 from MDC and MPP. The variants, aggregated in a weighted score, were tested for their association with fasting and after oral glucose tolerance test (OGTT) insulin and incretin levels (GIP, GLP-1) in the MDC, at fasting and after. Additionally, the association with glycemic traits (glucose level and HbA1c) and diabetes risk, was evaluated in all the three cohorts. **Results:** Two missense aminoacidic substitutions in *PAM* gene, Ser539Trp (MAF: 0.7%) and Asp563Gly (MAF: 5%), resulted as the lead variants of the association, which independently contribute to a decrease of 2.33 (p: $2.5E^{-140}$) and 0.98 (p: $1.12E^{-225}$) SD units of PAM-AMA, respectively.The score showed a strong association with lower insulin level at 30 min after OGGT (-0.07 [-0.10;-0.03]; p: $5.0E^{-4}$ for each SD increase of the score). A suggestive inverse association was found with fasting GLP-1 (-0.03 [-0.07,-0.001]; p: 0.05), while no association was reported for GIP.In addition, the score was associated with increased glucose level (0.007 [0.005,0.010]; p:$2.2E^{-5}$), HbA1c (0.012 [0.009,0.015]; p: 3.8^{E-14}), and with increased diabetic risk (OR [95%CI] 1.07 [1.06,1.08]; p: $8.08E^{-41}$). **Conclusion:** This study represents the largest population-based study reporting the role of the measured PAM amidating activity on insulin response, and diabetes risk. Individuals who carry PAM-AMA deleterious missense mutations, exhibit a lower insulin response, higher glycemic traits, and an overall higher diabetes risk. These findings highlight the importance of early identification of PAM LoFs carriers, as they could benefit from targeted anti-diabetic treatments.

Session Title: Molecular Effects of Genetic Variation Poster Session III

PB1266 The contribution of short tandem repeats to splicing variation in humans

Authors:

Y. Li[1], J. Margoliash[2], A. Goren[3], M. Gymrek[4]; [1]Univ. of California San Diego, San Diego, CA, [2]Univ. of California, San Diego, La Jolla, CA, [3]UCSD, San Diego, CA, [4]Univ California San Diego, La Jolla, CA

Short tandem repeats (STRs) are consecutively repeated sequences of 1-6 base pairs that constitute a significant source of human genetic variation. Multiple recent studies have revealed a broad contribution of STRs to gene regulation and complex traits. In addition to regulating gene expression, STRs have also been implicated in multiple splicing-related processes, including forming binding sites for splicing factors such as hnRNP-L. However, their contribution to splicing variation in humans has not been explored genome-wide. Here, we leveraged whole-genome sequencing and expression data from the Genotype-Tissue Expression (GTEx) Project to comprehensively assess the impact of STRs on splicing in humans. Using the tool HipSTR, we conducted multi-sample genotyping of 1.6 million STRs across 652 individuals. We examined associations between the repeat length of STRs and the percent spliced-in (PSI) values of exons within 100kb of these STRs while controlling for population structure, sex, and PEER factors. We tested 269,298 unique STR-exon pairs and discovered 46,958 significant associations (termed SpliceSTRs) at a false discovery rate of 10%. We applied the FINEMAP algorithm to conduct fine-mapping of the SpliceSTR associations against nearby SNPs. This yielded 5,393 unique spliceSTRs, termed FM-spliceSTRs, with a posterior inclusion probability of at least 50%. As an orthogonal validation, ANOVA analysis has shown that 45.3% FM-spliceSTRs explain exon splicing beyond the best SNPs. Compared to all the analyzed repeat units, we found significant enrichment of FM-spliceSTRs in specific repeat units, including AAAG/AAGG, GT/CG, ATCC, and AAAAC ($p<0.05$). FM-spliceSTRs close to splicing sites may directly modulate the splicing process. In one example, a GT dinucleotide repeat, positioned 17 base pairs downstream from the donor site of exon 2 of TMEM263 is negatively associated with the splicing of that exon. This finding is further supported by the predictions obtained from spliceAI, which show a correlation between the length of the STRs and the predicted donor site score. Overall, our study unveils the widespread impact of STRs on alternative splicing and identifies a set of highly credible FM-spliceSTRs. These findings provide valuable insight into the pathomechanism of genetic diseases.

Session Title: Molecular Effects of Genetic Variation Poster Session I

PB1267 The crossroads of fine-mapping, single-cell multiome and immunophenotyping to inform on the molecular drivers of diseases in FinnGen

Authors:

M. Kanai[1], T. M. Delorey[1], C. B. Porter[1], K. Nanki[1], H. Hwang[1], I. Teixeira-Soldano[1], J. Karjalainen[2], R. S. Rodosthenous[2], J. Halonen[3], J. Honkanen[3], J. Juvila[3], D. Chafamo[1], T. Yoshida[1], M. Arvas[3], M. Kurki[2], M. P. Reeve[1], W. Zhou[1], FinnGen, J. Partanen[3], A. Palotie[2], D. B. Graham[1], M. J. Daly[1], R. J. Xavier[1]; [1]Broad Inst. of MIT and Harvard, Cambridge, MA, [2]Univ. of Helsinki, Helsinki, Finland, [3]Finnish Red Cross Blood Service, Helsinki, Finland

Despite recent success in identifying causal variants of complex traits, translating their functional effects into mechanistic insights requires substantial efforts. In particular, population isolates such as those in Finland have excelled in uncovering novel causal variants but lagged in characterizing their functional effects due to the lack of samples carrying such high-impact, population-enriched variants.

To address this issue, we present a single-nuclei multiome atlas of gene expression and chromatin accessibility in peripheral blood mononuclear cells (PBMC) from Finnish blood donors. In collaboration with the Finnish Red Cross Blood Service, we collected 960 PBMC samples from the FinnGen participants, prioritizing carriers of Finnish-enriched variants. We utilized the 10x multiome to jointly profile gene expression and chromatin accessibility within the same nuclei, and the FACS immunophenotyping to quantify frequencies of immune cell types.

Here, we initially analyzed 192 samples for a pilot analysis. This pilot atlas comprises 1.4 million single nuclei of PBMCs. Using pseudobulk approach, we identified 8,629 unique eGenes (FDR < 5%) for at least one cell type. We found 8,805 unique *cis*-eQTL variants with > 2x Finnish enrichment, including 370 variants that belong to 95% credible sets (CS) in disease GWAS of FinnGen. For example, we found a 114x Finnish-enriched (MAF = 4.2%) inflammatory bowel disease (IBD) risk variant rs748670681 (*TNRC18* intron; OR = 2.2) showed significant *cis*-eQTL associations in CD4/8 T cells ($P = 1.0 \times 10^{-24}$; $\beta = -1.7$).

Leveraging the single-nuclei multiome, we constructed cell-type-specific enhancer-gene maps using SCENT (Sakaue et al, *medRxiv*). We identified 17,700 unique chromatin accessibility peaks significantly associated with *cis*-gene expression within the same nuclei for at least one cell type. Of these, 1,276 peaks overlap with GWAS fine-mapped variants in 95% CS, including rs748670681 that overlaps with CD4/8 T cell specific chromatin accessibility peak that positively regulates *TNRC18* expression.

Finally, we leverage the atlas to investigate the underlying molecular drivers of diseases in FinnGen. We annotate each disease GWAS loci whether they mediate chromatin accessibility and/or gene expression in *cis* and *trans*, and quantify disease heritability mediated via molecular traits. For example, we are able to dissect the molecular role of rs748670681 into its *cis*-eQTL effect and colocalization with TNRC18-regulatory peak while lacking caQTL effect with the colocalizing peak itself. Our findings underscore the potential of a biobank-scale atlas for characterizing Finnish-enriched variants.

Session Title: Molecular Effects of Genetic Variation Poster Session II

PB1268 The global transcriptomic response to genetic perturbations

Authors:

A. Nadig[1], E. Robinson[2], L. O'Connor[1]; [1]Broad Inst., Cambridge, MA, [2]Massachusetts Gen. Hosp., Boston, MA

Pooled perturbational screens with gene expression profiling provide a compelling opportunity to study thousands of genetic perturbations at once. As in GWAS, many genuine nonzero associations may not reach statistical significance, making them invisible to downstream analyses. A lack of methods to model these effects leaves open fundamental questions: (1) which perturbations cause the greatest transcriptomic change?, (2) what is the transcriptome-wide relationship between effects of different perturbations?, and (3) what proportion of effects are captured by significant differentially expressed genes? We present Transcriptome-wide Analysis of Differential Expression (TRADE), a suite of statistical tools for modeling distributions of differential expression effects from large-scale perturbational screens or any other differential expression experiment. TRADE fits a flexible mixture model to observed effects and standard errors to estimate the distribution of true differential expression effects. Applying these methods to a genome-wide CRISPRi screen of expressed genes in a K562 cell line, we find that, unexpectedly, most genetic perturbations have nonzero effects on a large swath of the transcriptome, with the median genetic perturbation affecting approximately a third of expressed genes. Very few of these effects are statistically significant individually, and significant effects capture only ~30% of the transcriptome-wide signal at current sample sizes, suggesting that significant gene lists are an incomplete summary of differential expression effects. Genes that are depleted of loss-of-function variants in the human population are also depleted of differential expression in response to most perturbations, implying that regulatory robustness parallels mutational constraint. Analyzing transcriptomic data from experiments that titrate mRNA and protein levels, we estimate dose-response curves for global transcriptomic effect, and find that transcriptome-wide effects are generally buffered against dosage changes: many critical genes have negligible transcriptome-wide effects at 50% dosage but large effects at lower dosage. We estimate the correlation of transcriptome-wide perturbational effects between cell types, and find that the median correlation of differential expression effects between K562 and RPE1 cells is 0.41, compared with a correlation of 0.93 for replicate experiments in the same cell type, suggesting substantial cell-type differences in gene regulatory networks. Our findings provide foundational insights into genome regulation and the interpretation of disease-associated genetic variation.

Session Title: Molecular Effects of Genetic Variation Poster Session III

PB1269 The nonalcoholic fatty liver disease promoting allele *GCKR(P446L)* promotes both lipid and glycogen accumulation in a human liver cell line.

Authors:

Y. Chen[1], B. Saini[1], J. Chen[2], B. Halligan[1], E. K. Speliotes[1], the GOLD consortium; [1]Univ. of Michigan, Ann Arbor, MI, [2]Northwestern Univ., Evanston, IL

Background: Nonalcoholic fatty liver disease (NAFLD) is a prevalent disease that affects about 30% of the US population. NAFLD is genetically influenced, and human genome wide association studies have identified a genetic variant in *GCKR(P446L)* that promotes the disease. *GCKR* encodes the glucokinase regulatory protein (GKRP) which is a negative regulator of glucokinase (GCK). Here we examine how *GCKR(P446L)* may promote liver disease. **Material and Methods:** Using lentiviral integration, we created stable human Huh-7 liver cell lines with wildtype and mutant P446L GCKR overexpressed or knocked down using shRNAs. We measured intracellular triglyceride and glycogen levels and GYS2, GCK, GKRP, FASN and ACC protein levels using Western blot analysis under various glucose concentrations and incubation durations. **Results:** *GCKR(P446L)* increased triglyceride and glycogen accumulation compared with *GCKR(WT)* overexpressed or control Huh-7 cells across different doses of glucose (0, 2, 5 and 25mM) after 8 hours of incubation ($p<0.0001$). The most significant difference in both lipid and glycogen accumulation between WT and P446L occurred at 2 mM glucose. We also found that overexpression of wild-type *GCKR* can significantly decreased glycogen content compared with control Huh-7 cells only at high glucose (25 mM) condition ($p<0.010$). Knockdown of *GCKR* also significantly increased cellular glycogen compared with non-target control Huh-7 cells at low concentrations of glucose (2 and 5 mM) after 8 hours of incubation. Western blot results showed that GCKR (P446L) variant significantly increased the protein level of glycogen synthase GYS2 about 2- and 3- fold as well as Fatty Acid Synthase (FASN) and Acetyl- CoA Carboxylase (ACC) and glucokinase (GCK) compared to cells expressing vector and GCKR(WT) at 2- and 5- mM glucose condition. Knockdown GCKR expression significantly increased GYS2 protein ~2 fold compared with non-target Huh-7 cells in 2 mM glucose ($p<0.0001$). **Conclusion:** Here we show that *GCKR (P446L)* increased both triglyceride and glycogen accumulation compared with *GCKR (WT)* in Huh-7 cells suggesting a cell autonomous effect on these endpoints. Overexpression of *GCKR (P446L)* mimics *GCKR* knockdown suggesting that the P446L mutation is a loss of function change. At least one-way *GCKR (P446L)* increases TG and glycogen synthesis in liver may be by increasing ACC and FASN levels to increase de novo lipogenesis and by up-regulating glycogen synthase GYS2 levels respectively. Further studies will determine how increased triglyceride and glycogen synthesis lead to the liver disease in patients carrying the *GCKR (P446L)* allele.

Session Title: Molecular Effects of Genetic Variation Poster Session I

PB1270 The outcome of next-generation sequencing in clinical diagnosis: Experience of Kuwait Medical Genetic Center

Authors:

L. Bastaki[1], R. M. Elshafie[1,2], M. Marafie[1], B. Albash[1], A. Alwadaani[1], S. BAHZAD[1], H. Alrohaif[1], S. Alkanderi[1], A. Alshammari[1], F. Mohammed[1], L. Aljeryan[1], M. Al-Ali[1], M. Al Ben Ali[1], N. M. Ali[1], D. Ibrahim[1], A. H. Aboelanine[1], S. Omar[1], S. Gouda[1], D. Marafi[1,3], H. Alsharhan[1,3], T. M. Kamal[1], L. AlDakheel[1], F. Alali[1], A. Kenawy[1], A. Alaqeel[1], A. Elmonairy[1], A. A. Ahmed[1], H. I. Albastawisy[1], A. Elshafey[1]; [1]Kuwait Med. Genetic Ctr., Ministry of Hlth., Sulaibikhat 80901, Kuwait, [2]Corresponding author, Kuwait city, Kuwait, [3]Dept. of Pediatrics, Faculty of Med., Kuwait Univ.., P.O. Box 24923, 13110 Safat, Kuwait

Introduction: Next-generation sequencing (NGS) techniques have facilitated human genome sequencing. The implementation of NGS in clinical practice has tremendously improved the diagnostic yield of rare disorders. Kuwait is an Arab Gulf country with a mixed population of native Kuwaitis, Arab and non-Arab residents. Before 2018, the Kuwait Medical Genetic Center (KMGC) at the Kuwait Ministry of Health offered a limited menu of genes for Sanger sequencing. In 2019, a collaborative effort was undertaken to develop a broad-based NGS sequencing menu serving the needs of clinicians all over Kuwait hospitals. This paper aims to share the experience of Kuwait Medical Genetic Center in the use of next-generation sequencing (NGS) gene panels and whole exome sequencing in the diagnosis of rare genetic disorders. To the best of our knowledge, no comprehensive data on the spectrum of variants associated with rare disorders in the Kuwait population has been published before. Methods and Subjects: This study is a retrospective review of 3857 sequencing tests, including both whole EXOME and NGS targeted panels, which have been performed at KMGC from February 2019 to May 2023. All patients had been referred to our laboratory for diagnostic testing, through clinical examination, and appropriate investigations. This study was approved by the Institutional Review Board of Kuwait Medical Genetic Center.

Results and Discussion: The data analysis results of this study have been tabulated, analyzed, and compared with that from other centers. Overall, across the 3857 tests, a definitive result was provided in 24.26%, a possible/probable result was provided in 22.66%, and a negative result was provided in 53%.

Conclusion: To the best of our knowledge, this is the first comprehensive report of this number of NGS tests in Kuwait. Our experience with 3857 sequencing tests suggests that analysis of similarly affected sibs with heterogeneous disorders reduces reporting of unnecessary variants of uncertain significance and facilitates the identification of candidate genes.

Session Title: Molecular Effects of Genetic Variation Poster Session II

PB1271 The small RNA transcriptome and its genetic regulation across human tissues

Authors:

P. Stojanov[1], T. Coorens[1], J. Fernandez del Castillo[1], S. Steelman[1], S. Young[1], C. Nussbaum[1], K. Ardlie[1], G. Getz[2], F. Aguet[3]; [1]Broad Inst. of MIT and Harvard, Cambridge, MA, [2]Broad Inst. at MIT and Harvard, Cambridge, MA, [3]Illumina, Foster City, CA

Population-based analyses mapping associations between genotypes and gene expression have identified expression quantitative trait loci (eQTLs) for almost all genes, enabling the identification of molecular mechanisms underlying genetic associations with complex traits and diseases. However, only a minority of complex trait associations have been linked to QTLs, partly due to limited power to detect regulatory effects in the relevant cell type, context, or molecular phenotype. Notably, standard RNA-sequencing protocols exclude small RNAs and preclude the study of thousands of small noncoding RNAs with essential roles in the post-transcriptional regulation of gene expression. Here, we present the characterization of small RNAs across 16,814 samples, 47 tissue sites and 978 donors in the GTEx Project, including microRNAs (miRNAs), Piwi-interacting RNAs (piRNAs), transfer RNAs, small nuclear RNAs, small nucleolar RNAs, Y RNAs, and others. We used personal transcriptomes and reference databases for alignment and applied stringent quality controls to minimize artifacts related to short sequences (~20-30nt), resulting in the quantification of 41,458 small RNAs. We used supervised classification to identify putative novel RNAs not present in references, and detected 57 novel high-confidence miRNAs. We observed strong tissue specificity for miRNAs, with other small RNAs showing homogenous expression across tissues. Small RNA expression is highly correlated with surrounding mRNAs or lncRNAs, suggesting shared transcriptional regulation. We mapped QTLs in cis and trans, identifying 100s to 1000s of cis-eQTLs for each RNA species. In line with tissue specificity of miRNA expression, miRNA cis-eQTLs also show a higher degree of tissue specificity than mRNA cis-eQTLs. For miRNAs and their predicted target genes, we used colocalization and mediation analyses to study how cis-regulatory effects on miRNA expression propagate to their target genes, in turn modulating their expression, and show how these effects link to complex trait associations.
In summary, we provide the largest resource of small RNA diversity and their eQTLs across human tissues. We demonstrate the importance of characterizing the full spectrum of small RNAs that play critical roles in the regulation of gene expression, including in development and disease.

Session Title: Molecular Effects of Genetic Variation Poster Session III

PB1272 The ulcerative colitis risk variant of adenylyl cyclase 7 has reduced cAMP signaling and promotes immune activation

Authors:

C. Cardinale[1], A. Strong[1], V. J. Watts[2], H. Hakonarson[1]; [1]Childrens Hosp. of Philadelphia, Philadelphia, PA, [2]Purdue Univ., West Lafayette, IN

Ulcerative colitis (UC) is an inflammatory condition of the large intestine. Genome-wide association studies identified a specific polymorphism in the adenylyl cyclase 7 gene (*ADCY7*; p.Asp439Glu) that confers a two-fold increased risk of UC, second in effect size only to the human leukocyte antigen (HLA) locus. The mechanism by which the p.Asp439Glu variant modulates disease risk is unknown. We now demonstrate that this variant has normal membrane localization in U-2 osteosarcoma cells, but impaired protein expression. Cyclic AMP production and CREB activation are also diminished ~40% by the risk variant. ADCY7 knockdown by siRNA, as a model for the risk variant in human primary CD4+ T cells, inhibits prostaglandin E2-mediated downregulation of proliferation, CD69 marker expression, and cytokine production. Our data suggest that the p.Asp439Glu variant confers increased risk of UC by inhibiting ADCY7-mediated cAMP production, thereby enhancing T cell activity. A direct-acting ADCY7 agonist may be a novel therapy for UC.

Session Title: Molecular Effects of Genetic Variation Poster Session I

PB1273 Tissue-specific mtDNA regulation of gene expression

Authors:

X. Giannoulis[1,2], S. Wengert[1], N. Cai[1,2]; [1]Helmholtz Pioneer Campus, Helmholtz Zentrum München, Neuherberg, Germany, Germany, [2]Sch. of Med., Technical Univ. of Munich (TUM), Munich, Germany, Germany

The mitochondrial DNA (mtDNA) encodes essential components of the oxidative phosphorylation pathway and RNAs that form parts of their translational machinery. However, gene regulatory effects of mtDNA and mtDNA-encoded genes have not been comprehensively explored. Using data from the GTEx consortium in 48 tissues from 684 individuals of European ancestry, we perform cis- and trans-eQTL analysis to investigate mtDNA-based gene regulation. First, we identify 309 significant mtDNA-cis-eQTLs (FDR 5%) for mtDNA-encoded genes across 38 out of 48 tissues tested, of which 108 are independent. Second, we identify 260 genome-wide significant associations (P<5x10-8) between nuclear DNA (nucDNA) SNPs and mtDNA gene expression (mtDNA-trans-eQTL), involving 210 unique nuclear loci. Of these, 150 are cis-eQTLs for nucDNA encoded genes (nuc-eGenes), and through mediation analysis, we find 117 of these nuc-eGenes mediate the mtDNA-trans-eQTL effects of their cis-eQTLs. For the rest, mtDNA-trans-eQTL effects are independent of their cis-eQTLs effects. Third, we find 108 mtDNA SNP associations with 68 nucDNA encoded genes (nucDNA-trans-eQTL) after controlling for the number of tissues and genes tested (P<1.25x10-9), across 25 out of the 48 tissues tested. Of these nucDNA-trans-eQTLs, 21 are cis-eQTLs for mtDNA encoded genes (mt-eGenes), and mediation analysis shows 20 of them mediate the nucDNA-trans-eQTLs effects of their cis-eQTLs. This demonstrates the potential existence of feedback-mediated co-regulation between mtDNA and nucDNA gene expression. Indeed, we find that nucDNA-eGenes with nucDNA-trans-eQTLs show significant enrichment in 75 pathways (72 of which are brain-specific), with top enrichments in RNA polymerase II activity, Biocarta-curated mitochondria pathways, and DNA-binding transcription factor activity. We also find that mtDNA-trans-eQTLs (across all tissues) overlap genes with significant GWAS hits on phenotypes such as height, BMI and metabolic syndrome, demonstrating the organismal relevance of these findings. Overall, our analysis builds the first comprehensive multi-tissue map of mtDNA and nucDNA cross-talk on gene expression regulation, a first step towards understanding the role of tissue-specific mito-nuclear interactions in health and disease.

Session Title: Molecular Effects of Genetic Variation Poster Session II

PB1274 Topic modeling uncovers cell states and context-specific genetic regulation in embryoid bodies

Authors:

M. Li[1], J. M. Popp[1], K. Rhodes[2], K. A. Barr[2], R. Jangi[1], B. J. Strober[3], K. Tayeb[2], Y. Gilad[2], A. Battle[1]; [1]Johns Hopkins Univ., Baltimore, MD, [2]Univ. of Chicago, Chicago, IL, [3]Harvard Sch. of Publ. Hlth., Boston, MA

Regulatory effects on gene expression often vary across cellular contexts such as cell types, states, and environments. However, discovery of context-specific regulatory effects is limited by our ability to efficiently assay diverse contexts of interest, such as those arising during cellular differentiation, as well as our ability to resolve the many biological processes which may be simultaneously active. Indeed, some interesting or disease-relevant cell states may not be known *a priori*. To address these challenges, we employed a unique experimental framework, embryoid bodies (EBs), in combination with topic modeling, an unsupervised machine learning method, to identify gene expression patterns reflective of diverse cell types and states. These EBs were derived from 53 individuals and are asynchronously differentiating cellular aggregates that give rise to numerous cell types. We collected single-cell RNA-sequencing data of these EBs that represent over 900,000 cells ranging from pluripotent stem cells to states resembling numerous differentiated fetal cell types. To model the complex, interwoven biological processes arising within this model system without attempting to define them *a priori*, we applied FastTopics, a tool for non-negative matrix factorization of count data, which learns patterns or "topics" in gene expression and models each cell as a weighted combination of these topics in a continuous and unbiased manner. We then characterized each topic using gene set enrichment analysis (GSEA) with the top differentially expressed genes. Our approach identified topics that represent both cell-type specific biological processes, such as glial cell differentiation, and processes shared across many cell types, such as mitosis. Finally, we applied CellRegMap using our learned topics to identify genotype-context interactions (GxC). Using this approach, we identified hundreds of eQTLs with evidence for GxC effects that are driven by multiple cellular contexts reflected by our topics. This study describes a flexible approach to identify latent cellular contexts in a uniquely broad *in vitro* model of cellular differentiation, and provides insights into regulatory changes governing diverse cell states.

Session Title: Molecular Effects of Genetic Variation Poster Session III

PB1275 Trajectory inference of multiple cellular lineages in embryoid bodies.

Authors:

R. Jangi[1], J. M. Popp[1], K. Rhodes[2], K. A. Barr[2], M. Li[1], B. Strober[3], K. Tayeb[2], Y. Gilad[2], A. Battle[1]; [1]Johns Hopkins Univ., Baltimore, MD, [2]Univ. of Chicago, Chicago, IL, [3]Harvard Sch. of Publ. Hlth., Brookline, MA

Experimental methods for studying cellular differentiation are rapidly advancing to the point where in vitro systems are capable of generating multiple cell types in a single framework. However, early and intermediate stages of differentiation for many cell types have remained understudied in human genomics due to limited analytic pipelines for singular trajectories generated from a multi-lineage system. Entire trajectories can reveal patterns of gene expression and regulation that are transient during early stages of differentiation, but have phenotypic consequences in later stages. With distinct trajectories, we can capture temporally dynamic genetic regulation of gene expression (dynamic expression quantitative trait loci or dynamic-eQTLs) specific to each lineage. To study these effects, computational methods to infer distinct trajectories from multi-lineage systems are needed.To capture diverse trajectories, we used embryoid bodies (EBs), three-dimensional cellular aggregates of pluripotent stem cells that differentiate into various cell types. As differentiation occurs at different speeds for these cells, we can characterize multiple developmental states and dynamic regulatory processes at a high resolution. We collected single-cell RNA sequencing data from human EBs across 53 donors, representing over 900,000 cells, and confirmed the presence of all 3 germ layers in EBs with canonical marker gene expression. We first verified the presence of multiple stages of differentiation based on marker gene expression established by previous work. We then developed a pipeline for identifying trajectories of interest based on scDRS, a method that generates cell-level scores by testing for enrichment of trajectory-relevant gene expression compared to mean and variance matched control gene expression. We applied this approach to a trajectory from each germ layer. We began with cardiomyocytes, a previously characterized cell type in a directed differentiation system, identifying cells along the cardiomyocyte trajectory.We mapped dynamic-eQTLs along this lineage and identified 303 significant dynamic-eQTLs, including regulatory effects that were replicated in a previous study and novel regulatory effects. We also identified a neuron trajectory, and characterized 574 dynamic effects. Finally, we identified a hepatoblast trajectory with downstream analysis in progress. We will expand this pipeline to identify trajectories for more cell types and identify novel dynamic-eQTLs. This work will capture transient regulatory effects in diverse developmental cellular trajectories to characterize the differentiation landscape.

Session Title: Molecular Effects of Genetic Variation Poster Session I

PB1276 Transcriptome wide Association Study of Activated Monocytes Identifies NETO2 and CRISPLD2 as Potential Alzheimer's disease Resilience Genes

Authors:

Y. Mustafa[1], W. Bush[2], T. Hohman[3], L. Dumitrescu[4], L. Main[5], J. Haines[2]; [1]Case Western Reserve Univ., Cleveland, OH, [2]Case Western Reserve Univ, Cleveland, OH, [3]Vanderbilt Univ Med Ctr., Nashville, TN, [4]Vanderbilt Univ. Med. Ctr., Nashville, TN, [5]Case Western Reserve Univ., Lakewood, OH

Alzheimer's disease (AD) is a highly heritable neurodegenerative disease of aging and the most common cause of dementia. A potential paradox of AD pathogenesis is that the hallmark protein aggregations consistent with AD pathology are commonplace in the elderly, yet only some manifest clinical symptoms. This fact, along with the limited clinical benefit seen from drugs which target these aggregations suggest that mechanisms of AD-resilience are a potential avenue for identifying new therapeutic targets. GWAS of AD have been extremely successful in identifying AD-risk loci, and have provided substantial evidence that innate immune cells, particularly myeloid cells, have a substantial role in AD development but are unexplored with respect to AD-resilience. We hypothesized that genetic variants influence the function and response of monocytes through changes in gene expression, contributing to differences in AD-resilience outcomes. To test this hypothesis, we performed a transcriptome-wide association study (TWAS) of monocytes in four activation states. Monocyte genetic and expression data (Fairfax et al. 2014) were processed and used to build elastic net-regression models of monocyte genetically-regulated expression in the naive state, 2- and 24-hours following LPS stimulation, and 24-hours following IFN-$\boldsymbol{\gamma}$; stimulation. These conditions are intended to simulate monocyte response to innate and adaptive immune challenges. Weights for eQTL effects were subsequently used by S-PrediXcan against summary statistics of an AD case-only meta-analysis of global resilience to AD (Dumitrescu et al. 2020). Using TWAS, we identified two genes with Bonferroni-significant association to AD-resilience (among cases only): NETO2 (LPS24h, $P < 5.0x10\text{-}7$) and CRISPLD2 (IFN24h, $P < 6.0x10\text{-}6$), each in a different context of inflammation. Both results suggest that genetic predisposition to lower expression of these genes in monocytes during inflammation is associated with increased global resilience to AD. We also validate our findings in a cohort of Indiana and Ohio Amish using a different but related "cognitive preservation" phenotype, with the NETO2 association replicating in this population with a consistent direction of effect ($P < 8.0x10\text{-}4$). Lastly, a colocalization analysis using SuSie demonstrated that TWAS-significant regions had a distinct eQTL structure that is not present in other cell types including microglia, neurons, endothelial cells, and oligodendrocytes--suggesting that this association is monocyte-specific. The association of these genes suggest a possible role of monocyte-driven pathogenic inflammation in AD.

Session Title: Molecular Effects of Genetic Variation Poster Session II

PB1277 Transferrin isoform analysis as a functional test for genetic variation in *PMM2*

Authors:

M. Schultz, K. Liedtke, C. Turgeon, A. White, G. pino, D. Peck, A. Studinski, P. L. Hall, D. Matern; Mayo Clinic, Rochester, MN

Introduction: Clinical testing of transferrin glycosylation is a standard screening test for patients with suspicion of an inborn error of N-linked glycosylation. The role of this test is expanding to include characterization of genetic variants detected in molecular testing. Individuals with the most common CDG, PMM2-CDG (70% of CDG cases), exhibit an abnormal transferrin profile; however, there are a few reports of false negative results. A better understanding of transferrin clinical sensitivity informs its value as a diagnostic test, functional evaluation for novel variants, and method to exclude potentially undetected variants. We present our experience with transferrin isoform analysis for PMM2-CDG to better understand its clinical sensitivity in the most common CDG type.
Methods: This cohort was ascertained by follow up of clinical results or by reported diagnosis. Clinical transferrin testing is performed by intact protein profiling by mass spectrometry. We make an implicit assumption that the laboratory would be contacted if a patient was diagnosed with PMM2-CDG after receiving negative transferrin results.
Results: We reviewed diagnostic testing results from individuals with PMM2-CDG (n=89). The transferrin profile was abnormal or indeterminate in 85/89 cases (clinical sensitivity = 95.5%). Two of the results were indeterminate and four were true false negatives by the assay's established cutoff. Closer examination of these individuals revealed known pathogenic variants, notably 3 individuals harboring c.722G>C, p.C241S, two with a splicing alteration c.178G>T, p.V60L, and one with a promoter alteration, c.-167G>T. In cases with a single pathogenic variant (n=18), none had positive transferrin results and were resolved as confirmed carriers.
Conclusions: These data demonstrate that transferrin analysis has >95% clinical sensitivity for PMM2-CDG and is a reliable functional test for genotypes of uncertain significance and to exclude the possibility of undetected second variants in *PMM2*. Three pathogenic variants observed in multiple cases with normal or indeterminate transferrin isoform profiles highlight an important limitation of the test which can be addressed by prior knowledge of their genotype-phenotype correlation. Provided there are no other clinically relevant molecular findings in the N-glycosylation pathway, positive transferrin results could be considered PS3 criteria for functional evidence in variant classification (for a VUS in trans with a P/LP variant) and consistent with a diagnosis of PMM2-CDG.

Session Title: Molecular Effects of Genetic Variation Poster Session III

PB1278 treX-QTL: A mixture-model for identification and characterization of trans-eQTLs

Authors:

C. Wu[1], T. Xu[1], D. Munro[1,2], P. Mohammadi[1,2], A. Palmer[1], A. Goren[1], M. Gymrek[1]; [1]Univ. of California, San Diego, San Diego, CA, [2]Univ. of Washington, Seattle, WA

Studies of expression quantitative trait loci (eQTLs) aim to discover genetic variants that explain variation in gene expression, and may be linked to complex traits or diseases. While thousands of cis-eQTLs have been identified, robust detection of trans-eQTL effects is challenging mainly due to insufficient statistical power. Thus, biological mechanisms and characteristics of trans-eQTLs remain largely unknown.

We introduce treX-QTL, a novel trans-eQTL detection method that improves statistical power over methods that test gene-by-variant pairs separately by jointly modeling effects of an individual variant across all genes. treX-QTL uses a biologically plausible mixture model of effect sizes that enables inference of the number of target genes of a variant. To benchmark and evaluate effects of trans-eQTLs detectable by our method, we developed a simulation framework that models parameters including effect sizes, number of target genes of trans-eQTLs, and effects of covariates. Using our simulation framework we show that widely used correction techniques such as PCA or PEER remove effects of true trans-eQTLs along with technical variation.

We tested treX-QTL with a publicly available yeast expression dataset of 1,012 meiotic segregants. We identify 367 unique variants acting as trans-eQTLs at FDR 5% and replicate 3 hotspots of trans-eQTLs, mostly found in transcription factors, which treX-QTL predicted to affect over half of expressed genes. Next, we applied treX-QTL to a RatGTEx brain dataset that consists of 340 rats. We identified dozens of novel trans-eQTLs with target genes throughout the genome. To identify candidate genes driving these signals, we performed motif analysis of the promoters of target genes. For one chr7 trans-eQTL hotspot overlapping the key neuronal transcription factor Neurod4, we found target genes were enriched for Neurod motifs (p=1e-11), supporting it as the causal gene.

Finally, we applied treX-QTL to expression data for 52 tissues in humans from the GTEx cohort. While this analysis did not return strong trans-eQTL candidates, our power analyses suggest the GTEx dataset is still substantially underpowered to detect even moderately strong trans-eQTLs. Overall, our framework provides an important resource for future trans-eQTL studies involving humans and other complex organisms.

The treX-QTL package and simulation framework can be found online.

Session Title: Molecular Effects of Genetic Variation Poster Session I

PB1279 Uncovering sources of human gene expression variation in a globally diverse cohort

Authors:

D. Taylor, M. Tassia, R. McCoy; Johns Hopkins Univ., Baltimore, MD

Gene expression variation accounts for a large proportion of phenotypic variation within and between species and is a major contributor to human traits and diseases. Previous large-scale studies of human gene expression have been invaluable for revealing mechanisms of genome function but have largely focused on individuals of European ancestries. This bias diminishes the generalizability of results to understudied populations and limits broader understanding of the diversity and evolution of gene expression across populations. To address these challenges, we performed RNA sequencing of lymphoblastoid cell lines derived from 731 globally-diverse individuals from the 1000 Genomes Project. Quantification of gene expression levels and alternative splicing revealed that the vast majority of gene expression variance (92%) and splicing variance (95%) is partitioned within as opposed to between populations-mirroring patterns of human genetic diversity. Additionally, gene expression variance within populations decreases with migratory distance from Africa, consistent with the serial founder effects during historical human expansion across the globe. Interestingly, gene expression variance was also lower for genes under stronger evolutionary constraint, supporting a model of stabilizing selection shaping expression variability. Using existing high-coverage whole-genome sequencing data from the same individuals, we leveraged the global diversity of our dataset to identify cis-eQTLs and cis-sQTLs in >15,000 and >7,000 genes, respectively. In addition to replicating the results of previous large-scale e/sQTL studies, we also identified novel associations, including many variants specific to Asian or Admixed American samples which were under-represented or absent in previous molecular association studies. Interestingly, we observed that genes with population-stratified expression are more likely to have population-stratified eQTLs, thereby explaining observed differential expression between populations. Taking advantage of the diverse history of recombination events in our sample, we performed fine-mapping to identify candidate causal expression and splice-altering variants, revealing evidence of widespread allelic heterogeneity (nearly 4,000 genes with >1 independent eQTL [40% of genes with finemapped eQTLs]) and strong enrichment within annotated regulatory elements. Together, our study reveals key sources of gene expression variation and its evolution across diverse human populations and provides a powerful resource for the field of human functional genomics.

Session Title: Molecular Effects of Genetic Variation Poster Session II

PB1280 Uncovering the brain-specific genetic regulation of splicing by mapping splicing quantitative trait loci in 13,061 post-mortem brain RNA-seq samples.

Authors:

A. Real[1], K. BP[2], B. Z. Muller[2], W. H. Cuddleston[2], J. Humphrey[2], T. Raj[2], D. A. Knowles[1,3]; [1]New York Genome Ctr., New York City, NY, [2]Icahn Sch. of Med. at Mount Sinai, New York City, NY, [3]Columbia Univ., New York City, NY

Alternative splicing enables genes to generate diverse transcript isoforms and multiple proteins with altered structural or functional properties, by selecting different combinations of exons. It is tightly regulated in a cell- and tissue-specific manner. In the brain, splicing contributes to neuronal diversity and plasticity across brain regions and its dysregulation is linked to neurological disorders, yet the underlying mechanisms remain unclear. We investigated splicing landscapes in the brain to understand their genetic regulation across tissues and involvement in neurodevelopmental and neurodegenerative processes. Our "BigBrain" resource harmonized 13,061 unique RNA-sequencing samples from 12 published datasets across 7 brain regions to allow meta-analyses of multiple brain regions and ancestry-specific cohorts for cis- and trans- expression and splicing quantitative trait loci (eQTL and sQTL). To detect splicing patterns between samples from different brain cohorts, we used the LeafCutter algorithm to cluster splicing junctions in a cohort- and tissue-aware approach. Initial meta-analysis mapped 58,735 cis-sQTLs for 28,957 genes (FDR 5%) in 4,342 individuals. However, principal component analysis identified cohorts and brain regions as key drivers of gene expression and splicing differences across the BigBrain tissues. To account for this variation systematically across all BigBrain cohorts and achieve optimal data harmonization, we compare batch correction and meta-analysis approaches, holding out one cohort to evaluate sQTL replication rate. Specifically, we compare 1) global batch correction across all datasets vs per tissue-cohort correction, and 2) fixed-effect, random-effect, and multivariate adaptive shrinkage-based meta-analysis vs mega-analysis across cohort-tissue pairs. Mega-analysis combines raw data from multiple studies, potentially resulting in a larger sample size and increased statistical power compared to meta-analysis, enabling more precise estimates and improved detection of subtle splicing effects. It allows exploration of subgroup effects and interactions not possible in a meta-analysis providing a comprehensive and robust approach to synthesizing data, yielding more reliable and generalizable results. Finally, we perform colocalization analysis to enhance the interpretation of neurological GWAS loci. Harmonizing published transcriptomics datasets into the BigBrain project will shed light on multiple aspects of gene expression, alternative splicing, and its genetic regulation, enhancing our understanding of splicing genetic regulation in different brain regions.

Session Title: Molecular Effects of Genetic Variation Poster Session III

PB1281 Uncovering the role of actin cytoskeletal dynamics in seizures caused by *TIAM1* loss

Authors:

S. Lu[1], M. Ma[1], F. A. Blanco[1], M. Gu[1], D. Dutta[1], Y. Huang[1], X. Pan[1], H. Li[1], W. K. Chung[2], K. F. Tolias[1], H. J. Bellen[1]; [1]Baylor Coll. of Med., Houston, TX, [2]Columbia Univ., New York, NY

Epilepsy, defined by recurrent seizures, is one of the most common neurological diseases affecting 50 million people worldwide. People with epilepsy often have comorbid conditions including learning and memory defects, anxiety and depression. The mechanism underlying epilepsy remains elusive. Since hyper-excitability of neuronal synaptic transmission can cause seizures, much work has focused on synaptic transmission defects in the pathogenesis of epilepsy. However, little attention has been given to the role of actin cytoskeleton in epileptogenesis. Actin cytoskeletal dynamics are essential for the physiological properties of neurons, and growing evidence has indicated that altered cytoskeletal dynamics may contribute to epilepsy in general. Aberrant expression levels of various actin cytoskeletal proteins are known to be involved in epilepsy. Moreover, mutations in actin-associated proteins, including Filamin A (OMIM: 300049), PAK1 (OMIM: 618158), alpha-II spectrin (OMIM: 613477), and DIAPH1 (OMIM: 616632), cause epilepsy. Recently, we have identified various proteins that are associated with seizures in humans, and these proteins are involved in the regulation of actin dynamics. Among them, bi-allelic loss-of-function (LoF) variants in *TIAM Rac1 associated GEF 1* (*TIAM1*) cause a neurodevelopmental disorder with language delay and seizures (NEDLDS, OMIM: 619908). TIAM1 is a Rac1-specific guanine nucleotide exchange factor (GEF), activating signaling pathways that regulate actin cytoskeleton organization within neurons. *TIAM1* and its fly ortholog, *still life* (*sif*), are enriched in the central nervous system (CNS). Loss of *sif* in flies leads to reduced survival and severe seizure-like behaviors. Especially, we found that *sif* is mainly expressed in neurons rather than glia in the fly CNS, and the seizure-like behaviors associated with *sif* LoF are mainly caused by neuronal loss of *sif*. Notably, *sif* loss in excitatory neurons, specifically glutamatergic neurons, causes seizure-like behaviors. Interestingly, *sif* loss in glutamatergic neurons resulted in significant mitochondrial accumulation in axons and increased mitochondria in synaptic terminals. These findings suggest that *TIAM1*/*sif* loss may cause mitochondrial trafficking defects by disrupting actin cytoskeletal dynamics, thus leading to aberrant neuronal excitability and the occurrence of seizures in NEDLDS patients. Further investigations into the molecular mechanisms may provide valuable insights into novel therapeutic strategies for the treatment of epilepsy.

Session Title: Molecular Effects of Genetic Variation Poster Session I

PB1282 Understanding the role of Hippo effectors Yap1 and Wwtr1 in optic fissure fusion.

Authors:

U. NEELATHI, D. Sanchez-Mendoza, B. Brooks; NIH, Bethesda, MD

The developing eye forms an optic fissure (OF) on its ventral side, the edges of which fuse to form a continuous layer of retinal pigment epithelium (RPE) and neural retina (NR). Failure of the OF to close causes uveal coloboma, which can affect iris, ciliary body, retina, choroid and optic nerve. It is genetic disorder which accounts for 10% cases of childhood blindness. Several genes and genetic pathways are linked to coloboma. Mutations in YAP1, an effector of Hippo signaling downstream of *FAT1* (another coloboma-associated gene), have been described in patients with syndromic and non-syndromic coloboma. WWTR1 (TAZ) is another effector of Hippo signaling and together with *YAP1* serves as a transcriptional coactivator via *TEAD* transcription factors. These observations prompted us to further study the role of Hippo signaling in OF fusion in a zebrafish model. *yap1* and *wwtr1* are expressed in RPE, OF and periocular mesenchyme (POM). *wwtr1* mutants exhibit grossly normal eye development without coloboma. *yap1* homozygous mutants have variable phenotype from bilateral coloboma to unilateral coloboma and/or microphthalmia. The double homozygous mutants have developmental arrest at approximately 16hpf, making it difficult to study OF fusion. Compared to *yap1-/-* mutants, *yap1-/-; wwtr1+/-* embryos have a higher penetrance of coloboma and pigmentation defects. In yap1-/-; wwtr1+/- mutants, expression of *aldh1a2* (dorsal), *vax2* (ventral), *foxG1a* (nasal) and *foxD1* (temporal) does not change compared to WT embryos, indicating the coloboma in these mutants is not due to defective optic cup morphogenesis. The OF lips appose but do not fuse in these mutants, as seen by the retention of laminin (basement membrane) at 48hpf and 72hpf. While characterizing the OF edges in these mutants by *ntn1a* expression, we have observed that there is a tissue between the OF edges. This was also confirmed by *dct* expression, where we can see the melanosomes being excluded at and around OF edges. The tissue between the OF was not optic stalk (OS) in origin as we did not see any upregulation of *fgf8* transcript. Expression of RPE markers *mitfA* and *tfec* is down regulated in and around the OF edges. Upregulation of early NR maker *pax6* at 24hpf and expression of late NR markers *Alcama and Huc/D,* which mark the retinal ganglion cells and amacrine cells respectively, in between the OF edges at 72hf, imply that the tissue in between the OF is NR in origin. Our observations imply that in *yap1-/-; wwtr1+/-* mutants the OF fails to fuse due to formation of NR in between the OF edges. We are investigating whether *yap1* and *wwtr1* directly regulate the expression of these RPE genes resulting in NR formation between the OF lips.

Session Title: Molecular Effects of Genetic Variation Poster Session II

PB1283 Unique genetic architecture of CSF and brain metabolites pinpoints the novel targets for the traits of human wellness

Authors:

C. Wang[1], D. Western[1], C. Yang[2], M. Ali[1], L. Wang[3], P. Gorijala[2], J. Timsina[1], A. Ruiz[4], P. Pastor[5], M. V. Fernandez[1], Dominantly Inherited Alzheimer Network (DIAN), Alzheimer's Disease Neuroimaging Initiative (ADNI), D. Panyard[6], C. Engelman[7], Y. Deming[7], M. Boada[4], A. Cano[4], P. Garcia-Gonzalez[4], N. Graff-Radford[8], H. Mori[9], J-H. Lee[10], R. Perrin[1], Y. Sung[1], C. Cruchaga[1]; [1]Washington Univ. in St Louis, St Louis, MO, [2]Washington Univ. in St Louis, st louis, MO, [3]Washington Univ., St Louis, MO, [4]Res. Ctr. and Memory Clinic, ACE Alzheimer Ctr. Barcelona, Barcelona, Spain, [5]Unit of Neurodegenerative diseases, Dept. of Neurology, Univ. Hosp. Germans Trias i Pujol, Barcelona, Spain, [6]Stanford Univ., Sunnyvale, CA, [7]Univ. of Wisconsin-Madison, Madison, WI, [8]Dept. of Neurology, Mayo Clinic, Jacksonville, FL, [9]Dept. of Clinical NeuiroSci., Faculty of medicie, Osaka, Japan, [10]Univ. of Ulsan Coll. of Med., Asan Med. Ctr., Seoul, Korea, Republic of

Unique genetic architecture of CSF and brain metabolites pinpoints the novel targets for the traits of human wellness

Background Perturbation in brain metabolism has been identified as an essential part of neurodegeneration and other brain-related disorders. Many metabolites and human traits harbor genetic components. Several studies have identified plasma and urine metabolites contributing to various traits and disorders based on genetic evidence. To study neurological traits, brain and cerebrospinal fluid (CSF) are tissues more relevant than plasma or urine. However, no study has thoroughly examined the genetic regulators of brain and CSF metabolite levels that could be used to identify causal metabolites of neurological disorders. The genetic architecture of metabolite levels in central nervous system can be different from those in circulating tissues due to blood-brain barrier and the metabolism of brain cells. We, therefore, pursued a large-scale brain and CSF metabolite genome-wide association studies (MGWAS) to identify metabolites contributing to 27 brain-related traits or disorders.

Methods We measured 440 metabolites in 2311 CSF samples (Knight ADRC, DIAN, ADNI, Barcelona-1, Fundació ACE) and 962 metabolites in 1016 brain samples (Knight ADRC, ROSMAP, MAYO). We identified MQTL and performed metabolome-wide association study (MWAS) to identify metabolites associated with traits. Mendelian randomization (MR) and colocalization were performed for causality and shared genetic regulations between metabolite abundance and 27 human wellness traits.

Results We identified 219 independent associations (59.8% novel) for 144 CSF metabolites and 36 independent associations (55.6% novel) for 34 brain metabolites. Most of the novel signals (97.7% and 70.0% in CSF and brain) were tissue specific. The nominated functional genes for 42.9% of CSF associations and 34.4% of brain associations encoded cis-acting proteins to the associated metabolites. By integrating MWAS-FUSION, MR, and colocalization, we identified eight metabolites to be causal for eight brain-related human wellness traits (11 relationships). We found that low mannose level was causal to bipolar disorder and as dietary supplement it may provide therapeutic benefits. Low galactosylglycerol level was found causal to Parkinson's disease.

Conclusion We conducted the first large-scale CSF and brain genome-wide association study. Our findings expand the knowledge of MQTL in central nervous system, provide insights into human wellness, and successfully demonstrates the utility of combined statistical approaches to inform interventions.

Session Title: Molecular Effects of Genetic Variation Poster Session III

PB1284 Unraveling Functional Effects of Curly Su Mutations in Drosophila Melanogaster: Computational Mutagenesis and Genetics Insights

Authors:

S. Teng, A. Sobitan; Howard Univ., Washington DC, DC

The Curly Su protein (dMPO), a homolog of human myeloperoxidase (hMPO), plays a crucial role in wing development in Drosophila melanogaster. Through its generation of reactive oxygen species (ROS), dMPO contributes to diverse cellular and physiological processes. Given the similarity between the sequences of dMPO and hMPO, dMPO presents an exceptional opportunity for experimental validation in studies related to development and immunity. In this study, we employed saturated computational mutagenesis to investigate the effects of specific mutations on both dMPO and hMPO. By analyzing predicted folding energy changes, we identified target mutations, with G378W emerging as the most destabilizing mutation for dMPO. To validate the impact of these mutations, including G378W and Deletion 305-687, we generated transgenic fruit flies using genome editing techniques. Our observations revealed that G378W influenced wing phenotypes and the overall lifespan of the samples during husbandry. To gain further insights, we performed transcriptome analysis using RNAseq on both transgenic and wild-type samples. By comparing differentially expressed genes (DEGs) between treatment groups, we conducted gene ontology analysis to elucidate the functional attributes of down-regulated and up-regulated genes. The combined use of computational tools and genetics experiments allowed for rapid discovery of the novel functional effects of missense mutations in target proteins. This integrated approach provides valuable insights into the intricate relationships between genetic alterations and phenotypic outcomes.

Session Title: Molecular Effects of Genetic Variation Poster Session I

PB1285 Unraveling the role of MPDZ SNP rs139926461 in osteoblast function and bone mineral density (BMD) using prime editing.

Authors:

A. Buo, K. Perumal, L. Dillard, C. Farber; Univ. of Virginia, Charlottesville, VA

BACKGROUND: Osteoporosis, defined by low bone mass and increased fracture risk, burdens 200+ million globally and incurs over $50 billion in healthcare costs due to fragility fractures. Research efforts have increasingly focused on elucidating the genetic underpinnings of bone health, with gene-editing methods offering potential to validate causal genetic variations. Despite this potential, these methods are underutilized in bone research.
The Multiple PDZ Domain Crumbs Cell Polarity Complex Component (*MPDZ*) has emerged as a candidate gene of interest due to compelling genetic evidence implicating its role as the causal driver of the bone mineral density (BMD) association at chromosome 9p23. Our prior research corroborated its role in regulating bone mass accrual in mice. However, the causal variant influencing *MPDZ* function is still unidentified, justifying gene editing of candidate variants.
METHODS: We used Prime Editing, a precise CRISPR-based DNA editing tool, to edit a BMD-associated *MPDZ* SNP, rs139926461. This SNP, an expression quantitative trait locus for *MPDZ* that is linked with BMD, is located in the *MPDZ* promoter. The C allele of rs139926461 is associated with lower MPDZ expression in comparison to the reference T allele. We changed the T to a C, hypothesizing that this substitution affects *MPDZ* transcription. Human fetal osteoblasts (hFOBs), which are homozygous for the T allele, were transfected with plasmids designed to direct the desired prime edited-mediated substitution. Cells were sorted for GFP to enrich the population of potentially edited cells.
RESULTS: Targeted sequencing of the *MPDZ* promoter region revealed that 20% of the DNA reads from these cells were successfully edited from T to C at rs139926461. Single-cell clonal populations were obtained and RNA was extracted from five clones for allele-specific expression analysis. Among these, only one clone, 3A11, showed a notable deviation in the C-to-T ratio at a separate heterozygous exonic SNP, rs34605667. In contrast to the approximate 50:50 C-to-T balance seen in other clones, the 3A11 clone showed a 36:64 C-to-T distribution, suggesting a distinct allelic imbalance. Sanger sequencing of this clone confirmed a heterozygous T-to-C edit at rs139926461, linking the observed allelic imbalance to the presence of the C at this SNP.
SUMMARY: These results provide evidence that the rs139926461 C allele impacts *MPDZ* gene transcription, thereby accounting for part of the BMD-associated osteoporosis risk. Further, this marks the first successful use of prime editing in bone cells, offering a promising approach for future exploration of osteoblast function and skeletal phenotypes.

Session Title: Molecular Effects of Genetic Variation Poster Session II

PB1286 Unravelling Pan-Cancer Susceptibility at 5p15.33: Functional Characterization of a Novel VNTR Variant

Authors:

A. O'Brien[1,2], H. Kong[1], M. Patel[1], K. Connelly[1], M. Xu[1], I. Collins[1], J. Zhong[1], J. Hoskins[1], M. Ho[1], B. Papenberg[1], H. Patel[1], T. Luong[1], International Lung Cancer Consortium (ILCCO), Bladder Cancer Consortium, The Pancreatic Cancer Case-Control Consortium (PanC4), The Pancreatic Cancer Cohort Consortium (PanScan), Melanoma Meta-Analysis Consortium, M. Iles[3], M. Law[4], M. T. Landi[5], R. Stolzenberg-Solomon[5], B. Wolpin[6], N. Orr[2], A. Klein[7], D. Li[8], W. Luo[9], K. Jones[9], S. Lindstroem[10], M-H. Stern[11], S. Chanock[5], J. Shi[5], L. Prokunina-Olsson[1], J. Choi[1], K. Brown[1], L. Amundadottir[1]; [1]Lab. of Translational Genomics, Natl. Cancer Inst., Div. of Cancer Epidemiology and Genetics, Rockville, MD, [2]Patrick G Johnston Ctr. for Cancer Res., Queen's Univ. Belfast, Belfast, United Kingdom, [3]Leeds Inst. for Data Analytics, Sch. of Med., Univ. of Leeds, Leeds, United Kingdom, [4]QIMR Berghofer Med. Res. Inst., Brisbane, Australia, [5]Natl. Cancer Inst. - Div. of Cancer Epidemiology and Genetics, Rockville, MD, [6]Dept. of Med. Oncology, Dana Farber Harvard Cancer Ctr., Boston, MA, [7]Dept. of Oncology Pathology and Med., Sidney Kimmel Comprehensive Cancer Ctr., Johns Hopkins Sch. of Med., Baltimore, Baltimore, MD, [8]Dept. of Gastrointestinal Med. Oncology, Univ. of Texas MD Anderson Cancer Ctr., Houston, TX, [9]Cancer Genomics Res. Lab. - Natl. Cancer Inst., Rockville, MD, [10]Univ. of Washington, Seattle, WA, [11]Inst. Curie, Paris, France

Genome-wide association studies (GWAS) have identified independent signals at 5p15.33 across numerous cancers, with protective alleles for one cancer often conferring risk for another. Many of these associations are thought to act via allele-specific alterations in the *cis*-regulation of target genes. eQTL and TWAS in multiple tissue types implicate two plausible target genes: *TERT* and *CLPTM1L*. However, the mechanisms linking cancer susceptibility alleles to their presumed target genes, as well as the divergent directions of association across different cancers, remain elusive. With this in mind, we employed a comprehensive approach integrating statistical fine-mapping of GWAS summary statistics with massively parallel reporter assays (MPRA) and a densely tiled CRISPRi screen implemented across eight cell lines, representing four distinct cancers associated with 5p15.33 alleles - melanoma, lung, pancreatic, and bladder cancer.

Integrative screening identified up to 30 variants across multiple signals in the locus with putative allele-specific transcriptional activity. Furthermore, a variable number tandem repeat (VNTR) in intron 9 of CLPTM1L emerged as a potent transcriptional enhancer in CRISPRi screening. Leveraging PacBio sequencing, VNTR alleles of European ancestry samples from the 1000 Genomes Project were genotyped and found to be highly polymorphic, with longer VNTRs linked to the haplotype tagged by rs31490-A (risk increasing for melanoma and pancreatic cancer, protective for lung cancer). Subsequent imputation into a pancreatic cancer GWAS of 9,040 cases and 12,496 controls revealed an association between longer VNTRs and risk of pancreatic cancer (OR for VNTR in the 75th percentile = 1.17, 95% CI = 1.13-1.22, P = 2.21x10-12). This association of the lead SNP rs31490 is no longer genome-wide significant upon conditioning on the VNTR genotype (P= 2.20x10-5), suggesting that these VNTR alleles merit consideration as part of the credible variant set within this multi-cancer-associated haplotype.

Initial observations suggest VNTR-mediated *cis*-regulation of *TERT* in pancreatic cancer cell lines. Ongoing experiments in relevant cell lines utilize luciferase reporter systems and DNA-protein binding assays with the aim of not only deciphering how longer VNTRs differentially exert their transcriptional effects compared to shorter alleles but also characterizing the divergent directions of effect between different cancer types. These findings underscore the proposition that cancer susceptibility at the 5p15.33 locus may be mediated by both SNPs and an uncharted class of variants, yet to be characterized in the post-GWAS landscape.

Session Title: Molecular Effects of Genetic Variation Poster Session III

PB1287 Unravelling the Genetic and Phenotypic Heterogeneity of Red Cell Membranopathies in Indian Population Using Next Generation Sequencing.

Authors:

T. MORE[1], P. Kedar[2]; [1]INDIAN COUNCIL OF Med. RESEARCH-Natl. Inst. OF IMMUNOHAEMATOLOGY, MUMBAI, India, [2]Natl. Inst. of Immunohematology (ICMR), MUMBAI, India

Red cell membranopathies refers to genetically and phenotypically heterogenous disorders due to defects in red cell membrane and skeletal proteins. Red cell membranopathy is a common cause of hemolytic anemia varying from mild to severe hemolysis due to defects in red cell membrane protein genes. Causative genes implicated in RBC membranopathies are numerous, making a gene-by-gene approach time consuming, expensive and labor intensive. Use of Next Generation Sequencing (NGS) targeted sequencing panel can expedite the molecular diagnosis after routine laboratory tests. This study reports the genetic and phenotypic heterogeneity of 124 Indian patients with red cell membranopathies using targeted next-generation sequencing and studied the correlation between the identified variants with their corresponding phenotypic features. We diagnosed 104 Hereditary Spherocytosis (HS) patients harbouring 39 *ANK1*, 47 *SPTB*, 6 *SLC4A1* AND 7 *SPTA1* variants. A total of 74 novel and 25 reported variants has been detected causing HS. Also, we identified 7 patients of Hereditary Xerocytosis (HX) harbouring 4 novel and 4 reported *PIEZO1* variants. Eight patients of Hereditary Elliptocytosis (HE) and 5 patients of Hereditary Pyropoikilocytosis (HPP) were diagnosed harbouring 9 novel and 4 reported *SPTA1* variants. All of the detected variants have damaging effect on the protein stability and function, as shown by the *in silico* analysis. The possible effect of the detected variants on the protein structure was studied using the HOPE software and DynaMut tools. This report mainly illustrates the molecular heterogeneity of red cell membranopathies in Indian patients. Comprehensive genetic and phenotypic evaluation assists in implementing the knowledge of genetic patterns and spectrum of variants, providing a molecular support for diagnosis of red cell membranopathies. Our experience demonstrates the high diagnostic yield of NGS for molecular diagnosis of RBC membrane disorders associated with defects in large genes. Timely detection of the disorder not only offers therapeutic benefits for patients but is likely to help in genetic counselling and future antenatal diagnosis, if required.

Session Title: Molecular Effects of Genetic Variation Poster Session I

PB1288 Unveiling the Functional Relevance of Protein Coding Short Tandem Repeats in APOB and E2F4 through Structural Modeling and Analysis

Authors:

X. Zhang, J. Margoliash, M. Gymrek; Univ. of California San Diego, La Jolla, CA

Short Tandem Repeats (STRs) are repetitive DNA sequences found throughout the human genome and are of significant interest due to their potential role in biological processes and diseases. We previously performed an STR-based genome wide association study (GWAS) in the UK Biobank and identified more than 90 polymorphic STRs whose repeat lengths are thought to be causally associated with blood and serum biomarkers in humans. While the majority of trait-associated STRs are non-coding, two high-impact protein-coding STRs were identified: an imperfect poly-leucine repeat in APOB associated with LDL cholesterol and a poly-serine repeat in E2F4 associated with multiple traits including red blood cell count. Fully exploring the structural consequences and functional implications of protein-coding STR variants has been challenging due to the repetitive nature of the RNA and proteins they encode. Here we explored using AlphaFold to generate structural models for variant alleles in these regions, allowing us to investigate their potential functional implications. For the APOB repeat, consistent with previous studies, our structural simulations indicate that the STR mutation affects the signal peptide region, potentially impacting protein secretion efficiency. We hypothesize that this locus primarily affects APOB protein levels by modulating its secretion rather than inducing structural changes. On the other hand, AlphaFold simulation suggests the poly-serine region in E2F4 lies in a flexible region. Modifying the poly-serine copy number did not directly affect the predicted E2F4 structure. However, when analyzing the complex formed between E2F4 and P130, a factor from the pRB family that acts as an inhibitor and stabilizes E2F4, we found that the shorter variant binds more tightly to the P130 co-factor. Knockdown/knockout of E2F4 has been reported to significantly affect cell proliferation. Our simulation suggests the longer variant reduced stability due to ineffective P130 binding, potentially leading to increased degradation susceptibility, consistent with the direction of effect observed in GWAS. In future studies, we will profile protein expression in cells containing variable poly-serine copy numbers to validate this finding. Overall, our analysis highlights how genetic variation in protein-coding STRs may affect protein structure and functional domains. While the accuracy of AlphaFold in SNP analysis is debated due to the minor impact of single point mutations on overall protein structure, our results suggest it may provide meaningful biological insights into potential function of trait-associated STRs.

Session Title: Molecular Effects of Genetic Variation Poster Session II

PB1289 Using CRISPR in the embryonic mouse brain to model patient variants in *MAP1B* associated with periventricular nodular heterotopia

Authors:

C. Kittock[1,2], L-J. Pilaz[1,2]; [1]Univ. of South Dakota, Sioux Falls, SD, [2]Sanford Res., Sioux Falls, SD

Recently, frameshift truncating and nonsense variants in *MAP1B* have been associated with periventricular nodular heterotopia (PVNH). PVNH is a neuronal migration disorder thought to arise from abnormal neuronal migration during embryonic development of the cerebral cortex. *MAP1B* encodes MAP1B, known for its role in regulating both microtubule and actin networks during embryonic brain development. However, the mechanism by which variants in *MAP1B* lead to PVNH is unknown. To study PVNH *in vivo*, we use Breasi-CRISPR, a method we developed to efficiently edit the genome of neural progenitor cells in the embryonic mouse cortex. By using Breasi-CRISPR to introduce *MAP1B*-associated PVNH patient variants and epitope tags in the endogenous *MAP1B* locus, we can directly model *MAP1B*-associated PVNH. Introduction of a *MAP1B*-associated PVNH variant via Breasi-CRISPR results in abnormal neuronal accumulation in the embryonic mouse cortex, thus partially recapitulating the patient phenotype. Live imaging in brain slices revealed that these neurons migrated slower and traveled shorter distances than wild type neurons. *MAP1B*-associated PVNH variants demonstrated robust protein expression via immunofluorescence and Western blotting. This suggests that *MAP1B* variants escape nonsense-mediated decay and that the truncated protein is stable. Due to the expression of truncated MAP1B, we hypothesized that the pathogenic mechanism underlying PVNH could be the truncated protein acting as a dominant negative. However, overexpression of the same *MAP1B*-associated PVNH variant did not recapitulate the migration delay seen when this same variant was introduced via Breasi-CRISPR. In conclusion, these data indicate that introduction of a *MAP1B*-associated PVNH variant in the endogenous *MAP1B* locus results in a neuronal migration delay in the embryonic mouse brain, but a dominant negative mechanism is unlikely to underlie this phenotype. Further experiments will test other potential mechanisms such as loss of certain *MAP1B* domains and/or haploinsufficiency. Altogether, these studies use cutting-edge tools to model PVNH *in vivo* in order to contribute to our understanding of neuronal migration disorders, the molecular mechanism underlying normal MAP1B function, and our general understanding of cortical development.

Session Title: Molecular Effects of Genetic Variation Poster Session III

PB1290 Using signatures of natural selection to inform functional significance of GWAS variants in Inflammatory Bowel Disease

Authors:

V. Gartner[1], N. Nishiyama[2], M. R. Schaner[1], K. Clough[1], S. Z. Sheikh[1], T. S. Furey[3]; [1]Univ. of North Carolina, Chapel Hill, Chapel Hill, NC, [2]UNC Chapel Hill, Chapel Hill, NC, [3]Univ North Carolina, Chapel Hill, Chapel Hill, NC

Inflammatory Bowel Disease (IBD), which includes Crohn's disease (CD) and ulcerative colitis (UC) as the two main disease subtypes, is a chronic disease of the gastrointestinal tract characterized by an inappropriate immune response to the enteric microbiota in genetically susceptible individuals. Genome Wide Association Studies (GWAS) have identified over 250 IBD risk loci, the majority of which are in non-coding regions of the genome and whose functional mechanisms contributing to disease are poorly understood. Colocalizing GWAS loci with expression and chromatin accessibility quantitative trait loci (eQTL and caQTL) can help identify putative causal variants and link risk loci with measurable phenotypes at the molecular level, but still leaves a gap in our understanding of the functional significance of these risk loci and association with disease phenotypes. Previous studies have linked adaptations in the immune system in defense of environmental pathogens with an increased risk of autoimmunity. Thus, for immune-related complex diseases such as IBD, the functional significance of disease risk loci may be related to an adaptive response to new pathogens that have acted as selective pressures in shaping population allele frequency. Towards this end, previous studies have shown that genomic signatures of natural selection can be used to identify causal SNPs within risk loci and help prioritize autoimmune disease risk variants. In this study, we sought to further test this hypothesis by investigating evolutionary pressures on candidate causal IBD variants and their associated target genes identified by colocalizing IBD GWAS variants with eQTL and caQTL results from colon tissue collected from 96 well-phenotyped IBD patients (90 CD and 6 UC) and 50 non-IBD patients. We are assessing signatures of selection of variants associated with IBD subtypes (CD and UC), as well as within clinical phenotypic sub-groups of these subtypes, to characterize subtype-specific disease pathways, assist with fine-mapping of GWAS loci, and prioritize candidate variants for functional follow-up studies. These results build on previous efforts to use human-specific signals of natural selection to understand the genetic basis of autoimmune and complex diseases to better inform effective diagnosis of IBD subtypes and potential treatments based on critical subtype-specific biological pathways.

Session Title: Molecular Effects of Genetic Variation Poster Session I

PB1291 Validating clinical interpretations of familial hypercholesterolemia variants using East Asian Biobank data

Authors:

M. Hwang[1,2], B-J. Kim[1], H-H. Won[2], Y. Kim[1]; [1]Korea NIH, Cheongju-si, Korea, Republic of, [2]Sungkyunkwan Univ, Seoul, Korea, Republic of

Key databases used for clinical interpretation include HGMD and ClinVar, which organize genetic variant information based on expert analysis, algorithmic predictions, and biological functions. Clinical interpretation of genetic variants is crucial for understanding their implications in disease. Despite the significance of these resources, inconsistencies among experts and limitations in precise functional predictions underscore the need for enhanced accuracy of variant interpretation. In recent years, large-scale biobanks have been established worldwide, including Korea Biobank, BioBank Japan, UK Biobank, Kadoorie Biobank, TOPMed, and others. These biobanks play a crucial role in researching disease-associated genetic variants and may provide an invaluable opportunity to validate the clinical interpretations through disease association analyses involving hundreds of thousands of participants. With a focus on familial hypercholesterolemia (FH), this study aims to validate the clinical interpretation of genetic variants registered in ClinVar. To achieve this goal, genetic variants and low-density lipoprotein (LDL) phenotypes from approximately 126,000 Korean participants were analyzed along with genome-wide association summary statistics for LDL from BioBank Japan, proposing a reclassification method based on the association results. The validation and reclassification process involved defining outliers in the genetic effect distribution from association results and reclassifying clinical interpretation of variants based on empirical effect size distribution considering allele frequency of the variants. Through this process, we validated one pathogenic variant and 66 benign variants classified according to ClinVar interpretation. Additionally, conflicts arising from 12 variants of uncertain significance (VUS) were resolved by reclassifying five pathogenic variants and seven benign variants. Notably, the reclassified pathogenic variants exhibited a higher prevalence of heart-related disease, including coronary artery disease and congestive heart failure, in comparison to the benign variants. Overall, this study emphasizes the pivotal role of large-scale biobanks in advancing our understanding of disease-associated genetic variants and improving the accuracy of clinical interpretations. The validation and reclassification of FH-related variants serve as a valuable step towards enhancing our knowledge and clinical practice in the field of precision medicine.

Session Title: Molecular Effects of Genetic Variation Poster Session II

PB1292 Variants in the SOX9 transactivation middle domain induce axial skeleton dysplasia and scoliosis

Authors:

L. Wang[1], Z. Liu[2], S. Zhao[3], K. Xu[1], C. Qiu[4], Z. Zhao[1], G. Li[5], J. Cai[1], Q. Li[1], Z. Wu[6], R. Gray[2], N. Wu[1]; [1]Peking Union Med. Coll. Hosp., Beijing, China, [2]The Univ. of Texas at Austin, Austin, TX, [3]Baylor Coll. of Med., Houston, TX, [4]Shandong Univ., Jinan, China, [5]Peking Union Med. Coll. Hosp. Chinese Academy of Med. Sci., beijing, China, [6]Peking Union Med. Coll. Hosp., Chinese Academy of Med. Sci., Beijing, China

SOX9 is an essential transcriptional regulator of cartilage development and homeostasis. In humans, dysregulation of *SOX9* is associated with a wide spectrum of skeletal disorders, including campomelic and acampomelic dysplasia, and scoliosis. The mechanism of how *SOX9* variants contribute to the spectrum of axial skeletal disorders is not well understood. Here, we report four novel pathogenic variants of *SOX9* identified in a large cohort of patients with congenital vertebral malformations. Three of these heterozygous variants are in the HMG and DIM domains, and for the first time, we report a pathogenic variant within the transactivation middle (TAM) domain of *SOX9*. Probands with these variants exhibit variable skeletal dysplasia, ranging from isolated vertebral malformation to acampomelic dysplasia. We also generated a *Sox9* hypomorphic mutant mouse model bearing a microdeletion within the TAM domain ($Sox9^{Asp272del}$). We demonstrated that disturbance of the TAM domain with missense mutation or microdeletion results in reduced protein stability but does not affect the transcriptional activity of SOX9. Homozygous $Sox9^{Asp272del}$ mice exhibited axial skeletal dysplasia including kinked tails, ribcage anomalies, and scoliosis, recapitulating phenotypes observed in human, while heterozygous mutants display a milder phenotype. Analysis of primary chondrocytes and the intervertebral discs in $Sox9^{Asp272del}$ mutant mice revealed dysregulation of a panel of genes with major contributions of the extracellular matrix, angiogenesis, and ossification-related processes. In summary, our work identified the first pathologic variant of *SOX9* within the TAM domain and demonstrated that this variant is associated with reduced SOX9 protein stability. Our finding suggests that reduced SOX9 stability caused by variants in the TAM domain may be responsible for the milder forms of axial skeleton dysplasia in humans.

Session Title: Molecular Effects of Genetic Variation Poster Session III

PB1293 Variant-to-function translation of obesity-associated loci through multi-omics data integration

Authors:

V. Diez-Obrero[1], R. A. J. Smit[1,2], M. Guindo-Martinez[1,2], R. J. F. Loos[1,2]; [1]Novo Nordisk Fndn. Ctr. for Basic Metabolic Res., Copenhagen, Denmark, [2]Icahn Sch. of Med. at Mount Sinai, New York, NY

Introduction: Genome-wide association studies (GWAS) have identified hundreds of loci associated with body mass index (BMI). However, the variant-to-function translation of only a handful of these loci has been successful. Here, we aim to identify candidate effector genes within BMI-associated loci, and reveal their molecular effects, to ultimately provide new biological insights into obesity pathogenesis.
Methods: We integrated GWAS summary statistics of BMI reported by the GIANT Consortium (Yengo et al, 2018) with multi-omics quantitative trait loci (QTLs), including plasma-derived gene expression, protein, and metabolite QTLs (e/p/metabQTLs), and brain and adipose tissue-derived eQTLs from different studies. We performed colocalization analyses to detect shared genetic signals between BMI and -omic traits. We used cis-e/pQTLs to identify candidate effector genes within loci, and trans-e/p/metabQTLs to pinpoint the molecular effects of cis-colocalizing loci. We implemented a two-sample Mendelian Randomization approach to assess the causal relationship between -omic traits and BMI.
Results: Cis-e/pQTLs of 916 genes colocalized at 259 of the 536 BMI-associated loci. The integration of multiple QTL datasets from different tissues maximized the power for discovery. In 18 of the 259 loci, the same genes colocalized both at the gene expression- and protein-level (e.g., TTC12 and LYZ). Trans-e/pQTLs and/or metabolite-QTLs colocalizing in 181 of the 259 cis-colocalizing loci pointed to molecular mechanisms. For example, we found that genetic variants in GIPR that are associated with higher BMI and cardiovascular disease, also associate with higher protein levels of GIP, lower protein levels of SCGB3A1, QPCTL, and MSMB, and lower levels of X-12818 metabolite in plasma. Furthermore, the protein levels in plasma of 62 genes, such as SNX1 and PRKCB, are causally associated with BMI, which may provide new biomarkers of obesity risk.
Conclusion: Integrating multi-omics data with GWAS results successfully provided new molecular insights into the variant-to-function translation of BMI-associated loci.

Session Title: Molecular Effects of Genetic Variation Poster Session I

PB1294 Vitamin D binding protein gene polymorphisms of the rs7041 and rs4588 minor allele are associated with HIV and tuberculosis co-infection.

Authors:

E. Acen; MAKERERE Univ., PHYSIOLOGY RESEARCH LABORATORY KAMPALA UGANDA, Uganda

Vitamin D binding protein gene polymorphisms of the rs7041 and rs4588 minor allele are associated with HIV and tuberculosis co-infection. Background: Studies show that the D binding protein (DBP) gene located at 4q12-q13 is highly polymorphic with over 120 variants (Bikle & Schwartz, 2019). These genetic variants affect the distribution of vitamin D in the circulatory system, leading to vitamin D deficiency and compromised immunity (Alhomsi et al., 2020). The two extensively studied non-synonymous DBP single nucleotide polymorphisms (SNPs) rs7041 and rs4588 occur in different populations (Santos et al., 2017). The wild-type Gc1F genotype is predominantly found in the African population, with a low frequency of Gc2 and Gc1S genotypes, and it is associated with low vitamin D levels in whites. According to previous research minor alleles exhibit tendencies of being risk alleles of disease. The aim of this study was to explore the DBP gene polymorphism and the risk of disease among TB patients and household contacts with and without HIV infection. Methods: This was across-sectional study with 53 active tuberculosis patients attending Kiruddu Referral Hospital, 23 latent tuberculosis individuals, and 27 individuals without tuberculosis infection from the KTB cohort. These were aged between 12 to 65 years. DNA extraction and PCR were performed and a product of 498 bp was obtained. We genotyped the DBP gene by Sanger sequencing and the single nucleotide polymorphisms were identified using the BioEdit tool. Results: The study frequency distributions of the DBP genotypes were reported as 97% Gc1F, 2% Gc2 and 1% Gc1S and Hardy-Weinberg equilibrium analysis was in equilibrium, D'= 0. Majority were female 63 (61.2 %) and of these 19 (18.4%) were HIV positive. The 2% of the participants had the Gc2 genotype and had HIV and TB co infection Conclusion: The DBP Gc1F genotype was predominantly found in the study population with the minor alleles associated with active and latent TB states. Additionally the Gc2 minor allele was associated with HIV and active TB confection in the Ugandan population.

Session Title: Molecular Effects of Genetic Variation Poster Session II

PB1295 Whole exome sequencing data reveals candidate loci for Type 1 diabetes in Kuwaiti-Arab families

Authors:

R. Nizam[1], P. Hebbar[2], S. John[3], D. Antony[1], M. Dashti[4], A. Channanath[1], A. Shaltout[1], H. Alkandari[1], H. Koistinen[5], J. Tuomilehto[6], O. Alsmadi[7], T. Thanaraj[1], F. Al-Mulla[8]; [1]Dasman Diabetes Inst., Kuwait, Kuwait, [2]Dasman Diabetes Inst., Shark, Kuwait, [3]Dasman Diabetes Inst., 15462, Dasman, Kuwait, [4]Dasman Diabetes Inst., Al Kuwayt, Kuwait, [5]Univ. of Helsinki and Helsinki Univ. Hosp., Helsinki, Finland, [6]Finnish Inst. for Hlth.and Welfare, Helsinki, Finland, Finland, Finland, [7]King Hussein Cancer Ctr., Amman, Jordan, [8]Dasman Diabetes Inst., Kuwait City, Kuwait

Prevalence of Type 1 diabetes (T1D) has been remarkably high in Kuwait and in the Middle Eastern region. International Diabetes Federation (2021) ranks Kuwait as third among countries with high prevalence of T1D. Consanguineous marriages are common in Kuwait (50-70%) leading to an excessive burden of recessive alleles. Despite this, limited studies have explored the genetic epidemiology of T1D from Kuwait. In the present study, we aim to identify genomic loci linked to T1D in Kuwaiti-Arab families which were previously unreported. Data and samples used in this study were obtained from the Childhood-Onset Diabetes eRegistry (CODeR), Kuwait. Linkage analysis was carried out on whole exome sequencing data from 18 unrelated T1D families consisting of 37 T1D cases and 49 unaffected first-degree relatives. Parametric two-point linkage analysis based on both recessive/dominant inheritance model was performed with PSEUDOMARKER 2.0 using default parameters. We identified three potential genomic regions linked to T1D in Kuwaiti-Arab families and these include chromosome 6p22.2 (LOD =3.72), 3p25.3 (LOD=3.010) and 4q31.3 (3.186) which showed evidence of significant linkage under a recessive mode of inheritance ($P \leq 2.2\times10^{-5}$). Fine mapping of these regions revealed candidate genes associated with uric acid metabolism. Supportive quantitative trait loci (QTL) data from GTEX and NephQTL indicates the likelihood for shortlisted genes to be a potential marker for nephropathy in T1D. In the first of its kind study from the region, our study highlights candidate loci linked to T1D in Kuwaiti-Arab families and their prospective role in disease etiology by regulating the expression of genes implicated in uric acid and metabolic events, as evidenced by publicly available QTL data.

Session Title: Molecular Effects of Genetic Variation Poster Session III

PB1296 Identification of suicide risk loci in eQTLs by whole-genome sequencing analysis.

Authors:

S. Han[1], E. Diblasi[1], E. Monson[1], A. Shabalin[1], D. Chen[1], A. Fraser[2], Z. Yu[2], M. Staley[3], W. Callor[3], E. Christensen[3], D. Crockett[4], Q. Li[5], V. Willour[6], E. Ferris[7], A. Bakian[1], B. Keeshin[1], A. Docherty[1], K. Eilbeck[1], H. Coon[1]; [1]Univ. of Utah Sch. of Med., Salt Lake City, UT, [2]Univ. of Utah Huntsman Cancer Inst., Salt Lake City, UT, [3]Utah Dept. of Hlth.and Human Services, Salt Lake City, UT, [4]Intermountain Hlth.care, Salt Lake City, UT, [5]Janssen R&D, LLC, Titusville, NJ, [6]Univ Iowa, Iowa City, IA, [7]Univ. of Utah Sch. of Med., Satl Lake City, UT

Recent large-scale genome-wide association studies (GWAS) have started to identify potential genetic risk loci associated with risk of suicide; however, a large portion of suicide-associated genetic factors affecting gene expression remain elusive. Dysregulated gene expression, not assessed by GWAS, may play a significant role in increasing the risk of suicide death. We performed the first comprehensive genomic association analysis prioritizing brain expression quantitative trait loci (eQTLs) within regulatory regions in suicide deaths from the Utah Suicide Genetic Risk Study (USGRS). 440,324 brain-regulatory eQTLs were obtained by integrating brain eQTLs, histone modification ChIP-seq, ATAC-seq, DNase-seq, and Hi-C results from publicly available data. Subsequent genomic analyses were conducted in whole-genome sequencing (WGS) data from 986 suicide deaths of non-Finnish European (NFE) ancestry and 415 ancestrally matched controls. Additional independent USGRS suicide deaths with genotyping array data (n=4,657) and controls from the Genome Aggregation Database were explored for WGS result replication. One significant eQTL locus, rs926308 (p=3.24e-06), was identified. The rs926308-*T* is associated with lower expression of *RFPL3S*, a gene important for neocortex development and implicated in arousal. Gene-based analyses performed using *Sherlock* Bayesian statistical integrative analysis also detected 20 genes with expression changes that may contribute to suicide risk. From analyzing publicly available transcriptomic data, nine of these genes have previous evidence of differential expression in psychiatric disorders that may be associated with suicide, including schizophrenia and autism (*ZNF501*, *CNN3*, *IGF1R*, *KLHL36*, *NBL1*, *PDCD6IP*, *SNX19*, *TBCA*, and *ARSA*). In summary, our study identified one risk locus and nine genes associated with suicide risk via gene expression, providing new insight into possible genetic and molecular mechanisms leading to suicide.

Session Title: Molecular Effects of Genetic Variation Poster Session I

PB1297 Widespread transposable element dysregulation in human brains with Alzheimer's disease

Authors:

Y. Feng[1], Y. Hou[1], Y. Zhou[1], J. B. Leverenz[2,3], C. Eng[1,2,4,5], A. A. Pieper[6,7,8,9,10], A. Goate[11,12,13], F. Cheng[1,2,5]; [1]Genomic Med. Inst., Lerner Res. Inst., Cleveland Clinic, Cleveland, OH, [2]Dept. of Molecular Med., Cleveland Clinic Lerner Coll. of Med., Case Western Reserve Univ., Cleveland, OH, [3]Lou Ruvo Ctr. for Brain Hlth., Neurological Inst., Cleveland Clinic, Cleveland, OH, [4]Dept. of Genetics and Genome Sci., Case Western Reserve Univ. Sch. of Med., Cleveland, OH, [5]Case Comprehensive Cancer Ctr., Case Western Reserve Univ. Sch. of Med., Cleveland, OH, [6]Harrington Discovery Inst., Univ. Hosp. Cleveland Med. Ctr., Cleveland, OH, [7]Dept. of Psychiatry, Case Western Reserve Univ., Cleveland, OH, [8]Geriatric Psychiatry, GRECC, Louis Stokes Cleveland VA Med. Ctr., Cleveland, OH, [9]Inst. for Transformative Molecular Med., Sch. of Med., Case Western Reserve Univ., Cleveland, OH, [10]Dept. of NeuroSci., Case Western Reserve Univ., Sch. of Med., Cleveland, OH, [11]Dept. of Genetics and Genomic Sci., Icahn Sch. of Med. at Mount Sinai, New York, NY, [12]Nash Dept. of NeuroSci., Icahn Sch. of Med. at Mount Sinai, New York, NY, [13]Ronald M. Loeb Ctr. for Alzheimer's Disease, Icahn Sch. of Med. at Mount Sinai, New York, NY

Transposable element (TE) dysregulation of quantitative trait loci (QTL) has not been well-characterized in human brains. Here, we leveraged large-scale RNA-sequencing (RNA-seq), whole-genome sequencing (WGS), and various QTL data (xQTL) from three brain biobanks to comprehensively characterize TE activation across diverse pathobiology of Alzheimer's disease (AD), including Tau and amyloid beta, APOE genotypes, and sex differences, using AI\ML and translational bioinformatics approaches. To identify genetic control of expressed TEs, we integrated expressed TEs with matched whole-genome sequencing (WGS) data and identified 38,398 genome-wide significant TE expression quantitative trait loci (teQTLs) in human brains. Compared with traditional various quantitative trait loci (xQTLs), we found that most teQTLs regulate distinct genes. We then used colocalization analysis to integrate six AD GWAS summary statistical datasets (278,950 cases versus 1,780,303 controls) with xQTLs, including teQTLs, gene expression QTLs (eQTLs), DNA methylation QTLs (meQTLs), and H3K27 histone acetylation QTLs (haQTLs), for identifying likely causal TEs involved in AD. This identified new AD risk genes, including complement C1q tumor necrosis factor-related protein 4 (*C1QTNF4*) and farnesyl-diphosphate farnesyltransferase 1 (*FDFT1*). A TE from Plat_L3 subfamily (chr11: 47605296-47605575) suppressed expression of *C1QTNF4* (its nearest gene) via long-range enhancer-promoter chromatin interaction in neurons. We identified that a TE from AluJb subfamily (chr8: 11840915-11841089) increased expression of *FDFT1* (its nearest gene) via microglia-specific long-range chromatin looping. Expression perturbation of *FDFT1* and *C1QTNF4* were significantly correlated with Braak staging score of AD patients. We further identified that individuals with AD risk allele APOE4 displayed elevated expression of AluJb and *FDFT1* in patient induced pluripotent stem cells (iPSC)-derived microglia and superior frontal gyrus (SFG) microglia compared to APOE3 individuals. These findings demonstrate widespread TE dysregulation in human aging brains and teQTLs offer a useful QTL analytic approach to identify AD risk genes and other complex diseases if broadly applied.

Session Title: Molecular Effects of Genetic Variation Poster Session II

PB1298 Wnt activity reveals context-specific genetic effects on gene regulation in neural progenitors.

Authors:

J. Valone, N. Matoba, B. D. Le, J. Wolter, J. Mory, D. Liang, N. Aygun, K. A. Broadaway, M. L. Bond, K. L. Mohlke, M. J. Zylka, M. I. Love, J. L. Stein; Univ. of North Carolina at Chapel Hill, Chapel Hill, NC

Common genetic variation has previously been associated with brain-relevant traits and risk for neuropsychiatric disorders; however, these loci are mostly found in non-coding regions without clear mechanisms of action. Regulatory mechanisms of non-coding loci are inferred by mapping the effects of genetic variation on regulatory element activity, marked by accessible chromatin peaks (chromatin accessibility quantitative trait loci or caQTL), or gene expression (eQTL). Studies in bulk post-mortem tissue have explained mechanisms for a subset of brain-trait associated loci through colocalization of causal variants, yet many brain-trait associated variants remain unexplained. This “missing regulation” may be due to variants impacting regulatory mechanisms only in specific contexts or under certain stimuli (response-QTLs). We hypothesized that stimulation of a developmental signaling pathway in a homogeneous neural cell type would reveal previously undetected functions of genetic variation.
We evaluated context-specific effects of genetic variation in a population of primary human neural progenitor cells (hNPCs; nmax= 82), a developmental cell type with regulatory elements enriched for genetic association signals in multiple-brain related traits and neuropsychiatric disorders. We measured chromatin accessibility via ATAC-seq and gene expression via RNA-seq in hNPCs following stimulation of the canonical Wnt pathway, which is known to impact neural progenitor proliferation, cortical patterning, and has previous associations with neuropsychiatric disorders.
Wnt-responsive regulatory elements (WREs) were enriched for variants associated with brain structure and neuropsychiatric disorders. Over 43,000 caQTLs regulating over 36,000 unique caPeaks were identified alongside over 3,000 unique eQTL-eGene pairs. Stimulation of the Wnt pathway increased the detection of genetically influenced REs/genes by 66.2%/52.7%, and led to the identification of 397 REs primed for effects on gene expression. Genetically influenced REs were enriched in regions under positive selection along the human lineage, including a previously validated enhancer of the Wnt receptor gene FZD8. We also found over 1,800 response-caQTLs and 102 response-eQTLs, where a significant genotype-by-condition interaction was detected. Context-specific molecular quantitative trait loci increased brain-trait colocalizations by up to 70%.
Our results characterize context-specific genetic effects in hNPCs that provide novel insights into neurodevelopmental gene regulatory mechanisms underlying brain trait-associated loci.

Session Title: Molecular Effects of Genetic Variation Poster Session III

PB1299 α-Synuclein pathology in *Drosophila melanogaster* is exacerbated by haploinsufficiency of *Rop*: Connecting STXBP1 encephalopathy with α-synucleinopathies.

Authors:

T. Matsuoka[1], H. Yoshida[2], T. Kasai[3], T. Tozawa[1], T. Chiyonobu[1,4]; [1]Dept. of Pediatrics, Graduate Sch. of Med. Sci., Kyoto Prefectural Univ. of Med., Kyoto, Japan, [2]Dept. of Applied Biology, Kyoto Inst. of Technology, Kyoto, Japan, [3]Dept. of Neurology, Graduate Sch. of Med. Sci., Kyoto Prefectural Univ. of Med., Kyoto, Japan, [4]Dept. of Molecular Diagnostics and Therapeutics, Graduate Sch. of Med. Sci., Kyoto Prefectural Univ. of Med., Kyoyo, Japan

STXBP1 haploinsufficiency causes STXBP1 encephalopathy (STXBP1-E), which encompasses neurological disturbances including epilepsy, neurodevelopmental disorders, and movement disorders. STXBP1 (also called Munc18-1) is involved in SNARE complex formation and contributes to synaptic vesicle exocytosis. A recent in vitro study proposed the role of STXBP1 as a molecular chaperone for α-synuclein, a key molecule in the pathogenesis of neurodegenerative disorders including Parkinson's disease. Patients with STXBP1-E have been reported to exhibit extrapyramidal symptoms, but no studies to date have shown α-synuclein pathology in model organisms or patients with STXBP1-E. Here, we report that haploinsufficiency of *Rop*, the *STXBP1* homolog in *Drosophila*, exacerbates α-synuclein pathology. Homozygosity for null variants of *Rop* is lethal at the embryonic stage, whereas flies that are heterozygous for null variants survive to adulthood, although their protein expression is about 20% lower than that of the wild type. In this *Drosophila* model of STXBP1-E, the volume of transgenic human α-synuclein in the 1% Triton-soluble fraction was significantly reduced compared to the control, while that in the 1% Triton-insoluble fraction was significantly increased, indicating that *Rop* haploinsufficiency aggravates α-synuclein aggregation. The α-synuclein-transgenic STXBP1-E model also showed a significant decrease in the number of dopamine neurons compared to control flies, and adults displayed reduced motor function. This model also displayed worsening compound eye degeneration. These results indicate that α-synuclein neurotoxicity is exacerbated in STXBP1-E model flies. We also showed that feeding trehalose to STXBP1-E model flies significantly reduced the 1% Triton-insoluble fraction of α-synuclein and restored adult locomotor function. This study is the first to demonstrate a molecular interaction between STXBP1 and α-synuclein using a model organism. This study contributes to understanding the mechanisms of neurodegeneration in patients with STXBP1-E and provides new insights into the pathogenesis of α-synucleinopathies.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1301 2-Sample Mendelian Randomization Reveal Widespread Lipidomic Dysregulation Underpinning Severe Obesity in Self-identified Mexican Americans

Authors:

M. Yaser[1], M. Graff[2], H-H. Chen[3], P. Sharma[4], J. E. Curran[5], P. J. meikle[6], W. Zhu[7], H. Highland[8], K. Young[2], L. Petty[9], A. Gutierrez[10], M. Lee[11], J. McCormick[11], S. P. Fisher-Hoch[12], J. E. Below[13], K. E. North[2]; [1]Univ. of North Carolina Chapel Hill, Chapel Hill, NC, [2]Univ North Carolina, Chapel Hill, NC, [3]Vanderbilt Unviersity, Nashville, TN, [4]Univ. of North Carolina, Chapel Hill, NC, [5]Univ. of Texas Rio Grande Valley, Brownsville, TX, [6]Univ. of North Carolina Chapel Hill, Baker Heart and Diabetes Institute, Australia, [7]Vanderbilt Univ., Nashville, TN, [8]Univ North Carolina at Chapel Hill, Chapel Hill, NC, [9]Vanderbilt Univ. Med. Ctr., Nashville, TN, [10]UT Hlth.Houston, Houston, TX, [11]UTH SPH Brownsville, Brownsville, TX, [12]UT SPH Brownsville, Brownsville, TX, [13]Univ. of Vanderbilt Med. Ctr., Nashville, TN

Background Severe obesity (SevO), a Body Mass Index (BMI) of ≥40 kg/m^2, is the fastest-growing metabolic disorder in the U.S. While lipids are significantly associated with SevO, the nature of the association between the heterogenous lipid species has not been queried, particularly in high-risk populations. We conducted a two-sample Mendelian Randomization (2SMR) to characterize the lipid species that underpin the risk of SevO in self-identified Mexican Americans, a population that has a SevO prevalence twice that of non-Hispanic Whites. **Methods** We conducted a phenotypic association analysis between the lipids and SevO in individuals from the Cameron County Hispanic Cohort (CCHC). We performed genome-wide associations (GWAS) for lipid species nominally ($p<0.05$) associated with SevO in 2,096 individuals from the CCHC. We also utilized a multi-population SevO-GWAS, comprised of 32,358 cases and 45,806 controls for our outcome. We extracted the associated genetic variants from the respective species' GWAS results, considering only variants surpassing a significance threshold of 5×10^{-6}. We then matched these variants' summary statistics to their corresponding pairs in the SevO GWAS. Subsequently, for each species, we ran a 2SMR using MR Egger to account for pleiotropy. To further assess clinical significance, we queried the instrumental variables from significant lipid species in the GWAS catalog. **Results** Analyses revealed that 34 classes (597 species) showed an association with SevO, out of 49 lipid classes (830 species). 2SMR identified 41 lipid species with potentially causal associations with SevO. Among these, 29 demonstrated a protective effect against SevO, including 3 ceramides, 3 trihexosylceramides (Hex3Cer), 6 lysophosphatidylcholines (LPC), 7 phosphocholines (PC), 2 sphingomyelins (SM), 1 lysophosphatidylethanolamine (LPE), 1 dimethyl-cholesteryl ester (dimethyl-CE), and 1 phosphatidylinositol (PI). On the other hand, 18 lipid species were identified as risk-increasing for SevO, including 3 diglycerols (DG), 12 triglycerols (TG), 1 sphingomyelin (SM), 1 phosphatidylinositol (PI), and 1 phosphocholine (PC). Genetic instruments associated with protective species were associated with polyunsaturated fatty acids, immunity cells, and coagulation. In contrast, the genetic instruments associated with the risk-increasing species were frequently linked to cholesterols, liver diseases, and cardiovascular diseases (CVD). **Conclusion** Our results support an important role of dysregulation of species of lipids in the pathogenesis of SevO. Such findings are important as they may inform novel targets and interventions for SevO.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1302 3D genomic features across over 50 diverse cell types reveal insights into the common and differing genomic architectures of sixteen autoimmune disorders

Authors:

K. Trang, M. C. Pahl, J. A. Pippin, P. K. Sharma, Z. Thompson, S. F. Grant, A. D. Wells; Ctr. for Spatial and Functional Genomics, The Children's Hosp. of Philadelphia, Philadelphia, PA

The risks for many autoimmune disorders are polygenic with a degree of overlap in susceptibility loci, suggesting a shared molecular etiology across diseases. Genome-wide association studies (GWAS) has shown genetic intercorrelation between such diseases, including ankylosing spondylitis (AS), Grave's disease, Hashimoto's thyroiditis, celiac disease, Crohn's disease, inflammatory bowel disease, ulcerative colitis, atopic eczema, juvenile idiopathic arthritis (JIA), multiple sclerosis, psoriasis, rheumatoid arthritis, systemic lupus erythematosus, type I diabetes mellitus, and vitiligo. However, remains uncovered the degree to which these diseases share underlying causal variants, corresponding effector genes, and cellular context. Leveraging our existing 3D genomic datasets consisting of high-resolution promoter-focused Capture-C/Hi-C, ATAC-seq, and RNA-seq across over 50 diverse human cell types enabled us to contrast the causal variants and the corresponding effector genes for these diseases. These include various innate and adaptive immune cells whose functions play crucial roles in the development and progression of autoimmune diseases, plus non-immune cell types spanning metabolic, neuronal, and pluripotential stem cell lineages. Using stratified LD regression, we calculated the proportion of genome-wide SNP heritability attributable to our derived cell type-specific features via the integration of the most recent GWAS summary statistics for these diseases. These analyses revealed a statistically significant enrichment (P<0.05) for 13 diseases across all immune cell types. Interestingly, Grave's disease also showed enrichment in several metabolic, neural, and stem cell types. AS, JIA and psoriasis were exceptions, with such enrichments limited to distinct subsets of T cell and plasmacytoid dendritic cells. Subsequent application of our chromatin contact-based ‘variant-to-gene’ mapping of these loci in these cell types identified “hub” genes including members of the NF−kappa B binding and integrin binding pathways that appeared in almost every cell type for all disorders, supporting the notion that mediation of cell-cell and cell-extracellular matrix adhesion plays crucial roles in autoimmune etiology. Other immune functional pathways and biological processes were uniquely associated with a given cell type or with a specific disorder. In conclusion, our comprehensive appraisal of 3D genomic datasets in a large panel of different cell types drives greater genomic understanding of both the similarities and the differences across autoimmune disease pathogenesis.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1303 3D modelling predicts reduced stability and protein binding for a biallelic missense variant in the Sonic Hedgehog gene in autism spectrum disorder

Authors:

T. Muhammad[1], A. Rabia[2], R. Harripaul[1], A. Mikhailov[3], M. Carter[4], S. Mahmood[2], S. Maqbool[5], J. Vincent[3], M. Ayub[6]; [1]Univ. of Toronto, Toronto, ON, Canada, [2]Univ. of Hlth.Sci., Lahore, Pakistan, Lahore, Pakistan, [3]Ctr. for Addiction and Mental Hlth., Toronto, ON, Canada, [4]Dept. of Genetics, Children's Hosp. of Eastern Ontario,, Ottawa, ON, Canada, [5]Children's Hosp. & The Inst. of Child Hlth., Lahore, Pakistan, Lahore, Pakistan, [6]Dept. of Psychiatry, Queens Univ. Kingston,, Kengston, ON, Canada

Background: Dissecting the genetics of autism spectrum disorders (ASD) has indicated the importance of *de novo* variants in single genes as well as copy number variants involving one or more genes. Recently, studies have shown the importance of biallelic pathogenic variants in genes with autosomal recessive (AR) inheritance, particularly in populations in which consanguineous marriages are common. However, it is also noteworthy that biallelic variants may have quite different phenotypic and clinical consequences to heterozygous variants in the same gene.**Objectives:** The goal was to explore the phenotype/genotype relationship for a rare biallelic missense mutation in *Sonic Hedgehog*, though clinical assessments of the patient, and molecular genetic and *in silico* analyses to examine the underline genetic and proteomic alterations.**Methods:** Clinical assessments were performed for developmental, medical and family history, current functioning, problematic behaviors, psychiatric symptoms and ASD using ADI-R as framework. Biometrics and dysmorphology were assessed. Genetic analysis were carried out by microarray genotyping and whole exome sequencing. Functional predictions for deleteriousness using various well known prediction tools. For *in silico* analysis, molecular docking MD simulations were performed using MOE and Desmond.**Results:** Here we report on a proband with ASD, who inherited a biallelic missense variant in the gene encoding the developmental morphogen, Sonic Hedgehog (*SHH*; NM_000193.4: c.206A>T: p. Asn69Ile). Heterozygous missense or loss-of-function variants in *SHH* have previously been reported in association with the autosomal dominant (AD) forebrain and midface developmental disorder, holoprosencephaly-3 (HPE3; MIM 142945). Here we describe the clinical features identified with this biallelic missense variant of a conserved amino acid residue, for comparison and contrast with the features of HPE3 and related phenotypes. We also demonstrate a predicted reduction in stability of the mutant SHH protein, also impacting its binding with hedgehog interacting protein (HHIP).
Conclusion: Here, we report on a family with a biallelic missense mutation in SHH, with ASD but no sign of HPE, and present supporting evidence for the pathogenicity of the mutation from 3D-modelling analysis that predicts reduced functionality and/or stability of SHH.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1304 A comparison of the obesity-associated transcriptome in skeletal muscle of African Americans and Europeans

Authors:

S. Das[1], S. S. Das[2]; [1]Wake Forest Univ. Sch. of Med., Winston-salem, NC, [2]The Sch. of Biotechnology at Atkins, Atkins Academic and Technology High, Winston-salem, NC

Obesity is a common disease with a higher prevalence among African Americans. Obesity alters cellular function in many tissues, including skeletal muscle, and is a risk factor for many life-threatening diseases, including cardiovascular disease and diabetes. Utilizing publicly available summary statistics data from transcriptome-wide analyses on skeletal muscle tissues from well-powered human cohorts, in this study, we compared genes and biological pathways affected by obesity in European and African ancestry populations. Results from the linear regression analysis on the FUSION (European, N= 301) and AAGMEx (African American, N=256) cohorts identified 2569 body mass index (BMI)-associated transcripts ($q<0.05$), of which 970 genes (at $p<0.05$) are associated in both cohorts, and the majority showed the same direction of effect on BMI. Biological pathway analyses, including over-representation and gene-set enrichment analyses, identified enrichment of protein synthesis pathways (e.g., ribosomal function) and the ceramide signaling pathway in both cohorts among BMI-associated down- and up-regulated transcripts, respectively. A comparison using the IPA-tool suggested activation of inflammation pathways only in Europeans with obesity. Interestingly, these analyses suggested repression of the mitochondrial oxidative phosphorylation pathway in Europeans but showed its activation in African Americans. Integration of SNP-to-Gene analyses-predicted target gene-lists for obesity-associated genetic variants (GWAS-identified SNPs) and BMI-associated transcripts suggested that these SNPs might cause obesity by altering the expression of critical target genes (e.g., *GRB14*) in the muscle. This study provides replication of obesity-associated transcripts and biological pathways in skeletal muscle across ethnicities, but also identifies obesity-associated processes unique in either African or European ancestry populations.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1305 A contextual genome-wide perspective on physical activity and its relationship to health and illness

Authors:

M. Galimberti[1,2], D. F. Levey[1,2], J. D. Deak[1,2], K. Adhikari[1,2], C. Overstreet[1,2], H. Zhou[1,2], N. J. Lake[1], K. Harrington[3,4], M. Lek[1], J. Gaziano[4,5,6], M. B. Stein[7,8], J. Gelernter[1,2]; [1]Yale Univ. Sch. of Med., New Haven, CT, [2]VA Connecticut Hlth.care System, West Haven, CT, [3]Boston Univ. Sch. of Med., Boston, MA, [4]VA Boston Hlth.care System, Boston, MA, [5]Brigham and Women's Hosp., Boston, MA, [6]Harvard Med. Sch., Boston, MA, [7]Univ. of California San Diego, La Jolla, CA, [8]VA San Diego Hlth.care System, San Diego, CA

Physical activity (PA) is beneficial for health and can be useful for prevention against numerous diseases. Here, we considered PA from a genome-wide perspective, identifying risk variants and genes, and studying its relationship with other traits of interest.

Using data from the Million Veteran Program (MVP), we created a quantitative physical activity rank (PAR) score, combining different PA intensities -- vigorous, moderate, and light - by summing their frequencies. For PAR phenotype during leisure time, we conducted GWAS analyses for European (EUR) (n=189,812; heritability (h^2)= 0.083±0.005), African (AFR) (n=27,044; h^2= 0.034±0.017), and Latin-American (AMR) (n=10,263; h^2= 0.083±0.036) ancestries. We also completed GWAS in MVP for different contexts of vigorous activity - leisure (V-Leisure), work (V-Work), home (V-Home) - in EUR. We ran a meta-analysis including MVP PAR EUR data and UK Biobank (UKB) summary statistics (only EUR available) on strenuous sport and other exercises (124,842 cases, 225,650 controls; h^2= 0.061±0.003). These two traits had high genetic correlation (rg=0.76±0.03) and the meta-analysis had h^2= 0.060±0.003. A cross-ancestry meta-analysis included UKB and MVP data for EUR, AFR, and AMR ancestry. The EUR and cross-ancestry meta-analyses identified 67 and 70 independent SNPs, respectively. In both cases the strongest association was on chromosome 3 at *CADM2**rs62253088 (EUR *p*-value=2.0×10^{-21}; cross-ancestry *p*-value=6.1×10^{-20}). Genetic correlations and Mendelian Randomization (MR) analyses between the PAR traits and other health outcomes confirmed the protective effects of PA on cardiovascular and respiratory system diseases, metabolic traits, aging, osteoarthritis, and gastroesophageal reflux disease. A multivariable MR analysis evaluated the possibility of body mass index as a confounder confirmed a protective role of PAR versus hospitalization caused by COVID19 (β=-0.067±0.016; *p*-value=2.8×10^{-5}). There were strong genetic differences between the vigorous activity traits. We calculated genetic correlations between V-Leisure, V-Work, and V-Home with previously analyzed traits of interest. We observed 15 traits with significant statistical differences between V-Leisure and V-Work, 12 between V-Leisure and V-Home, and 3 between V-Home and V-Work. Income, leisure screen time, and educational attainment showed significant statistical differences for all three pairs of traits. In conclusion, we provide new insights into the biology of PA, showing genetic differences and advantages according to the context, and the health benefits of PA genetic liability in protecting against several diseases.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1306 A genetic analysis of familial aggregation in inflammatory bowel disease multiplex families.

Authors:

I. Cleynen[1], D. Jans[1], H-S. Lee[2], M. Ferrante[3], S. Vermeire[3]; [1]KU Leuven, Leuven, Belgium, [2]Dept. of Biochemistry and Molecular Biology, Seoul, Korea, Republic of, [3]UZ Leuven, Leuven, Belgium

Inflammatory bowel disease (IBD), encompassing Crohn's disease (CD) and ulcerative colitis (UC), is characterized by chronic inflammation of the gut. In some families, a remarkably high prevalence of disease is seen. Here, we aim to determine the role of the known IBD risk loci in familial aggregation of disease, as well as to study if there are common variants specific to familial IBD. We used a single-centre dataset of imputed immunochip genotypes of 53 IBD families that have at least three affected first-degree relatives (157 CD, 31 UC, 129 unaffected first-degree relatives, 9 healthy spouses) and a sporadic dataset for comparison (1684 CD, 908 UC, 869 controls). For each individual, a polygenic risk score (PRS) was calculated with PRSice based on the IBD summary statistics of de Lange et al (2017), a p-value threshold ≤ 0.01 and MAF ≥ 0.01. As a group, unaffected first-degree relatives have a lower PRS than affected relatives (p=6.27e-4, β(se)=-0.56(0.14)), but still higher than unrelated controls (p=6.74e-4, β(se)=0.44(0.11)). Twenty-four families have a familial PRS - ie the mean PRS of all relatives in a family - that is higher than the mean PRS of sporadic cases; while seven families have a familial PRS that is lower than the mean PRS of unrelated controls (low PRS families). Although in most families the affected members have a higher mean PRS than the unaffected members, this relationship is sometimes reversed, especially in the low PRS families. Family-based association analysis with SAIGE found nine variants to be associated (p<1e-4) with familial IBD. Two independently associated variants, rs2241130 (p=1.19 e-5, β(se)=1.21(0.28)) and rs144641193 (p=8.23e-5, β(se)=1.29(0.33)), reside in a known IBD locus (*IL1RL1-IL18RAP*). Some other variants were also located in genes previously implicated in IBD, either through GWAS [rs72781786 in *PRKCQ* (p=2.93e-5, β(se)=-0.95(0.23))], or functional studies [rs2242601 in *EPHA1* (p=1.43e-5, β(se)=0.70(0.16)) and rs2272766 in *CTSB* (p=5.98e-5, β(se)=0.65(0.16))]. One of the four new loci is rs10772102 (p=6.41e-5, β(se)=-0.64(0.16)), located between *CLECL1*, that mediates immune cell-cell interactions, and *CD69*, that plays a role in T-cell differentiation. Our analysis indicates that common IBD risk variants play an important role in familial aggregation of disease in many multiplex families. Some families however have a very low polygenic risk, indicating shared environmental factors might be relatively more important, or genetic factors not captured by the score, e.g. rare variants, may have contributed. Also, a few loci are specifically implicated in familial IBD, some related to the immune system.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1307 A genome-wide interaction study identified genetic variants modifying the associations of fish oil supplementation with circulating polyunsaturated fatty acids.

Authors:

S. Ihejirika, A. Chiang, A. Singh, K. Ye; Univ. of Georgia, Athens, GA

Fish oil supplements (FOS) are a readily available source of omega-3 polyunsaturated fatty acids (PUFAs), an essential component of human diet. Omega-3 and omega-6 PUFAs have been linked to reduced risks and improved prognosis of a variety of chronic diseases. Previous observational association studies and randomized controlled trials indicate the presence of varying individual responses when administered dietary PUFAs or supplements. Genetic factors may be partially responsible for these varying responses. Here, we performed a genome-wide interaction study to identify genetic variants that modify the associations of fish oil supplementation with circulating PUFAs. We examined 14 phenotypes, including the absolute circulating levels of total PUFAs, omega-3 PUFAs, omega-6 PUFAs, linoleic acid (LA), docosahexaenoic acid (DHA), their relative percentages in total fatty acids, and their ratios. In a cohort of unrelated European-ancestry individuals from the UK Biobank, 113003 individuals took fish oil supplements and 243632 did not, based on the initial recruitment assessment. We identified three interaction loci reaching the genome-wide significance of $P < 5 \times 10\text{-}8$. One locus was identified for the omega-6:omega-3 ratio surrounding the *FADS1-FADS2-FADS3* gene cluster. The minor allele of the top variant (rs174535; T>C, minor allele frequency, MAF = 0.35) is associated with a stronger increase of the omega-6:omega-3 ratio in individuals that take fish oil supplements than those that do not. *FADS1* and *FADS2* are known to be involved in the biosynthesis of long-chain PUFAs. The second locus was for LA% and mapped to the *SLC35F4* gene. The minor allele of the lead variant (rs117393536; C>T; MAF = 0.038) is associated with higher LA% in participants with fish oil supplementation, but with lower LA% in those without. The third significant locus was for total PUFA levels and around the *PLVAP* gene. The minor allele of the top SNP (rs7258084, (G>A; MAF = 0.16) is associated with higher total PUFAs in those taking fish oil supplements, but with lower total PUFAs in those not taking the supplements. Additionally, we will apply the GCTA-GREML to estimate the proportion of phenotypic variance explained by genome-wide interactions with fish oil supplementation. These findings highlight the importance of gene-environment interaction studies to identify more variants explaining the missing heritability and to develop genome-informed personalized nutrition.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1308 A genome-wide meta-analysis connects iron homeostasis to metabolic disease through poly-unsaturated fatty acid synthesis.

Authors:

S. Praggastis[1], A. Gilly[2], B. Crowell[1], L. Ivanova[1], Regeneron Genetics Center, GHS-RGC DiscovEHR Collaboration, Mayo Clinic Project Generation, Penn Medicine BioBank, A. N. Economides[1], S. Hatsell[1], H. E. Lob[1], J. Bovijn[3], L. A. Lotta[4], T. De[5]; [1]Regeneron Pharmaceuticals, Tarrytown, NY, [2]Regeneron Genetics Ctr., Regeneron Pharmaceuticals, Tarrytown, NY, [3]Regeneron Genetics Ctr., Oxford, United Kingdom, [4]Regeneron Genetics Ctr., Tarrytown, NY, [5]Regeneron Genetics Ctr., Arlington Heights, IL

Individuals with genetic iron overload diseases have a higher incidence of type 2 diabetes. In individuals without clinically apparent iron overload serum levels of the iron biomarkers ferritin and transferrin positively correlate with type 2 diabetes. The molecular mechanisms connecting iron and metabolic dysfunction remain unknown. The goal of this study was to identify genetic variations that connect iron homeostasis with metabolism. We performed a large-scale genome-wide association study (GWAS) meta-analysis of iron, glucose-related biomarkers (N = 480,389), and type 2 diabetes (58,379 cases, 530,072 controls) in UKB, GHS, MALMO, UPENN, and SINAI cohorts.. We identified a variant, rs174560, in the *FADS1/FADS2* locus associated with both reduced iron levels and reduced glucose levels (Total Iron Binding Capacity, beta (95%CI) = 0.044 SD (0.035, 0.052), $p = 2.0 \times 10^{-23}$, Glucose, beta (95%CI) = -0.023 SD (-0.027, -0.018), $p = 7.5 \times 10^{-27}$). This variant is a multi-tissue expression quantitative trait locus (eQTL) resulting in reduced expression of *FADS1*, in high linkage disequilibrium with another *FADS1* variant associated with reduced iron in a previous iron GWAS. *FADS1* plays a critical role in long chain polyunsaturated fatty acid (PUFA) synthesis, hepatic depletion of this gene improves glucose clearance and limits adipose tissue expansion *in vivo*. *In vitro FADS1* expression is induced by iron supplementation. Variants in the *FADS1* locus have been strongly associated with blood glucose levels in previous GWAS, the mechanism underlying this association is unclear. Using CRISPR/Cas9 and siRNA technologies, we investigated this connection by targeting *FADS1* expression in the liver of C57Bl/6 mice. Specific reduction of *FADS1* impaired PUFA synthesis, reduced serum iron, and elevated *Hamp* expression confirming *FADS1* involvement in iron homeostasis. Additionally, reducing *FADS1* expression in $HFE^{-/-}$ mice, a genetic model of iron overload, significantly reduced serum aspartate aminotransferase levels (AST). High circulating AST levels are commonly used to diagnose liver damage or disease. This improvement in AST levels indicates reducing *FADS1* expression may be protective against liver damage. Future studies will investigate the molecular connection between *FADS1* and *Hamp* gene expression, as well as glucose metabolism.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1309 A genome-wide search for pleiotropy in more than 100,000 harmonized longitudinal cognitive domain scores

Authors:

M. Kang[1], T. Ang[1], S. A. Devine[1], R. Sherva[1], S. Mukherjee[2], E. H. Trittschuh[2], L. E. Gibbons[2], P. Scollard[2], M. Lee[2], S-E. Choi[2], B. Klinedinst[2], C. Nakano[2], L. C. Dumitrescu[3], A. Durant[3], T. J. Hohman[3], M. L. Cuccaro[4], A. J. Saykin[5], W. A. Kukull[6], D. A. Bennett[7], L-S. Wang[8], R. P. Mayeux[9], J. L. Haines[10], M. A. Pericak-Vance[4], G. D. Schellenberg[8], P. K. Crane[2], R. Au[1], K. L. Lunetta[11], J. B. Mez[1], L. A. Farrer[1]; [1]Boston Univ. Sch. of Med., Boston, MA, [2]Univ. of Washington Sch. of Med., Seattle, WA, [3]Vanderbilt Univ. Med. Ctr., Nashville, TN, [4]Univ. of Miami Sch. of Med., Miami, FL, [5]Indiana Univ. Sch. of Med., Indianapolis, IN, [6]Univ. of Washington Sch. of Publ. Hlth., Seattle, WA, [7]Rush Univ. Med. Ctr., Chicago, IL, [8]Univ. of Pennsylvania Sch. of Med., Philadelphia, PA, [9]Columbia Univ. Sch. of Med., New York, NY, [10]Case Western Reserve Univ. Sch. of Med., Cleveland, OH, [11]Boston Univ. Sch. of Publ. Hlth., Boston, MA

Background: More than 75 common variant loci account for only a portion of the heritability for Alzheimer's disease (AD). A more complete understanding of the genetic basis of AD can be deduced by exploring associations with AD-related endophenotypes. **Methods:** We conducted genome-wide scans for cognitive domain performance using harmonized and co-calibrated scores derived by confirmatory factor analyses for executive function, language, and memory. We analyzed 103,796 longitudinal observations from 23,066 members of community-based (FHS, ACT, and ROSMAP) and clinic-based (ADRC and ADNI) cohorts using generalized linear mixed models including terms for SNP, age, SNP×age interaction, sex, education, and five ancestry principal components. Significance was determined based on a joint test of the SNP's main effect and interaction with age. Results across datasets were combined using inverse-variance meta-analysis. Genome-wide tests of pleiotropy for each domain pair as the outcome were performed using PLACO software. **Results:** Individual domain and pleiotropy analyses revealed genome-wide significant (GWS) associations with five established loci for AD and AD-related disorders (*BIN1*, *CR1*, *GRN*, *MS4A6A*, and *APOE*) and eight novel loci. *ULK2* was associated with executive function in the community-based cohorts (rs157405, $P=2.19\times10^{-9}$). GWS associations for language were identified with *CDK14* in the clinic-based cohorts (rs705353, $P=1.73\times10^{-8}$) and *LINC02712* in the total sample (rs145012974, $P=3.66\times10^{-8}$). *GRN* (rs5848, $P=4.21\times10^{-8}$) and *PURG* (rs117523305, $P=1.73\times10^{-8}$) were associated with memory in the total and community-based cohorts, respectively. GWS pleiotropy was observed for language and memory with *LOC107984373* (rs73005629, $P=3.12\times10^{-8}$) in the clinic-based cohorts, and with *NCALD* (rs56162098, $P=1.23\times10^{-9}$) and *PTPRD* (rs145989094, $P=8.34\times10^{-9}$) in the community-based cohorts. GWS pleiotropy was also found for executive function and memory with *OSGIN1* (rs12447050, $P=4.09\times10^{-8}$) and *PTPRD* (rs145989094, $P=3.85\times10^{-8}$) in the community-based cohorts. Functional studies have previously linked AD to *ULK2*, *NCALD*, and *PTPRD*. **Conclusion:** Our results provide some insight into biological pathways underlying processes leading to domain-specific cognitive impairment and AD, as well as a conduit toward a syndrome-specific precision medicine approach to AD. Increasing the number of participants with harmonized cognitive domain scores will enhance the discovery of additional genetic factors of cognitive decline leading to AD and related dementias.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1310 A haploblock involving dozens of variable and/or differentially expressed proteins contributes to the large effect of HLA-B57 in HIV-1 control

Authors:

S. Le Clerc[1], M. Rahmouni[1], L. De Marco[1], J-l. Spadoni[1], M. Tison[1], R. Medina-Santos[1], T. Labib[1], J. Noirel[1], J. Fellay[2], P. McLaren[3], J-F. Zagury[1]; [1]CNAM, Paris, France, [2]EPFL & CHUV, Lausanne, Switzerland, [3]Publ. Hlth.Agency of Canada, Winnipeg, MB, Canada

We have conducted a Genome-Wide Association Study (GWAS) to compare a cohort of 543 HIV Elite Controllers (ECs, maintaining a viral load below 1000 copies/ml), with 3272 uninfected control subjects of European descent. By leveraging the latest database, we performed imputation of 35552 Single Nucleotide Polymorphisms (SNPs) within the Major Histocompatibility Complex (MHC) region and identified 2626 significant signals ($p<5.10\text{-}8$). Importantly, we confirmed the presence of well-established MHC signals, including rs2395029, a marker of *HLA-B*57:01*, and rs4418214, associated with the *MICA* gene.In our search for genetic markers associated with super elite control (defined as viral load below 100 copies/ml) in a subset of 217 subjects, we found a large ancestral haploblock within the MHC region, present in all individuals carrying the HLA-B*57:01 allele. This haploblock spans a length of 1.9 MB (1894493 pb) and consists of 379 SNPs exhibiting high linkage disequilibrium (LD). It encompasses a significant portion of the MHC region and is associated with damaging variations in proteins such as NOTCH4 or DXO. Moreover, according to GTEx, this haploblock is associated with differential mRNA expression patterns for several genes such as *MICB*, *HLA-B*, *ZBTB12*, *ZFP57*, *HLA-S*, which could potentially contribute to the control of HIV-1 replication. We extended our investigation to two cohorts of seropositive African-American subjects and examined the SNP rs1131446, in high LD with rs2395029 in Europeans, which tags *HLA-B*57:03* in African populations and is associated with the control of viral load. We found that this SNP together with 43 SNPs in LD form a small haploblock spanning 128 kb in African Americans. The GTEx mRNA expression profile associated with this haploblock in PBMCs from African subjects was highly similar to that observed in Europeans. These findings suggest a potential common mechanism underlying the massive effect of the *HLA-B*57:01/B*57:03* alleles for the control of HIV-1 replication. Overall, our study not only revealed a novel haploblock and molecular mechanisms involved in HIV-1 control but also highlighted the interplay of various factors, including direct anti-HIV-1 activity, and innate/adaptive responses, that could contribute to the elite control phenotype.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1311 A haplotype based extreme phenotype GWAS to functional investigation revealed the function of *CNTNAP5* in glaucomatous neurodegeneration

Authors:

S. Chakraborty, S. Saha Roy, J. Sarma, S. Das, S. Bagchi, S. Saha, S. Bhattacharjee, M. Acharya; Natl. Inst. of BioMed. Genomics, Kalyani, India

Primary angle closure glaucoma is one of the main leading causes of blindness (PACG). As per clinical reports, in India, ~30% of people show a narrow iridocorneal angle (<15°), but out of these only 0.5-1% people develop PACG. In order to rule out anatomical heterogeneity and age as a confounder, we conducted a haplotype based GWAS, where we compared early-onset PACG patients with progressive angle closure (PACG: age <50 years) to older anatomically predisposed (narrow angle <15°) non-glaucomatous individuals (PACS: age >60 years). In our discovery GWAS, we identified 13 SNPs of *CNTNAP5* that were associated with PACG. Subsequent conditional analyses prioritize SNP rs780010 of *CNTNAP5*, where G allele, was significantly (P=0.0024) associated with higher cup-to-disc ratio, which is a clinical parameter directly correlated with glaucomatous neurodegeneration. We further validated rs780010, in a separate replication cohort and observed a significant association (odds ratio=2.307, P=0.012). Integrative bioinformatic analysis showed that the associated intronic region of *CNTNAP5* shows higher transcriptional regulation activity with active enhancer marks. For functional follow-up of risk allele of rs780010 in *CNTNAP5*, we performed a dual-luciferase assay in HEK293T cells and we observed that G allele showed higher luciference activity. As per genome-wide expression data CNTNAP5 is highly expressed in retinal neurones. In order to determine the function of *CNTNAP5* in ocular development and the morphology of the retinal nerve in zebrafish, we thus used morpholino to knockdown *CNTNAP5* in zebrafish. Immunofluorescence (IF) analyses using acetylated tubulin antibody showed significant eye size reductions and retinal nerve thinning in zebrafish morphants (*CNTNAP5*-MO) compared to 5bp-mismatch morpholino injected control fish. Moreover, in spontaneous movement behavioural analysis, *CNTNAP5*-MO zebrafish have a significantly lower average total distance moved when light was on compared to 5bp-mismatch controls whereas no significant difference in movement was observed when light was off, might indicates peripheral vision loss. Additionally, colocalization IF using cleaved-caspase3 and NeuN antibody showed significant neurodegeneration *CNTNAP5*-MO zebrafish. Therefore, our GWAS results not only indicate a genomic association of the *CNTNAP5* with PACG but also imply that it might play in glaucomatous neurodegeneration. Further, post-GWAS functionalization led us to believe that *CNTNAP5* is an important player for neurodegeneration that further leads to retinal nerve thinning, thereby increasing the risk of PACG-associated vision loss.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1312 A key neuroendocrine stress response gene linked with post traumatic stress disorder is rapidly evolving in humans.

Authors:

J. Pecon-Slattery[1,2], D. O'Rourke[1]; [1]Cornerstone Genomics INC, Frederick, MD, [2]Colorado State Univ., Fort Collins, CO

We investigate the evolution of human exome variation within the Primate Order. Our project focused on identification of those genes that rapidly evolve only in humans compared to our extended family of the Primate Order. We queried the CodeXome® database created from exome data sequenced from 55 of 77 exotic genera within Primates, rooted by Scandentia, and anchored to human exome build HG38 coordinates. Our search identified those genes with the highest number of human-specific mutations relative to the rest of the primate phylogeny. Our results included *ADCYAP1*, adenylate cyclase activating polypeptide 1, a gene that encodes a neuropeptide linked with adrenal stress response. We found this gene to have a unique pattern of evolution within the exome and provide detailed insights on rates of mutation, patterns of selection and potential impact in mediating neuroendocrine stress response in primates. This study represents the first comprehensive comparative evidence for how the gene evolves, and baselines natural variation for interpreting potential variants liked with PTSD.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1313 A Large Meta-Analysis Identifies Genes Associated with Anterior Uveitis

Authors:

S. Gelfman[1], A. Moscati[2], S. Huergo[3], R. Wang[4], V. Rajagopal[1], N. Parikshak[4], V. Kumar[1], E. Chen[5], M. Leblanc[6], R. Hazlewood[1], J. Freudenberg[7], B. Cooper[2], A. Ligocki[2], C. G. Miller[1], T. van Zyl[2], J. Weyne[2], C. Romano[2], B. Sagdullaev[1], O. Melander[8], Penn Medicine BioBank, DiscovEHR, Colorado Center for Personalized Medicine - RGC Collaboration, UCLA-RGC ATLAS Collaboration, Mayo Clinic Project Generation, A. Baras[4], Regeneron Genetics Center, E. Stahl[4], G. Coppola[4]; [1]Regeneron Genetic Ctr., Tarryttown, NY, [2]Regeneron Pharmaceuticals, Tarrytown, NY, [3]Regeneron Pharmaceuticals, Tarryttown, NY, [4]Regeneron Genetics Ctr., Tarrytown, NY, [5]Regeneron, North Brunswick, NJ, [6]Regeneron, Bronx, NY, [7]Regeneron, Tarrytown, NY, [8]Malmo Univ Hosp, Malmo, Sweden

Anterior Uveitis (AU) is the inflammation of the anterior part of the eye, the iris and ciliary body, and accounts for ~80% of Non-Infectious Uveitis. AU is a shared complication of Human Leukocyte Antigen (HLA) class-I-opathies including Ankylosing Spondylitis and Psoriatic Arthritis, and is strongly associated with HLA-B*27 (~20-50% of AU cases are HLA-B*27 positive). We performed the largest AU exome sequencing and meta-analysis to date, including eight independent cohorts consisting of 3,850 ICD defined cases and 916,549 controls. We identified common genome-wide significant loci in *HLA-B* (OR=3.37, p=1.03e-196) and *ERAP1* (OR=0.86, p=1.1e-08). In addition, we found *IPMK* (OR=9.4, p=4.42e-09) and *IDO2* (OR=3.61, p=6.16e-08) as genome-wide significant risk-associated genes based on the burden of rare coding variants. We further divided the cohort into HLA-B*27 positive and HLA-B*27 negative individuals and deciphered the genetic risks for AU in each subgroup.
The *ERAP1* haplotype was strongly protective only for B*27-positive AU (OR=0.73, p=5.2e-10), raising *ERAP1*'s therapeutic potential in the management of AU. Genetic factors affecting the risk for B*27-negative AU also became evident: a novel genome-wide significant common signal near HLA-DPB1 (rs3117230, OR=1.26, p=2.7e-08), *IPMK* and *IDO2* as novel risk genes for B*27-negative AU specifically, and several additional candidate risk genes, including *ADGFR5*, *STXBP2*, and *ACHE*. Taken together, using the largest cohort of AU to date, we deciphered the underlying genetics of B*27-positive and B*27-negative AU and identified novel rare and common genetic signals for both sub-types of disease.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1314 A Locus for Cognitive Preservation on Chromosome 2p11.2-13.1 in the Midwestern Amish

Authors:

L. Main[1], Y. Song[2], R. Laux[2], K. Miskimen[2], M. Cuccaro[3], P. Ogrocki[2], A. Lerner[2], J. Vance[4], M. Fuzzell[2], S. Fuzzell[2], S. Hochstetler[2], M. Osterman[2], A. Lynn[2], D. Dorfsman[4], L. Caywood[4], M. Prough[5], L. Adams[4], J. Clouse[4], S. Herington[6], W. Scott[4], M. Pericak-Vance[7], J. Haines[8]; [1]Case Western Reserve Univ., Lakewood, OH, [2]Case Western Reserve Univ., Cleveland, OH, [3]John P. Hussman Inst. for Human Genomics, Miami, FL, [4]Univ. of Miami, Miami, FL, [5]Univ. of Miami, Bristol, TN, [6]Univ. of Miami Miller Sch. of Med. Hussman Inst. for Human Genomics, Elkhart, IN, [7]Univ. of Miami Miller Sch. of Med., Miami, FL, [8]Case Western Reserve Univ, Cleveland, OH

Most genomic therapeutic targets for Alzheimer disease (AD) focus on previously discovered genetic risk variants. However, protective loci against disease formation are alternative targets for pharmaceutical development. To investigate factors driving cognitive preservation, we examined genetic variation in older cognitively unimpaired individuals at high risk for developing AD, defined as having a first degree relative with dementia. Cognitive status was determined as either cognitively impaired (CI) or cognitively unimpaired (CU) using multiple neurocognitive assessments and confirmed by a clinical adjudication board. We performed genome-wide linkage (GWLS) and association (GWAS) studies in the Midwestern Amish, a closed founder population. 946 study individuals were connected into a 14-generation all-connecting pedigree of 8,222 individuals. This pedigree was used to correct our GWAS for relatedness and generate sub-pedigrees for GWLS. The GWAS was performed using the GENESIS R package for autosomes, and XWAS for the X chromosome. 11 single nucleotide polymorphisms (SNPs) were suggestive (p-value<$1x10^{-4}$) in our association analyses, but no SNP reached genome-wide significance. The GWLS were performed using MERLIN for autosomes and MINX for the X chromosome. These were performed using dominant and recessive parametric models and nonparametric models, repeated in both two-point and multipoint analyses resulting in 106 significant (HLOD>3.3) SNPs. We identified the sub-pedigrees driving these significant results for several loci. Most of these locus-specific sub-pedigrees could be combined into larger pedigrees for localized Markov chain Monte Carlo (MCMC) linkage analyses. The chromosome 2p11.2-13.1 region significant in multiple of our linkage analyses retained robust results (LOD=4.87). This region was not seen as suggestive in our GWAS. After further investigation, a biallelic SNP, rs1402906, defined a binding site for the transcription factor (TF) POU1F1a in this region. The segregating minor allele destroys the binding for POU1F1a. This TF regulates a nearby gene, *LRRTM1*, located in an intronic region of the *CTNNA2* gene, which was detected in previous analyses as a risk locus for AD in a subset of this Amish population. Additionally, shared promoters of *LRRTM1* and *CTNNA2* within our significant linkage region regulate alternative isoforms of *CTNNA2* involved in synapse formation and turnover, highlighting this locus as potentially important in the etiology of AD.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1315 A machine learning association model for hypertension development using gene-based polygenic risk scores (PRS) and lifestyle factors.

Authors:

Y. Hrytsenko[1,2,3], B. Shea[1,2], N. Kurniansyah[2], B. Spitzer[1], T. Kelly[4], X. Zhu[5], H. Chen[6], K. Wiggins[7], X. Guo[8], J. Haessler[9], S. Smoller[10], R. Casanova[11], C. Kooperberg[12], A. Carson[13], A. Morisson[14], S. Redline[2], M. Fornage[15], L. Hou[16], R. Kaplan[10], B. Psaty[7], J. Rotter[17], S. Rich[18], D. Levy[19], T. Sofer[3,1,2]; [1]Beth Israel Deaconess Med. Ctr., Boston, MA, [2]Brigham and Women's Hosp., Boston, MA, [3]Harvard Med. Sch., Boston, MA, [4]Univ. of Illinois at Chicago, Chicago, IL, [5]Case Western Reserve Univ., Shaker Heights, OH, [6]The Univ of Texas Hlth.Sci. Ctr. at Houston, Houston, TX, [7]Univ. of Washington, Seattle, WA, [8]The Lundquist Inst., Torrance, CA, [9]Fred Hutch Cancer Ctr., Seattle, WA, [10]Albert Einstein Coll. of Med., Bronx, NY, [11]Wake Forest Univ. Sch. of Med., Winston-Salem, NC, [12]Fred Hutchinson Cancer Ctr., Seattle, WA, [13]Univ. of Mississippi Med. Ctr., Jackson, MS, [14]The Univ. of Texas Hlth.Sci. Ctr. at Houston Sch. of Publ. Hlth.at Houston, Houston, TX, [15]Univ. of Texas Hlth.Sci. Ctr. at Houst, Houston, TX, [16]Northwestern Univ., Evanston, IL, [17]Lundquist Inst., Harbor-UCLA Med Ctr, Torrance, CA, [18]Univ. of Virginia, Scottsville, VA, [19]NHLBI/NIH, Framingham, MA

Introduction: We constructed and evaluated prediction models for incident hypertension using gene-based polygenic risk scores (PRS) and lifestyle variables.
Methods: We used a gradient boosting trees algorithm (LightGBM) machine learning framework to estimate the probability of incident hypertension, 3-6 years after baseline examination. The data included 12,334 unrelated TOPMed individuals of diverse race/ethnicities, free of hypertension at baseline. We identified genes associated with systolic and diastolic blood pressure (SBP, DBP) by location of associated SNPs (p-value < 5x10-8) from regions identified in multiple multi-ethnic GWAS studies (UKBB+ICBP, MVP, and BBJ). We used PRS-CS to construct PRSs localized to the gene regions. We also included lifestyle factors (smoking status, sleep duration), covariates (age, sex, BMI) and time between the baseline clinic visit to the next exam. We trained four LightGBM models: all models included age, sex, and BMI at baseline. Each of the models did not use one type of variable: 1) used lifestyle measures (smoking status, sleep duration) but did not use PRSs; 2) used PRSs and smoking status but not sleep duration; 3) used PRSs and sleep duration but not smoking status; a final model 4) used lifestyle measures and PRSs. There were N=9,867 individuals in the training dataset, and N=2,467 individuals in the test dataset, in which model performance was assessed. The primary prediction performance metrics was the F1 score, which optimizes both the precision and recall of a prediction model. We evaluated model performance by strata of individuals.
Results: We constructed 231 gene-based SBP and DBP PRSs. Model performance increased with addition of measures, with gene-based PRS having the largest contribution to prediction accuracy: in the combined testing dataset, model 1) without PRSs had F1 score of 2%, model 2) without sleep duration had F1 score of 15%, model 3) without smoking status had F1 score of 13%, while model 4) including all these measures had F1 score of 48%. Model performance varied by strata, with, for model 4), the highest performance being in obese individuals (BMI>30; test dataset N=527; F1 score = 59%), and the lowest performance being in individuals defined with healthy weight (BMI<25; test dataset N=1,010; F1 score = 40%).
Conclusion: Including both lifestyle and genetic measures improves prediction of hypertension development.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1316 A Mendelian Randomization Study of Haematological Traits and Type-2 Diabetes Mellitus in African Ancestry Individuals

Authors:

C. Soremekun[1,2,3], D. Jjingo[4], D. Kateete[1], O. Nash[3], M. Nyirenda[5], D. Gill[6], T. Chikowore[7,8], C. Batini[9,10], O. Soremekun[11,2], S. Fatumo[12,2,13]; [1]Dept. of Immunology and Molecular Biology, Sch. of BioMed. Sci.,Makerere Univ., Kampala, Uganda, [2]The African Computational Genomics (TACG) Res. Group, MRC/UVRI and LSHTM Uganda Res. Unit., Entebbe, Uganda, [3]H3Africa Bioinformatics Network (H3ABioNet) Node, Ctr. for Genomics Res. and Innovation, NABDA/FMST, Abuja, Nigeria, [4]African Ctr. of Excellence in Bioinformatics, Makerere Univ., Kampala, Uganda, [5]MRC/UVRI and LSHTM Uganda Res. Unit, Entebbe, Uganda, [6]Imperial Coll. London - Dept. of Epidemiology and Biostatistics, London, United Kingdom, [7]Med. Res. Council/Wits Univ. Rural Publ. Hlth.and Hlth.Transitions Res. Unit (Agincourt), Sch. of Publ. Hlth., Faculty of Hlth.Sci., Univ. of the Witwatersrand, Johannesburg, South Africa, [8]Wits Donald Gordon Med. Ctr., Sch. of Clinical Med., Faculty of Hlth.Sci., Univ. of the Witwatersrand, Johannesburg, South Africa, [9]Dept. of Population Hlth.Sci., Univ. of Leicester, Leicester, United Kingdom, [10]Leicester Natl. Inst. for Hlth.and Care Res., BioMed. Res. Ctr., Glenfield Hosp., Leicester, United Kingdom, [11]11Inst. of Translational Genomics, Helmholtz Zentrum München – German Res. Ctr. for Environmental Hlth., Neuherberg, Munich, Germany, [12]MRC/UVRI & LSHTM Uganda Res. Unit, Entebbe, Uganda, [13]Dept. of Non-Communicable Disease Epidemiology, London Sch. of Hygiene and Tropical Med., London, United Kingdom

Background: Observational studies have shown associations between haematological traits and type-2 diabetes mellitus (T2D), however, these findings are limited by cofounders. Our study utilizes Mendelian randomization (MR) approach which addresses the above limitation, to investigate the effect of haematological traits like white blood cell (WBC) count, haemoglobin (Hb) levels, and red blood cell (RBC) count on T2D in African ancestry individuals. **Method:** The participating cohort includes participants of African ancestry in the Blood Cell consortium (n = 15,171) and the Million Veteran Program dataset (n = 53, 445). We applied a univariable, bidirectional, and multivariable 2-sample MR to estimate the causal relationship between 15 haematological traits and T2D using GWAS summary statistics. Genetic instruments for haematological traits and T2D were selected and these included only the variants that attained a genome-wide significant level of $P \leq 5 \times 10^{-8}$. The inverse-variance weighted MR approach was used to estimate causal effects in the main analysis, in addition to sensitivity analyses that are robust to pleiotropic variants. **Results:** In the main inverse-variance weighted (IVW) MR estimates, genetically determined mean corpuscular hemoglobin concentration (MCHC), mean corpuscular hemoglobin (MCH) and mean corpuscular volume (MCV) levels were associated with decreased risk of T2D with an OR_{IVW} of 0.867 per g/dL difference in MCHC levels (95% CI: 0.803-0.937; $P_{IVW} = 0.00028$), OR_{IVW} of 0.888 per pg difference in MCH levels (95% CI: 0.835-0.943; $P_{IVW} = 0.00012$) and OR_{IVW} of 0.911 per fL difference in MCV levels (95% CI: 0.856-0.969; $P_{IVW} = 0.0313$) respectively. While genetically high levels of RBC were significantly associated with an increased risk of T2D with OR_{IVW} of 1.105 per 10^{12}/L difference in RBC levels (95% CI: 1.027-1.189; $P_{IVW} = 0.0076$). For WBC traits, there was evidence of protective effects of total WBC count (WBC) [OR= 0.941 (CI: 0.909-0.974, P= 0.0006] and neutrophil count (NEU) [OR= 0.940 (CI: 0.917-0.965, P= .95E-06] on T2D. In the reverse analysis, genetically determined T2D was associated with lower levels of RBC distribution width (RDW) and elevated levels of lymphocyte count (LYM). The multivariable analysis showed that genetically predicted MCH, MCHC, RDW, MCV, and lower levels of WBC were significantly associated with a decreased risk of T2D. **Conclusion:** These findings suggest that haematological traits play a role in the development of T2D. However, further research is needed to validate and explore the biological pathways involved in these associations.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1317 A Multi-Ancestry Meta-Analysis of GWAS for Mitral Valve Prolapse Identifies 34 Novel Genomic Regions

Authors:

A. Small[1], A. Georges[2], J. Lee[3], C. Roselli[4], D. Milan[5], P. T. Ellinor[6], Y. Sun[7], K. Cho[8], P. Wilson[7], A. Ganna[9], G. Peloso[10], N. Bouatia-Naji[11], P. Natarajan[12]; [1]Brigham and Women's Hosp., Boston, MA, [2]Université de Paris, Inserm U970, Paris, France, [3]Broad Inst. of MIT and Harvard, Finnish Inst. for Molecular Med., Cambridge, MA, [4]Broad Inst., Cambridge, MA, [5]Leducq Fndn., Boston, MA, [6]Massachusetts Gen. Hosp. / The Broad Inst., Boston, MA, [7]Emory Univ., Atlanta, GA, [8]VA Boston Hlth.care System, Boston, MA, [9]Helsinki Inst. of Life Sci. HiLIFE, Helsinki, Finland, [10]Boston Univ., Boston, MA, [11]INSERM, Paris, France, [12]Broad Inst. of Harvard and MIT, Cambridge, MA

Introduction: Mitral valve prolapse (MVP) is a common pathology of the mitral valve characterized by thick, redundant mitral valve leaflet tissue. MVP manifests clinically as mitral regurgitation, and carries significant morbidity including valve replacement, heart failure, and sudden cardiac death. The genetics of MVP are only partially understood.
Methods: We performed a meta-analysis of genome-wide association studies (GWAS) for MVP incorporating 14,733 MVP cases among 1,485,696 individuals from the Million Veteran Program, Finngen, and a recently published GWAS of MVP by Roselli et al. We functionally annotated potential causal variants at MVP loci by evaluating for missense variants and variants with potential regulatory functions in mitral valve tissue (both normal and diseased) or cardiac related tissues (heart and aorta). Results were extended with transcriptome wide association analysis (TWAS) incorporating cardiovascular relevant tissues from the GTEx database.
Results: There were 49 unique genome-wide significant (GWS) genomic regions, of which 34 were novel. Potential causal variants with relevance in mitral valve tissue or heart tissue were identified in 47 out of 49 genomic regions. Of these, 25 loci included either missense variants or variants overlapping a promoter active in valve tissue as suggested by ATAC-seq data. MVP-associated variants overlapping candidate distal cis-regulatory elements active in either mitral valve tissue or heart tissue were identified in the remaining 22 loci. TWAS corroborated gene-expression associations for 24 loci, and further prioritized several sub-GWS loci, including *MFAP2*, which encodes an important extracellular matrix (ECM) protein. Novel genes corresponded to key pathways in TGF-ß signaling (*SMAD3*), ECM remodeling (*TIMP3*, *ADAMTS10*, *PLEC*, *LOXL1*, *SPTBN1*), and inflammation (*PDLIM7*).
Conclusions: We performed the largest multi-ancestry GWAS to-date of MVP, identifying 34 novel genomic regions with key functions in TGF-ß signaling, ECM remodeling, and inflammation.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1318 A new gene expression environmental decomposition method applied for the study of T2D in Mexican Americans

Authors:

M. Almeida, E. G. MANUSOV, V. Diego, X. Mao, S. Laston, A. Leandro, J. Blangero, S. Williams-Blangero; Univ. of Texas Rio Grande Valley, Brownsville, TX

Background: Type 2 diabetes (T2D) is a complex metabolic disease and Mexican Americans are specially affected by this condition. Many studies were proposed for the identification of candidate genes associated with T2D risk, but their combined contribution substantially fail to account for the total T2D's heritability estimate. Gene expression, like any other quantitative phenotype, is the sum of genetic and environmental variances. Gene-by-environment (GxE) interaction analysis may contribute to the identification of additional candidate genes. We propose a new method to estimate the environmental contribution to the expression of candidate genes and the identification of T2D risk genes on a large cohort of Mexican Americans in south Texas.
Methods: A total of 525 Mexican American participants from an extended family study were evaluated and assessed for obesity, diabetes, hypertension, hyperlipidemia, and depression. mRNA sequencing was performed on blood samples collected from participants. For gene expression, we wrote a linear model $\boldsymbol{y=XB + g + e}$ where $\boldsymbol{X;}$ is a matrix of covariates, $\boldsymbol{B}$ is the intercept, $\boldsymbol{g}$ and $\boldsymbol{e;}$ are respectively genetic and environmental effects. The inverse normalized residuals are denoted $\boldsymbol{D= (y - XB)}$**.** The phenotypic covariance matrix $\boldsymbol{\Sigma}$; defined as: $\boldsymbol{\Sigma; = Kh^2 + I(1-h^2)}$, where **K**; and **I**; respectively give genetic relationship and identity matrices, and $\boldsymbol{h^2}$ is the heritability. The random genetic value of individual *I* is defined as $\boldsymbol{gi= h^2 K_i \Sigma;^{-1}_{-i} D_{-i}}$ where $\boldsymbol{k_i}$ and $\boldsymbol{\Sigma;^{-1}_{-i}}$ where terms where individual i was excluded. The environmental effects vector (EEV) of each gene defined as the subtraction of genetic effects term from the residuals vector.**Results:** Significant associations were found between PNMA1 ($1.2*10^{-5}$), TPRX1($7.3*10^{-5}$), GRIK3 ($1.1 * 10^{-4}$) and CTLA4 ($1.7* 10^{-4}$) and T2D status. These findings support the role of genetic factors interacting with the environment and influencing T2D status of Mexican Americans individuals.**Discussion:** CTLA4 gene is a member of the immunoglobulin superfamily and encodes a protein which transmits inhibitory signals to T cells. Mutations in this gene have been associated with insulin-dependent diabetes mellitus. The GxE interactions identified in this study provide insight into the complex interplay between environment (lifestyle and diet) and the expression of target genes. CTLA4 expression may serve as a potential diagnostic marker for T2D in Mexican American population.**Conclusion:** Our study highlights the importance of GxE interactions in T2D among Mexican Americans. The associations between CTLA4 and T2D provide valuable insights into the inflammation pathways to this condition.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1319 A novel non-recurrent CNV deletion involving *TBX4* leaving *TBX2* intact, in a newborn with congenital alveolar dysplasia.

Authors:

J. Karolak[1], K. Bzdęga[1], M. Biela[2], G. Deutsch[3], J. Kitzmiller[4], M. Rydzanicz[5], R. Płoski[5], J. Whitsett[4], R. Śmigiel[2]; [1]Poznan Univ. of Med. Sci., Poznan, Poland, [2]Med. Univ. of Wroclaw, Wroclaw, Poland, [3]Univ. of Washington Sch. of Med., Seattle, WA, [4]Cincinnati Children's Hosp. Med. Ctr., Cincinnati, OH, [5]Med. Univ. of Warsaw, Warsaw, Poland

Congenital alveolar dysplasia (CAD) belongs to rare lethal lung developmental disorders (LLDDs) in neonates that manifest as acute progressive respiratory failure and pulmonary arterial hypertension refractory to treatment. While largely genetically undefined, recently, a number of cases of CAD have been reported to result from abnormalities involving 17q23.1q23.2 or 5p12. Most copy-number variant (CNV) deletions at 17q23.1q23.2 were recurrent and encompassed two closely located genes *TBX4* and *TBX2*. In mice, depletion of *Tbx4* and *Tbx2* were found to lead to hypoplastic lungs. However, one intragenic frameshifting CNV deletion and four single-nucleotide variants (SNVs) involving *TBX4*, but not *TBX2*, have been described in LLDD neonates. Our recent histopathological and transcriptomic analyses of *TBX4*-deficient LLDD lungs have revealed that TBX4 can regulate expression of the lung endothelial *TMEM100* gene that is also involved in lung development. Here, using molecular and histopathological evaluation, we studied a newborn with clinically suspected LLDD. A male proband born from uncomplicated pregnancy in good condition and without any structural malformations, from the first hours of life manifested severe respiratory insufficiency refractory to treatment and he died at 27th day of life. Family trio genome sequencing revealed in the proband a novel de novo ~1.07 Mb heterozygous non-recurrent CNV deletion at 17q23.2 (chr17:61,451,736-62,521,987, hg38), encompassing *TBX4* but leaving *TBX2* intact. Histopathological evaluation demonstrated lobular maldevelopment with lung growth arrest along the spectrum of CAD, pulmonary arterial hypertrophy, and superimposed acute alveolar inflammation. Immunofluorescence confocal microscopy revealed lack of sacculation and atypical lung tubules lined by AT2-like cells staining intensely for SFTPB and ABCA3; PECAM staining demonstrated abnormal pulmonary vessels without close apposition to respiratory epithelial cells. Immunostaining for TMEM100 in capillaries was significantly reduced compared to a control lung. Our results indicate that perturbations of *TBX4*, rather than *TBX2*, can cause severe lung phenotypes in humans and further imply the possible molecular interactions between TBX4 and TMEM100 that, when disrupted, can lead to lung disease.

Support: National Science Centre, Poland, 2019/35/D/NZ5/02896

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1320 A Pilot project: Examining gender-specific transcriptional alterations in Myalgic Encephalomyelitis/Chronic Fatigue Syndrome patients under stressful conditions.

Authors:

R. Pangeni[1], J. Gamer[1], D. Van Booven[2], O. Zarnowski[1], M. Perez[3], N. Klimas[1], E. Oltra[4], L. Nathanson[1]; [1]Nova Southeastern Univ., Fort Lauderdale, FL, [2]Univ Miami Miller Sch Med, Miami, FL, [3]3321 Coll. Ave.,, Fort Lauderdale, FL, [4]Catholic Univ. of Valencia, Valencia, Spain

Background Myalgic Encephalomyelitis/Chronic Fatigue Syndrome (ME/CFS) is characterized by symptoms such as profound fatigue, cognitive difficulties, muscle and joint pain, gastrointestinal disturbances, neurological impairments, and hormonal irregularities. The precise mechanisms involved in the development and advancement of this disorder are not yet fully understood. **Objectives** This pilot research project aims to assess how sex differences in individuals with ME/CFS impact their response to stress, specifically through an exercise challenge. Additionally, it highlights the importance of considering sex differences in the diagnosis and treatment of ME/CFS. **Methods** We utilized RNA-seq to examine the differential gene expression (DG) in peripheral blood mononuclear cells of 20 female ME/CFS patients, 14 male ME/CFS patients, and 40 healthy controls (HC) who were matched in for sex, age, and BMI. We evaluated the gene expression changes at three time points: T0 (baseline), T1 (at maximal exertion), and T2 (4 hours into recovery after T1). To ensure consistency, we excluded genes located on the sex chromosomes and performed separate analyses for men and women. Our focus was to compare the response to exercise within each group (ME/CFS patients and HC) between different time points. **Results** In the male ME/CFS cohort, pathways associated with immune cell signaling, particularly IL-12, and natural killer cell cytotoxicity were activated during exercise. However, the female ME/CFS patients did not exhibit significant changes in gene expression that met the study's criteria for inclusion. During the recovery period after exercise, male ME/CFS patients demonstrated distinct alterations in the regulation of specific cytokine signals, including IL-1β. On the other hand, female ME/CFS patients displayed significant changes in gene networks related to cell stress, response to herpes viruses, and NF-kβ signaling. These functional pathways and differentially expressed genes identified in this pilot project provide valuable insights into the sex-specific pathophysiology of ME/CFS. **Conclusion** The identification of sex-specific biomarkers and therapeutic targets in ME/CFS holds great potential for understanding the distinct onset and progression of this disease in different sexes.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1321 A polygenic risk score explains variation in monogenic autoimmunity presentation.

Authors:

A. Luckett, G. Hawkes, M. Weedon, R. Oram, M. Johnson; Univ. of Exeter, Exeter, United Kingdom

Introduction: Rare mutations in single genes can lead to syndromes of autoimmunity. Individuals with those mutations present with a variety of autoimmune disorders, including early-onset autoimmune diabetes (diagnosis typically <2 years), though not all develop the disease. We aimed to assess whether polygenic risk, which has recently been shown to have a modifying effect in other monogenic disorders, could explain this variable phenotypic presentation.
Method: We calculated a type 1 diabetes genetic risk score (T1D-GRS) in individuals in the Extremely Early-onset Type 1 diabetes (EXE-T1D) cohort with monogenic autoimmune diabetes (n=71, *AIRE* n=3, *FOXP3* n=36, *IL2RA* n=5, *LRBA* n=11, *STAT3* n=7, *TNFAP3* n=1 and Trisomy 21 n=8), monogenic autoimmunity without diabetes (n=56, *AIRE* n= 54, *FOXP3* n= 1, *STAT3* n=1), and non-autoimmune monogenic neonatal diabetes (NDM) (n= 200, *ABCC8* n=22, *EIF2AK3* n= 34, *INS*= 74, *KCNJ11* n=70). The T1D-GRS model included 67 SNPs (35 HLA and 32 non-HLA variants). We investigated the constituent parts of the T1D-GRS model to assess the contribution of HLA and non-HLA common variants to the phenotype of monogenic autoimmune diabetes.
Results: Individuals with monogenic autoimmune diabetes had higher T1D-GRS (mean = 11.6, 95% CI 11.1-12.1) compared to those with monogenic autoimmunity without diabetes (mean=10.8, 95% CI 10.3-11.3, *P*=0.03) and non-autoimmune NDM (mean= 10.0, 95% CI 9.7-10.3, *P*=1.3e-07). Individuals with monogenic autoimmune diabetes had a higher HLA class II genetic risk score compared to those with non-autoimmune NDM (beta = 0.82 SD, 95% CI 0.48-1.16, *P*= 1.9e-06).
Conclusions: A polygenic risk score for type 1 diabetes explains variation in monogenic autoimmunity expressivity, partly through higher HLA class II risk. How the monogenic and polygenic risks interact is an area for future study and may lead to improved understanding of disease aetiology.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1322 A risk locus on chromosome 12 is suggested by genome-wide association study in Puerto Ricans for Alzheimer disease

Authors:

K. Hamilton-Nelson[1], B. Feliciano-Astacio[2], B. Akgun[1], J. Rivero[1], K. Celis[1], L. Adams[1], V. Rodriguez[1], P. Mena[1], S. Tejada[1], P. Whitehead[1], P. Bussies[1], M. Prough[1], H. Acosta[3], A. Griswold[1,4], C. Dalgard[5], G. Beecham[1,4], K. McInerney[6], J. Vance[1,4,6], M. Cuccaro[1,4], F. Rajabli[1,4], M. Pericak-Vance[1,4,6]; [1]John P Hussman Inst. for Human Genomics, Univ. of Miami Miller Sch. of Med., Miami, FL, [2]Univ. Central del Caribe, Bayamon, PR, [3]Clinica de la Memoria, San Juan, PR, [4]Dr John T Macdonald Fndn. Dept. of Human Genetics, Univ. of Miami Miller Sch. of Med., Miami, FL, [5]Dept. of Anatomy, Physiology, and Genetics, Uniformed Services Univ. of Hlth.Sci., Bethesda, MD, [6]Dept of Neurology Univ. of Miami Miller Sch. of Med., Miami, FL

Background: Increasing ancestral diversity in genetic studies is important to define the genetic background of Alzheimer disease (AD). Hispanic populations are underrepresented in AD genetic studies. The Puerto Rican (PR) population, the second largest Hispanic group in the continental US, provides an opportunity to assess Hispanic risk for AD. In addition, the PR population is admixed, with average ancestry values of 69% European (EU), 17% African (AF), and 14% Native American (AI). We performed a genome-wide association study (GWAS) in the PR population to identify novel AD susceptibility loci and characterize known AD genetic risk loci.
Methods: The PR dataset includes Whole Genome Sequencing (WGS) and phenotype data from 639 individuals (334 cases; 305 cognitively unimpaired (CU)). We employed a generalized linear mixed model for the GWAS analysis using SAIGE software. Two separate models were analyzed; the first model accounted for sex, age, and population substructure, and the second model also included the dosage of *APOEe4*. A genetic relationship matrix was included as a random effect to account for relatedness in both models. We assessed the polygenic risk score (PRS) using the effect sizes from summary statistics from the non-Hispanic White (NHW) study.
Results: Variants at *APOE* locus ($pv=7.2x10^{-8}$; chr19) and 10 additional loci reached suggestive significance ($pv<1x10^{-6}$); however, there was no genome-wide significance ($pv<5x10^{-8}$) region and only 3 loci were significant in both models: *NACC2*, *AKR1C2*, and *SCN8A*. In addition, 13 known AD loci - *ADAM17*, *ANKH*, *RASGEF1C*, *TREM2*, *SEC61G*, *CLU*, *TSPAN14*, *TPCN1*, *FERMT2*, *DOC2A*, *PRDM7*, *SCIMP*, and *WNT3*- showed nominal significance. NHW-derived PRS has a good prediction power (AUC=0.62) in the PR dataset.
Conclusions: PR GWAS identified a promising AD risk locus in *SCN8A*, a gene associated with reduced pathogenesis of AD in previous mouse model studies. Our results further corroborated chr 12 findings from previous linkage studies. In addition, our study replicated 13 known AD loci, and showed good predictive power with NHW-derived PRS, suggesting that may be related to the high EU ancestral background of the PR population. Including genetic studies for underrepresented and ancestral-heterogeneous populations such as the PR, provides an important opportunity to evaluate the role of admixed populations in AD.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1323 A sex-stratified analysis of the genetic architecture of human brain anatomy

Authors:

r. shafee[1], D. Moraczewski[1], S. Liu[2], T. Mallard[3], A. Thomas[1], A. Raznahan[4]; [1]Natl. Inst. for Mental Hlth.(NIMH), Bethesda, MD, [2]Natl. Inst.s of Mental Hlth., Bethesda, MD, [3]Massachusetts Gen. Hosp., Boston, MA, [4]NIMH Intramural Res. Program, NIH, Bethesda, MD

Large open biobanks have dramatically advanced our understanding of genetic influences on human brain anatomy - but much of this work has combined rather than compared males and females. Here, we screen for potential sex-biases in the influence of common genetic variants on brain anatomy.T1-weighted magnetic resonance images in the UK Biobank (14301 males, 16054 females of “white British ancestry”, age = 64.3 +/- 7.5 yrs) were used to obtain (FreeSurfer) gray matter volume (GMV), surface area (SA) and cortical thickness (CT) for 360 cortical regions (HCP atlas, covaried for total GMV, SA and CT respectively). Additionally, 23 subcortical volumes and mean cortical thickness, total surface area and total brain volume were included. Genotypic data consisted of 17 million SNPs after standard quality control. For each phenotype, I) sex-specific relatedness matrices (GRMs) in GCTA were used to calculate SNP-heritability (h^2) for each phenotype in males and females separately; II) GCTA was also used to estimate autosomal genetic correlation (r_g) between the sexes, and III) sex-stratified GWASs were performed in PLINK to estimate SNP-level sex-differences. SNPs with significant sex-differentiated effects were mapped to genes using FUMA which in turn were analyzed for enrichment of GO categories. Two p-values cutoffs were used for significance at this step: 1.4e-10 ("stringent" - corrected genome wide and across phenotypes) and 5e-8 ("relaxed" - genome wide).After correction for multiple comparisons, no individual phenotype showed a statistically significant sex-difference in h^2. However, paired t-tests between the sexes revealed that mean h^2 for regional GMV and SA tended to be higher in females than males (p = 6.5e-10 , p = 4.3e-6 respectively). In GCTA bivariate analyses, 2 phenotypes showed $r_g < 1$ between the sexes surviving correction for multiple comparisons (Posterior Insula (left) r_g = 046 +/- 0.11, Brodmann area 5m r_g = 0.50+/-0.11). Of the two SNPs that showed significant sex-difference in GWAS at the most stringent cut-off, one (chr16:rs113078989; SA of supplementary motor region; male-specific effect) mapped to RBFOX1, a gene linked to multiple neuropsychiatric and neurodevelopmental disorders. Combined enrichment analysis of the genes mapped from SNPs passing the "relaxed" threshold did not show enrichment of any gene ontology category. Notably, sex-biased SNPs were not enriched on the X-chromosome or in sex-steroid related pathways. This work suggests that common variant influences on human brain anatomy are largely convergent between males and females, with a few exceptions that will guide future research as biobanks continue to grow in size.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1324 A systematic analysis of the shared and specific genetic signals among 15 autoimmune diseases.

Authors:

X. Dang[1], Y. Wang[2], W. Yang[1]; [1]The Univ. of Hong Kong, Hong Kong, China, [2]The Chinese Univ. of Hong Kong, Shenzhen(CUHK-SZ), China

Genome-wide association studies (GWAS) have discovered thousands of genetic loci associated with various autoimmune diseases, and nearly half of these loci are shared by multiple traits. These shared loci provide insights into the common genetic factors contributing to the development of autoimmune diseases. However, the specific or shared association mechanisms for these diseases require further investigation.

Our study conducted a comprehensive analysis of the association signals for the 15 most common autoimmune diseases using data from the GWAS catalog. In particular, we obtained 4,017 disease-associated variants and defined loci and shared and independent association signals based on linkage disequilibrium(LD). Our analysis identified 353 loci that are associated with at least two autoimmune diseases out of 620 loci for all 15 autoimmune diseases, and then clustered 325 pleiotropy signals and 1,261 disease-specific signals(R2 >= 0.8). Since the majority of variants are located within noncoding regions, it is challenging to identify target genes, cell types, and biological functions. Therefore, we integrated multimodal genomic data and used five evidence-based strategies to identify and prioritize the candidate target genes of these signals. We also applied pathway analysis and gene network construction to analyze these target genes of pleiotropic and disease-specific signals, and compared the difference and similarity of autoimmune diseases.

Through evaluating all signal-target gene pairs, we analyzed the shared and specific genes, pathways, and cell types among these 15 autoimmune diseases. We identified that about 41.2% and 27.4% of the target genes were shared by at least two and three autoimmune diseases, respectively, in which *STAT4, IL12RB2,* and *BACH2* were the most pleiotropic genes shared by at least nine diseases. Analyzing all these signal-linked target genes, we found they were enriched in the T-cell differentiation, immunological disease, and virus infection pathways. Our systematic analysis of association signals from target genes, regulated immune cells, and pathways using multimodal genomic data for 15 autoimmune diseases reveals the similarities and differences among these diseases, and open a window for facilitating further functional characterizations, drug repurposing and discovering novel drug targets.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1325 A trans-ancestry clinical validation of a polygenic score for height in a cohort of 13,194 children

Authors:

J. Shelley[1], M. Shi[2], S. Van Driest[3], J. Simmons[2], J. Mosley[2]; [1]Vanderbilt Univ. Sch. of Med., Nashville, TN, [2]Vanderbilt Univ. Med. Ctr., Nashville, TN, [3]NIH, Bethesda, MD

Background: Short stature in children, defined as a height more than 2 standard deviations below expectations, may be an early manifestation of systemic disease. The first step in the evaluation for short stature is to determine whether the child has familial/genetic short stature, defined by a deviation between the child's measured height and expected height based on the parental heights. Contemporary polygenic scores (PGS) for height capture up to 40% of variance in adult heights. We tested whether a trans-ancestry PGS for height associates with meeting short stature criteria and whether the PGS associates with familial short stature.

Methods: We utilized a published trans-ancestry PGS (967,127 SNPs) from a GWAS of height in 5.4 million adults. We curated a retrospective cohort of 13,194 children (ages 2-19) of diverse ancestries (68.9% European [EUR], 22.8% African [AFR], 5.9% Hispanic [HIS], and 2.4% Asian [AS]) using VUMC's EHR-linked biobank (BioVU). We used linear and logistic regression to test the association between the PGS and (1) measured height and (2) having a height qualifying as short stature, respectively. AUROC analyses were used to test the ability of a PGS to discriminate between children with and without short stature. We then identified 432 children with parental height measurements who met the criteria for short stature. We used logistic regression to test whether the PGS and mid-parental height (MPH) was significantly different between familial and non-familial short stature. Asian ancestry was excluded from this analysis as there were fewer than 10 individuals with parental heights available. All analyses were adjusted for age, sex, and within-ancestry principal components (PCs).

Results: The PGS explained 15.3% (EUR), 6.8% (AFR), 7.5% (HIS), and 3.9% (AS) of the variation in children's heights. The PGS was significantly associated with short stature: EUR (OR per s.d. decrease in PGS=1.74 [95%CI: 1.61-1.89], AUC=0.68), AFR (1.79 [1.37-2.35], AUC=0.64), HIS (2.11 [1.43-3.17], AUC=0.68), and AS (1.93 [1.16-3.22], AUC=0.68). In the cohort with parental height data available, 97 (22.4%) met criteria for familial short stature (FSS). The PGS was not significantly associated with FSS in children of European (OR per s.d. decrease in PGS=1.02 [0.78-1.32], p=0.89), African (2.17 [0.58-1.20], p=0.29) or Hispanic (1.33 [0.20-9.14], p=0.75) ancestries.

Conclusion: While a trans-ancestry PGS associates with short stature and can discriminate between those with and without short stature, it does not associate with familial short stature among a cohort of children with short stature.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1326 A weak effect of *PIEZO1* short tandem repeat deletions on severe malaria susceptibility estimated in 18,000 samples.

Authors:

E. Leffler[1,2,3], G. Band[2,3], A. E. Jeffreys[2], K. Rowlands[2], C. Hubbart[2], K. Bojang[4], D. J. Conway[5], M. Jallow[6], F. Sisay-Joof[4], U. D'Alessandro[4], C. M. Ndila[7], A. macharia[7], K. Marsh[7,8], T. N. Williams[7,9], D. Kachala[10], M. Molyneux[10,11], V. Nyirongo[10], T. Taylor[10,12], K. Rockett[2,3], D. P. Kwiatkowski[2,3,13]; [1]Univ. of Utah, Salt Lake City, UT, [2]Wellcome Ctr. Human Genetics, Univ. of Oxford, Oxford, United Kingdom, [3]Wellcome Sanger Inst., Hinxton, United Kingdom, [4]MRC Unit, The Gambia at LSHTM, Fajara, Gambia, [5]London Sch. of Hygiene and Tropical Med., London, United Kingdom, [6]Royal Victoria Teaching Hosp., Banjul, Gambia, [7]Kenyan Med. Res. Inst. (KEMRI)--Wellcome Trust Res. Programme, Kilifi, Kenya, [8]Nuffield Dept. of Med., Univ. of Oxford, Oxford, United Kingdom, [9]Dept. of Med., Imperial Coll., London, United Kingdom, [10]Malawi-Liverpool--Wellcome Trust Clinical Res. Programme, Queen Elizabeth Central Hosp., Coll. of Med., Blantyre, Malawi, [11]Liverpool Sch. of Tropical Med., Liverpool, United Kingdom, [12]Coll. of Osteopathic Med., Michigan State Univ., East Lansing, MI, [13]Big Data Inst., Univ. of Oxford, Oxford, United Kingdom

PIEZO1 is a mechanosensitive ion channel involved in regulating hydration levels in red blood cells, an important factor in their susceptibility to infection by malaria parasites. E756del, a deletion allele within a short tandem repeat (STR) in *PIEZO1*, is common in many African populations and has been proposed to influence susceptibility to malarial disease, but association evidence has been inconsistent. Here, we use Illumina sequencing of amplicons covering the *PIEZO1* STR to genotype 5,198 severe malaria cases and 8,176 population controls from The Gambia, Kenya, and Malawi. We find no strong evidence for association but estimate the same direction of effect for E756del as two previous studies. A meta-analysis across all three (7,864 cases and 10,105 controls) indicates weak evidence for a protective effect (OR=0.92, P=0.011). This unifies previous reports and supports a much smaller effect than originally suggested. Finally, although E756del is at higher frequency in African populations than other regions including Europe, which could be consistent with a selective effect, we show that this frequency differentiation is not exceptional compared with STR variants genome wide. Ultimately, our study reinforces the importance of large association studies to achieve sufficient power to estimate smaller effects and suggests that many such variants likely remain to be uncovered for malaria susceptibility.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1327 ABO gene, *Faecalicatena lactaris*, and Covid-19 Susceptibility and Clinical Outcomes

Authors:

W. Xian[1], V. Napolioni[2], J. Huang[1,3]; [1]Southern Univ. of Sci. and Technology, Shenzhen, China, [2]Sch. of BioSci.s and Vet. Med., Univ. of Camerino, Camerino, Italy, [3]Inst. for Global Hlth.and Dev., Peking Univ., Beijing, China

Background: The gut microbiota has been reported to be associated with both non-communicable diseases and infectious diseases including Covid-19. Emerging large GWAS on tens of microbiota identified strong host genetics loci including ABO, which has been reported for significant association with Covid-19 susceptibility and clinical outcomes. **Methods**: We conducted two-sample Mendelian randomization (MR) studies using latest genome-wide association study (GWAS) summary statistics. For the exposure, we focused on *Faecalicatena lactaris*, *Bifidobacterium* and *Enterococcus faecalis*, which are strongly influenced by host genetics. For the instrumental variable, we used the lead single nucleotide polymorphism (SNP) that is associated with each of these three bacteria. For Covid-19 outcome, we included GWAS summary statistics of three phenotypes (susceptibility, hospitalization, severity) made publicly available by the Covid-19 host genetics initiative (HGI, release 7), and two Covid-19 GWAS with over 24,000 Covid-19 critical illness cases (non-HGI). TwoSampleMR R package (v0.5.6) was used for the data harmonization, and MR Base (v0.5.0) was used for calculating the point estimate (Wald Ratio) of causation. For sensitivity analyses, we included more instrumental variants by lowering P-value threshold and by applying methods including inverse variance weighted method, maximum likelihood method, weighted median method, and MR-Egger. We further applied colocalization analyses in the ABO locus to establish the causal link down to the expression quantitative trait locus (eQTL) level. **Results**: Among the 15 pairs (3 bacteria x 5 Covid-19 phenotypes) we tested, we found significant causal association between the *Faecalicatena lactaris* and all five Covid-19 phenotypes: HGI susceptibility ($p = 1.48\times10^{-11}$), HGI hospitalization ($p = 3.95\times10^{-23}$), HGI severity ($p = 4.45\times10^{-80}$), non-HGI critical illness ($p = 7.03\times10^{-12}$), non-HGI infection ($p = 4.47\times10^{-19}$). The effect size is all positive, indicating the abundance of *Faecalicatena lactaris* increases the likelihood of Covid-19 susceptibility and clinical outcomes. Sensitivity analyses did not detect any heterogeneity and horizontal pleiotropy. Colocalization analyses showed that the *Faecalicatena lactaris* is mainly regulated by the expression of ABO locus. **Conclusion**: Based on latest GWAS of microbiota and Covid-19, this study points to a plausible mechanism on how*Faecalicatena lactaris* affects directly Covid-19 susceptibility and clinical outcome and also mediates the effect of ABO blood groups.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1328 Abstract: Investigating the causal relationship between glycemic trait and cardiac structure/function using Mendelian randomization

Authors:

T. Daria[1,2], H. Alkhairo[1,2], M. G. Levin[3,4], S. M. Damrauer[3,5], P. S. Tsao[6,7], J. R. Priest[7,8], J. P. Pirruccello[9], J. B. Echouffo Tcheugui[10], C. Tcheandjieu[11,12]; [1]Gladstone Inst., San Francisco, CA, [2]Univ. of California San Francisco, San Francisco, CA, [3]Corporal Michael Crescenz VA Med. Ctr., Philadelphia, PA, [4]Div. of Cardiovascular Med., Dept. of Med., Univ. of Pennsylvania Perelman Sch. of Med., Philadelphia, PA, [5]Dept. of Surgery and Genetics, Univ. of Pennsylvania Perelman Sch. of Med., Philadelphia, PA, [6]VAPAHCS/Stanford Univ., Palo Alto, CA, [7]Dept. of Med., Stanford Univ., Stanford, CA, [8]Tenaya Therapeautics, South San Francisco, CA, [9]Div. of Cardiology, Univ. of California San Francisco, San Francisco, CA, [10]Dept. of Med., John Hopkins Univ. Sch. of Med., Baltimore, MD, [11]Gladstone Inst., San Francisco, CA, [12]Dept. of Epidemiology and Biostatistics, Univ. of California of San Francisco, San Francisco, CA

Impaired glucose regulation is an established risk factor for cardiovascular diseases. However, it remains unclear whether this relationship is causal. Here, we investigated the relationship between glucose exposure and MRI-derived measures of cardiac structure and function using Mendelian randomization (MR).
We used an inverse-variance weighted method to perform a two-sample MR using summary statistics from genome-wide association studies of glycemic traits (fasting insulin FI, fasting glucose FG, glycated hemoglobin HbA1c, 2-hour glucose after oral challenge 2hGlu), type 2 diabetes (t2d) and MRI-derived measures of cardiac structure/function (left ventricular ejection fraction (LVEF), left ventricular end-systolic volume (LVESV), diastolic (LVEDV) or stroke volumes (LVSV) myocardial interstitial fibrosis (MIF)). We used the MR-Egger method for sensitivity analysis and assessed pleiotropy using the MR-PRESSO method.
We found evidence for associations between glycemic traits and cardiac structure/function measures. Genetically predicted increased FG and HbA1c were associated with increased LVEF (beta=0.11, p=0.02, beta=0.15, p=0.03 for FG and HbA1C respectively) and decreased LVESV (beta=-0.11, p=0.01, beta=-0.16, p=0.009 for FG and HbA1C respectively). FI was associated with decreased LVESV (beta=-0.25, p=0.007), LVEDV (beta=-0.21, p=0.05) and MIF (beta=-0.2, p=0.02). Lastly, t2d was associated with a decrease in LVESV (beta=-0.03, p=0.01), LVEDV (beta=-0.03, p=0.004) and LVSV (beta=-0.03, p=0.004).
Genetic evidence supports the effects of glycemic traits on left ventricular structure. These findings provide additional causal evidence to support the epidemiological evidence that high glucose exposure is associated with cardiac dysfunction and the beneficial effects of glucose-lowering medications on cardiac function recently demonstrated in large randomized controlled trials. Further research is needed to better understand the mechanisms behind these relationships.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1329 Accelerating psychiatric genetics research in Latin American populations, Latin American Genomics Consortia efforts and Novel findings

Authors:

J. Montalvo-Ortiz[1], P. Giusti-Rodriguez[2], E. Atkinson[3], Latin American Genomics Consortium; [1]Yale Univ., Orange, CT, [2]Univ. of Florida Coll. of Med., Gainesville, FL, [3]Baylor Coll. of Med., Houston, TX

There is a pressing need to engage traditionally underrepresented research participants and researchers to diversify both genomics studies and the genetics workforce. For this, collaborative and coordinated efforts are crucial. The Latin American Genomics Consortium (LAGC), established in 2019, aims to accelerate psychiatric genetics research and expand the genomics workforce of Latin America. The LAGC currently has over 150 active members, representing 9 different Latin American countries. LAGC has disorder-specific working groups as well as a methods working group focused on the analyzing admixed Latin American populations. Network analysis of our LAGC membership shows a high research productivity, but efforts should be made to integrate researchers from South America in global and equitable collaborations. Ongoing LAGC research efforts include a meta-genome-wide association study of alcohol consumption in 55,518 individuals of Latin American origin from US- and Latin America-based cohorts. We identified 4 independent loci associated with alcohol consumption mapping to the *ADH1B* and *ADH1C* genes. A transcriptome-wide association analysis also identified six significantly associated genes, including well-known associations such as *ADH1B*, *ADH1C*, and *ADH7*, and three potentially novel top hits mapping to *IQCF1*, *VEGFC*, and *FRG2B*. Our findings revealed both well-known and potentially novel genetic associations for alcohol consumption in Latin American populations. Our consortium initiative not only empowers Latin American researchers to conduct genomic analyses, addressing the enormous underrepresentation in the psychiatric genomics workforce but also diversifies GWAS to ensure that the benefits of GWAS findings are shared beyond European populations.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1331 † Activating B cells Facilitates the Discovery of Latent Disease-associated Variants

Authors:

D. Murphy, Y. Fu, J. Kelly, R. Pelikan, G. Wiley, P. Gaffney; Oklahoma Med. Res. Fndn., Oklahoma City, OK

Although millions of disease-associated variants have been discovered using GWAS, most of these occur in non-coding regions and their mechanistic link to disease is unclear. Further, separating causal from neutral variants within haplotypes remains challenging. Analyses of expression quantitative trait loci (eQTLs) in healthy tissues only explain some GWAS hits due to the different mechanisms driving variant discovery in these two approaches. Epigenomic phenotypes, like chromatin accessibility, paired with information about long range chromatin contacts encoded in the 3D genome, may reveal regulatory mechanisms not detected by traditional eQTL analysis. Moreover, latent disease-specific mechanisms encoded in the epigenome may be elicited by stimulating healthy tissues. To this end, we gathered expression data using RNA-seq, epigenomic data using ATAC-seq and CUTNRUN (for H3K27-acetylation and CTCF-binding), and 3D-genome contact data using Hi-C, from B cells in 81 genotyped, healthy individuals for both a resting and immune-stimulated condition. We identified strong effects of stimulation across these “omic” layers, including thousands of differentially expressed genes and tens of thousands of differentially activated epigenomic modifications triggered by stimulation. Pairing this omics data with polymorphism data from our sample, we identified over 300,000 candidate QTLs, with over 1/3 of these exclusive to the stimulated state. We used the Activity-By-Contact (ABC) method to connect epigenomic modifications to gene expression via the 3D-genome HiC contact data. Using this approach, we identified tens of thousands of candidate enhancer-gene (E-G) pairs, including nearly 14,000 E-G pairs revealed exclusively in the stimulated state. Taking advantage of variation in our sample, we were able to validate ABC predictions at 654 E-G pairs where levels of enhancer activity (E) and gene expression (G) had a significant correlation. Finally, we examined 2,320 autoimmune-related GWAS variants and found that 84 of these variants co-localized to QTLs, with an additional 561 variants found to lie within candidate enhancer elements identified by ABC (connecting non-coding disease-associated variants to putative effector genes). Collectively, these results demonstrate the utility of a multi-omic, stimulation-augmented approach for annotating non-coding GWAS variants and identifying regulatory epigenomic modifications.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1332 Acute effects of glucocorticoid treatment on gene expression during oral glucose tolerance testglucose tolerance test

Authors:

J. Gamboa, A. Oeser, J. F. Solus, M. Ormseth, Q. Sheng, S-C. Chen, F. Ye, M. Stein, V. Kawai; Vanderbilt Univ. Med. Ctr., Nashville, TN

Background Glucocorticoids (GCs) are potent anti-inflammatory and immunosuppressant drugs commonly prescribed to treat a wide range of illnesses. GCs are associated with multiple metabolic side effects including GC-induced hyperglycemia, which is more evident during non-fasting states. Nevertheless, the underlying mechanisms of this side effect are poorly understood. We hypothesize that acute administration of prednisone affects multiple gene pathways associated with glucose metabolism during an oral glucose tolerance test (OGTT) **Methods**: Eighteen healthy volunteers were enrolled in a clinical study involving two study visits for OGTTs. Participants fasted overnight before each study visit. On the first visit (off-GC), a finger-stick sample was collected at 7:00 am and a standard OGTT was performed at 11:00 am. On the second study visit, a single oral dose of 60 mg prednisone was administered immediately after collecting the finger-stick sample (7:00 am) and the OGTT was performed at 11:00 am. The area under the curve (AUC, median [IQR]) for blood glucose concentrations during the OGTT was calculated. Total RNA was extracted from samples obtained at the beginning (t0) and 120 minutes (t120) after the initiation of the OGTT with (on-GC) and without (off-GC) prednisone treatment. RNA was used to construct libraries for sequencing using Illumina NovaSeq 6000 system. Differentially expressed genes (DEGs) were identified with a false discovery rate (FDR) <0.05 and fold change > 1.5. Gene Set Enrichment Analysis (GSEA) was performed using WebGestaltR v0.4.4 to identify overrepresented pathways. **Results** Acute administration of prednisone increased the AUC for glucose from 13704 [12741, 1500] to 21525 [16983, 23937] mg/dl x min; p<0.001). The median peak glucose was 205 [163, 218] and 131 [109, 140] mg/dl, on-GC and off-GC, respectively. We identified 2889 DEGs at t0 and 2841 DEGs at t120 of the OGTTT between on-GC and off-GC. GSEA showed that the DEGs were enriched for genes involved in **insulin response** [58 DEGs (with enrichment ratio (ER) of 1.7, FDR< $1x10^{-3}$) at t0, and 51 DEGs (ER 1.5, FDR=0.01) at t120]; in **insulin receptor signaling** [32 DEGs (ER 1.8, FDR=0.007) at t0, and 29 DEGs (ER 1.6, FDR=0.03) at t120]; and in **cellular response to insulin stimulus** [44 DEGs (ER 1.6, FDR=0.008) at t0, and 42 DEGs (ER 1.5, FDR=0.02) at t120]. DEGs common to these pathways are *PID1*, *SOCS1*, *ENHO*, and *IRS2*, among others. **Conclusion:** Acute administration of prednisone induced glucose intolerance and changes in the expression of genes involved in insulin and glucose metabolism. This gene signature profile may be useful to understand the underlying mechanisms of GC-induced hyperglycemia.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1333 AD-by-proxy Polygenic Risk Scores: A New Predictor for Alzheimer's Disease and Age of Onset

Authors:

X. Gao[1], M. Chiariglione[1], K. Nuytemans[2], Y-J. Li[3], E. Martin[4]; [1]The Ohio State Univ., Columbus, OH, [2]Univ. of Miami, Miami, FL, [3]Duke Univ., Durham, NC, [4]Univ. of Miami Miller Sch. of Med., Miami, FL

Alzheimer's disease (AD) is the most common late-onset neurodegenerative disease, affecting approximately six million individuals in the United States. Currently, there is no cure for AD, making early detection crucial. Polygenic risk scores (PRSs) have been constructed for AD and the age at onset (AAO) of AD, using genome-wide association study (GWAS) summary statistics derived from clinically diagnosed AD. In this study, we tested whether PRSs constructed using AD-by-proxy, based on parental AD status, are significantly associated with AD and its AAO. Using 368,865 participants of white British ancestry from the UK Biobank dataset and the REGENIE software, we carried out a GWAS of AD-by-proxy (57,976 proxy AD cases and 310,889 non-AD proxy controls). Subsequently, we constructed an AD-by-proxy PRS for the Alzheimer Disease Genetics Consortium (ADGC) participants (9,205 AD cases and 10,334 non-AD controls) and tested their associations with AD and its AAO. To mitigate the known effects of the *APOE* gene, we excluded *APOE* markers (chr19: 45409053 - 45412650) from our PRS calculation. Logistic regression and case-only linear mixed-effects analyses were utilized to examine the associations of PRS with AD and AAO, respectively. The PRS showed a significant association with AD (odds ratio [95% CI] = 1.99 [1.92, 2.06], $P = 6.59 \times 10^{-310}$, after adjusting for age, sex, principal components [PCs] of genetic ancestry) and its AAO (beta [95% CI] = -1.52 [-1.66, -1.38], $P = 6.51 \times 10^{-101}$, after adjusting for sex, PCs, and cohorts), achieving a performance comparable to PRSs constructed using known genetic markers for AD. These associations remained significant for AD ($P = 6.67 \times 10^{-231}$) and its AAO ($P = 7.38 \times 10^{-82}$) after further adjusting for *APOE*-e4 dosage. AD-by-proxy PRSs could potentially be used to aid in predicting AD and its AAO.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1334 Admixture mapping reveals Amerindian-specific Alzheimer disease risk loci in Puerto Ricans.

Authors:

B. Akgun[1], B. E. Feliciano-Astacio[2], K. L. Hamilton-Nelson[1], J. Rivero[1], K. Celis[1], L. D. Adams[1], V. C. Rodriguez[1], P. R. Mena[1], S. Tejada[1], P. G. Whitehead[1], P. L. Bussies[1], M. Prough[1], H. Acosta[3], A. J. Griswold[1,4], C. L. Dalgard[5], K. F. McInerney[6], J. M. Vance[1,4,6], M. L. Cuccaro[1,4], G. W. Beecham[1,4], M. A. Pericak-Vance[1,4,6], F. Rajabli[1,4]; [1]John P. Hussman Inst. for Human Genetics, Univ. of Miami Miller Sch. of Med., Miami, FL, [2]Univ. Central del Caribe, Bayamón, PR, [3]Clinica de la Memoria, San Juan, PR, [4]Dr. John T. Macdonald Fndn. Dept. of Human Genetics, Univ. of Miami Miller Sch. of Med., Miami, FL, [5]Dept. of Anatomy, Physiology, and Genetics, Uniformed Services Univ. of the Hlth.Sci., Bethesda, MD, [6]Dept. of Neurology, Univ. of Miami Miller Sch. of Med., Miami, FL

Background: Puerto Ricans (PR), the second largest Hispanic group in the mainland US, are underrepresented in Alzheimer disease (AD) genetic studies. Their admixed genetic ancestry (Southern European [SEu], African [AF], and Amerindian [AI]) provides a unique opportunity to identify ancestry specific genetic factors associated with AD. Admixture mapping (AM) in comparison to SNP-based genome-wide association studies (GWAS), offers a more powerful approach in such admixed populations, partly due to a lower multiple testing burden. In this study, we employed AM to identify genetic loci associated with AD in PR individuals.
Methods: Our study included 639 individuals (334 cases; 305 cognitively unimpaired [CU]) ascertained through the Puerto Rico Alzheimer Disease Initiative (PRADI). We estimated global ancestry (GA) using the EIGENSTRAT software. To infer local ancestry (LA), we merged the target PR dataset with appropriate reference-population samples from HGDP reference panel. LA was estimated using SHAPEIT, followed by RFMix. Subsequently, we conducted univariate AM (SEu, AF, and AI separately) and joint AM [SEu and AF joint] analysis using the GENESIS software. Finally, we followed up the significant regions using Whole Genome Sequencing (WGS) data with a regression model, including genotype as main effects along with GA, sex, and age as covariates.
Results: Univariate AI AM analysis identified two genome-wide significant ($pv<1.2x10^{-4}$, Kizil et al.) ancestral blocks located on chromosomes 1q41 ($pv=4.4x10^{-5}$) and 3p25 ($pv=3.1x10^{-5}$). Joint AM [SEu and AF] analysis replicated genome-wide significant signal on chromosome 3p25.2. We observed a significant increase of AI ancestry relative to expected local-ancestry proportions at the 1q41 locus, while AI ancestry decreased at the 3p25 locus. The follow-up regional SNP-based regression analysis revealed strong associations on chromosome 1q41 (rs4240935; $pv=1.7x10^{-6}$).
Conclusions: In this study, we identified novel significant loci on chromosomes 1q41 and 3p25 associated with AD in PRs. Our follow-up SNP-based regression analysis further corroborated the findings for the 1p41 locus. Results underscore the intriguing contribution of AI ancestry at these identified novel risk loci. These findings emphasize the importance of including diverse genetic backgrounds in AD research and demonstrate the potential of AM in identifying ancestry-specific genetic risk loci.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1335 Advancing genomic research in the Korean population through comprehensive genomic scans using the Korea Biobank Array v2.0

Authors:

B-J. Kim, M. Hwang, I. Jeong, K. Yoon, H-Y. Park, Y. Kim; Korea NIH, Cheongju-si, Korea, Republic of

Over the past decade, there has been a rapid accumulation of genetic factors linked to various complex diseases through the analysis of hundreds of thousands of samples. To conduct comprehensive genomic scans on such large sample sizes, single nucleotide polymorphism (SNP) microarray and genotype imputation analysis have become widely utilized. However, the microarray approach often lacks rare functional or clinical variants, which limits precision medicine research that relies on genomic data obtained from microarray analysis. To address this limitation, the National Institute of Health, Republic of Korea, launched the Korea Biobank Array (KBA) project. This project involved genotyping 200K Korean samples using the KBA v1.1 array, which included around 830K variants specifically optimized for Korean genome research. In the subsequent phase of the KBA project, the KBA v2.0 was employed, incorporating 1.68 million markers. This updated version aimed to provide comprehensive Korean genome information, including approximately 132K clinically actionable variants and 489K functional variants extracted from publicly available clinical variant databases and approximately 28,000 whole genome sequences of East Asians and Koreans. During the pilot phase, approximately 10,000 samples underwent genotyping using the KBA v2.0. Following stringent quality control measures, the KBA v2.0 dataset yielded around 1.4 million non-monomorphic variants. Among these variants, about 424K were categorized as functional or clinical variants, including those classified as pathogenic or likely pathogenic variants documented in databases such as ClinVar or LOVD. Among the 143K clinically related variants, 0.5% were found to be common, (MAF >= 1%), while 0.9% were within MAF 0.1-1% and the remaining 98.6% were MAF < 1% in the Korean population. Additionally, the KBA v2.0 provided extensive genomic coverage, with 99% for variants with MAF >= 5%, and 86% for variants with MAF 1-5%. In a comprehensive genome-wide association study focusing on lipid traits, previously known loci were reaffirmed, and additionally, four newly associated rare genetic variants with strong genetic effects specific to East Asians were identified. These new variants were located at the ALOX5, SLFN14, APOB, and CLPN3 regions, providing further insights into the genetic underpinnings of lipid-related traits in this population. In summary, this study underscored the significance of conducting a comprehensive genomic scan using microarray technology, highlighting its early results and its potential as a valuable resource for precision medicine research within the Korean population.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1336 African ancestry individuals with higher educational attainment are resilient to Alzheimer disease pathology specific blood biomarker pTau181.

Authors:

F. Rajabli[1], B. Akgun[1], L. Adams[1], J. Inciute[1], T. D. Starks[2], R. Laux[3], M. Cuccaro[1], J. M. Vance[1], J. L. Haines[3], G. S. Byrd[2], A. Seixas[4], M. Pericak-Vance[1]; [1]John P. Hussman Inst. for Human Genomics, Univ. of Miami Miller Sch. of Med., Miami, FL, [2]Maya Angelou Ctr. for Hlth.Equity, Wake Forest Univ., Winston-Salem, NC, [3]Dept. of Population and Quantitative Hlth.Sci., Case Western Reserve Univ., Cleveland, OH, [4](2) Dept. of Psychiatry and Behavioral Sci., Univ. of Miami Miller Sch. of Med., Miami, FL

Background Alzheimer disease (AD) risk is influenced by a wide range of factors, including genetic, psychosocial, environmental, and behavioral elements. *APOE ε4* is the strongest genetic risk factor for late-onset AD. Previous studies suggested that individuals with higher levels of education tend to demonstrate better cognitive capabilities compared to those with less education, even when having comparable degrees of advanced AD-related pathology (ADP). However, it is yet to be determined if educational attainment (EA) has a modifiable effect on the functional decline in individuals with the ε4 risk allele and ADP. Blood pTau181 is a biomarker specific to ADP. Using African American (AA) individuals with elevated blood pTau181 levels, we investigated whether EA promotes functional resilience differently between *APOE ε4* allele carriers and non-carriers. **Methods** We studied 410 AAs with known years of completed education, blood pTau181 levels, and *APOE* genotypes. We constructed a composite score, CDR-FUNC, from the four non-memory components of the Clinical Dementia Rating (CDR) to measure functional difficulties, with a score of 0 indicating no impairment and 12 indicating severe impairment. We divided EA into two categories: low (≤ 8 years), and high (>8 years). pTau181 levels were log-transformed, and individuals with advanced ADP-related pTau181 levels were defined as those whose log10(pTau181) level was greater than one standard deviation above the mean. Further, individuals were categorized based on their *APOE* ε4 carrier status. We utilized the non-parametric Mann-Whitney U test to evaluate the association between EA and CDR-FUNC in individuals with advanced ADP-related pTau181 levels and *APOE* e4 allele. **Results** The results showed that EA promotes resilience to functional problems in AA individuals with advanced blood pTau181 levels, such that individuals with high EA are more likely to have better functional ability compared to those with lower EA (p=0.0007). Additionally, we found that the effect of high EA on functional resilience was stronger in ε4 non-carriers compared to ε4 carriers (p=0.022). **Conclusion** This study found that years of education are associated with better functional abilities in individuals exhibiting high levels of ADP-related blood pTau181 biomarkers. This effect was observed in both *APOE* ε4 carriers and non-carriers, but the protective effect of high EA was stronger in non-carriers. It can be inferred that higher education can help individuals maintain their functional abilities even with AD pathology.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1337 Age and parasitemia explain most of the variation in host and parasite gene expression among Malian children infected with *P. falciparum*

Authors:

K. Tebben; Univ. of Maryland, Baltimore, Baltimore, MD

Plasmodium falciparum caused over 600,000 deaths in 2021, primarily in young children. In Bandiagara, Mali, children experience 2 malaria episodes per season, on average. However, even in the same transmission area, the parasitemia and number of uncomplicated, symptomatic infections vary widely between children. The host and parasite factors contributing to this variation remain incompletely understood. To begin to understand this variation, we characterized the host and parasite transcriptomes from 136 blood samples from children symptomatic for falciparum malaria, studying the expression of 2,484 *Plasmodium* and 9,205 human genes. We used gene expression deconvolution to estimate the proportion of parasite stages and immune cells in each sample to investigate differences in cell proportions between children and correct our differential gene expression analyses for these differences. Parasitemia and host age explained most of the variation in both host and parasite gene expression, while few genes were associated with the number of symptomatic infections in the study period, complexity of infection or the sex of the child. Gene expression differences associated with parasitemia were driven by differences in cell composition. Higher parasitemia infections had more ring-stage parasites and neutrophils and fewer trophozoites and T-cells, suggesting parasitemia-dependent parasite sequestration or regulation of the *Plasmodium* life cycle and/or parasitemia-dependent T-cell suppression. Similarly, parasite genes associated with the child's age resulted from differences in stage composition, with older children having proportionally more male gametocytes. By contrast, the host gene expression associated with age was not completely explained by differences in immune cell composition. In particular, many genes involved in innate response (TLR and NLR signaling) were more expressed in younger children while genes involved in adaptive immunity (TCR and BCR signaling) were higher in older children, supporting that the immune response to malaria changes over time. These analyses broaden our understanding of the variations in pathogenesis of *P. falciparum* even within similar clinical syndromes and can provide insight for targeted prevention and treatment strategies based on age and/or parasitemia of infected children.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1338 Age of asthma onset and risk of cardiovascular disease: A Mendelian randomization study

Authors:

H-Y. Chung, N. Y. Ng, J. Zhao, C. Mak, S. Lee; The Univ. of Hong Kong, Hong Kong, China

Background Asthma is a heterogeneous, non-communicable disease, with clinical presentations varying with age-of-onset. Longitudinal observational studies have shown consistent associations between asthma as a collective disease and cardiovascular diseases (CVDs). However, observational associations of childhood-onset asthma with CVDs are inconsistent, and causality has not been confirmed. We conducted a Mendelian randomization (MR) study to elucidate the different causal impacts of childhood and adulthood-onset asthma on four CVDs (coronary artery disease, heart failure, atrial fibrillation, and stroke). **Methods** Genetic instruments (instrumental variables) were identified from a published genome-wide association study (GWAS) on childhood-onset asthma (cases: 13,962; controls: 300,671) and adulthood-onset asthma (cases: 26,582; controls: 300,671) using a p-value threshold of $P<3x10^{-8}$. MR analyses were conducted using the largest available GWAS obtained from genome-wide association meta-analysis studies or large consortia for each CVD outcome. Odds ratios (ORs) were calculated using the inverse variance weighted method. Sensitivity analyses (MR-Egger, MR-PRESSO and weighted median methods) were conducted to evaluate the effects of horizontal pleiotropy to imply causality. **Findings** Under the weighted median method, we found a 3.3% increased risk of heart failure per log-odds increase in the risk of childhood-onset asthma ($p=0.032$). Under the MR-Egger method, we found a 5.7% increased risk of atrial fibrillation per log-odds increase in the risk of childhood-onset asthma ($p=0.039$). Genetically predicted adulthood-onset asthma was not causal for any CVD outcomes across all analyses. **Conclusions** Our results reveal potentially different causal effects of asthma subtypes on the risk of CVDs. Clinical interventions directly targeting childhood-onset asthma and its downstream effects may prove effective in reducing the risk of heart failure and atrial fibrillation. These findings should be triangulated with clinical evidence to support the distinct management of asthma subtypes in reducing CVD risk.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1339 Alzheimer's disease APOE4 variant associated with metabolites while Alzheimer's disease polygenic risk score lacks association with metabolites.

Authors:

E. Drzymalla[1], L. Raffield[2], C. Avery[1]; [1]Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, [2]UNC - Chapel Hill, CHAPEL HILL, NC

Mechanisms underlying Alzheimer's disease (AD) are largely unclear, in part reflecting challenges in phenotype classification and biological pathway identification as well as a late disease onset. Multi-omics studies that integrate metabolomics and AD polygenic risk scores (PRS) can help address these challenges by measuring diverse biological mechanisms (via metabolomics) and identifying individuals at elevated AD risk (via AD PRS) decades before disease onset. AD PRS was constructed using external European-derived weights for 83 variants, excluding the *APOE4* locus which would have a disproportionately high weight given its strong known effects on AD risk. This PRS has been validated previously in external cohorts. To validate AD PRS performance (continuous or dichotomized at the 90th %ile) in UK Biobank, we used Cox proportional hazard models with time to AD diagnosis, defined by ICD-10 G30 or F00 or ICD-9 331.0, as the outcome, AD PRS as the exposure, and adjusted for age, sex, education, first ten principal components (PCs) and self-reported ethnicity. To identify mechanisms by which increased AD genetic risk (continuous AD PRS, dichotomized at the 90th %ile AD PRS, or rs429358 *APOE4* AD risk variant alone) may increase AD incidence, we used linear regression and regressed each of 161 metabolites outcomes (measured using the Nightingale Health NMR platform) on the AD PRS exposure, adjusting for age, sex, education, first ten PCs and self-reported ethnicity, and applying a 5% FDR correction. There were 88,412 UK Biobank participants (mean age = 57, 55% female, 671 AD cases) with complete outcome, follow-up time, AD PRS, covariate, and metabolite data. The AD PRS was as expected associated with AD in UK Biobank (continuous PRS: per SD HR = 1.46, 95% CI = 1.36-1.57, p-value = <2x10-16, binary PRS: HR = 2.05, 95% CI = 1.69-2.49, p-value = 3.99x10-13). No metabolites were significant in the continuous or binary PRS model. Sensitivity analysis restricted to European ancestry had consistent results. However, 134 (83%) metabolites such as cholesterol and phospholipids were associated with *APOE4*. *APOE4* is a strong predictor of AD and has been previously associated with lipid metabolites in plasma including phospholipids and cholesterol, which we confirm here. The lack of association between the AD PRS and metabolites suggests the *APOE4* locus is the main driver between AD genetic risk and the metabolome. However, we analyzed a targeted metabolite platform, and the AD PRS explains only a modest portion of AD risk. We plan to expand this analysis to proteins in the UK Biobank as well as broader metabolite panels in other cohort studies.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1340 Alzheimer's disease is associated with reduced survival across generations in *APOE4* carriers and non-carriers.

Authors:

R. Holmes, H. Duan, K. Arbeev, D. Wu, A. Yashin, S. Ukraintseva; Duke Univ., Durham, NC

Introduction: A strong genetic risk factor for Alzheimer's Disease (AD), *APOE4*, has also been associated with reduced longevity. However, trans-generational impact of predisposition to AD on longevity, including in *APOE4* non-carriers, is less explored. We compared lifespans of parents of UK Biobank (UKB) participants with and without AD diagnosis, considering *APOE4* carrier status. **Method:** We estimated differences in frequencies of parental survival to ages 80+ and 90+ for UKB participants with and without AD. To ensure that parental data are not affected by participants' current AD status, all participants who had AD diagnosis upon entering the study were removed from analysis. We used two binary phenotypes of parental survival: 1) survived age 80+ vs. died before 80; and 2) survived age 90+ vs. died between 80 and 90. SAS procedure PROC FREQ was used. Chi-square test was used to compare significance between groups. **Results:** The proportion of parents who survived age 80+ was significantly (FDR<0.05) lower for participants with AD than without AD, in the total sample (mother's survival: p=1.43E-06, AD: 50.85% vs. NoAD: 55.11%; father's survival: p=1.83E-13, AD: 29.47% vs. NoAD: 35.79%). The results were similar for males, females, and Whites, and non-significant for Blacks. The difference in parental survival to age 80+ was similar for *APOE4* carriers and non-carriers with and without AD (*APOE4* carriers: mother's survival: p=0.026, AD: 51.13% vs. NoAD: 54.32%; father's survival: p=2.85E-06, AD: 28.61% vs. NoAD: 35.06%; *APOE4* non-carriers: mother's survival: p=0.004, AD: 51.37% vs. NoAD: 56.07%; father's survival: p=4.67E-05, AD: 30.19% vs. NoAD: 36.62%), though the proportion of parents surviving age 80+ was overall lower in *APOE4* carriers. Mother's survival to age 90+ was significantly (FDR<0.05) associated with AD diagnosis in the total sample (p=4.52E-08, AD: 26.68% vs. NoAD: 33.55%), with similar results for males, females, and Whites. The difference between participants with and without AD in survival of their mothers was larger for survival to age 90+ than to age 80+, especially for *APOE4* carriers (p=1.82E-05, AD: 22.70% vs. NoAD: 31.12%). **Conclusion:** Parents of UKB participants with AD had lower chances to survive ages 80+ and 90+ than parents of participants without AD. This was true for participants who were *APOE4* carriers, as well as non-carriers. Our results suggest that predisposition to AD may negatively affect survival and longevity across generations and that risk factors beyond *APOE4* may play role in this. They also suggest a larger detrimental impact of predisposition to AD on chances of longevity 90+ than on conventional survival (80+).

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1341 Alzheimer's disease genetic risk and Cognitive "SuperAging" in the Midwestern Amish

Authors:

D. Dorfsman[1,2], M. B. Prough[1], L. J. Caywood[1], J. E. Clouse[1], S. D. Herington[1], S. H. Slifer[1], L. D. Adams[1], R. A. Laux[3,4], Y. E. Song[3,4], A. Lynn[3,4], S. L. Fuzzell[3,4], S. D. Hochstetler[3,4], K. Miskimen[3,4], L. R. Main[4,5], P. Wang[3], Y. Liu[4], P. Ogrocki[6], A. J. Lerner[5], J. M. Vance[1,2], M. L. Cuccaro[1,2], J. L. Haines[3,4], M. A. Pericak-Vance[1,2], W. K. Scott[1,2]; [1]John P. Hussman Inst. for Human Genomics, Univ. of Miami Miller Sch. of Med., Miami, FL, [2]Dr. John T. Macdonald Fndn. of Human Genetics, Univ. of Miami Miller Sch. of Med., Miami, FL, [3]Dept. of Population and Quantitative Hlth.Sci., Case Western Reserve Univ., Cleveland, OH, [4]Cleveland Inst. for Computational Biology, Case Western Reserve Univ., Cleveland, OH, [5]Case Western Reserve Univ. Sch. of Med., Cleveland, OH, [6]Univ. Hosp. Cleveland Med. Ctr., Cleveland, OH

Alzheimer's disease (AD) is a highly heritable neurodegenerative disorder that accounts for 60-80% of all dementia cases. While cognitive performance often declines with increasing age, some individuals appear to resist this trajectory by maintaining youthful brain capacity well into their later years. One such group, described as SuperAgers (SA), are ≥ 80 years old with episodic memory function at or above the level expected for a middle-aged adult. Studies of SA may provide key insights into the heritable factors that predispose to, as well as those that protect against, AD and cognitive impairment (CI). The purpose of this study is to use an Amish-derived sample to compare SA to those who are cognitively unimpaired (CU) aged ≥ 80 but do not meet the criteria for SA, to assess whether differences between these groups are explained by known genetic risk factors for AD. Enrolled subjects (N=428) are adult members of the Midwestern Amish population, an endogamous group with a shared lifestyle. Subjects were evaluated and adjudicated using clinical information that included a comprehensive neuropsychological test battery that measures performance across multiple cognitive domains. Using normative data, subjects aged ≥ 80 with episodic memory (CERAD Word List Recall) at or above the mean for ages 35-44, and non-memory cognitive test performance (Trail Making Test B, Category Fluency, MINT) within one standard deviation or better for the subject's age, were categorized as SA (n=85). Using called whole-genome sequence variants, an AD genetic risk score (GRS) was calculated for each subject by utilizing effect sizes for 83 genome-wide significant AD-associated variants, excluding *APOE*, published in Bellenguez et al. (2022). The effect of known AD genetic risk factors on CI was confirmed through multivariable logistic regression analysis, which included *APOE* E4 allele count ($p<0.001$), GRS ($p<0.001$), age ($p<0.001$), and sex ($p=0.3$) as predictors of disease status (case: AD, control: SA and CU non-SA). The effect of these predictors on SA was then modeled using an updated case status (case: SA, control: CU non-SA aged ≥ 80). In this scenario, *APOE* E4 allele count ($p=0.7$), GRS ($p=0.6$), age ($p=0.09$), and sex ($p=0.2$) did not exhibit significant differences between the two groups. Moreover, *APOE* E2 allele count, a known protective factor against AD, was not significantly different between the two groups ($p=0.6$). These results suggest that differences between SA and unimpaired non-SA individuals are not due to known AD genetic risk factors in the Amish population. Ongoing studies aim to identify genetic factors that promote SA and enhanced cognitive function at older ages.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1343 † An Alzheimer's Disease risk variant in *TTC3* modifies the actin cytoskeleton organization and PI3K-Akt signaling in iPSC-derived forebrain neurons.

Authors:

H. Cukier[1], C. L. Duarte[1], J. Laverde-Paz[1], S. A. Simon[1], D. J. Van Booven[1], A. T. Miyares[1], P. L. Whitehead[1], K. L. Hamilton-Nelson[1], L. D. Adams[1], R. M. Carney[2], M. L. Cuccaro[1], J. M. Vance[1], M. A. Pericak-Vance[1], A. J. Griswold[1], D. M. Dykxhoorn[1]; [1]Univ. of Miami Miller Sch. of Med., Miami, FL, [2]Bruce W. Carter VA Med. Ctr., Miami, FL

We identified a rare, nonsynonymous variant in the *tetratricopeptide repeat domain 3* (*TTC3*) gene that segregated in a multigenerational non-Hispanic white late onset Alzheimer disease (LOAD) family (Kohli, et al, 2016). This missense alteration, rs377155188 (p.S1038C), is predicted to be deleterious and is extremely rare. Studies have reported that cortical *TTC3* expression is reduced in LOAD patients and negatively correlated with AD neuropathology. To understand the mechanism by which the TTC3 p.S1038C may contribute to LOAD risk, CRISPR/Cas9 genome edited induced pluripotent stem cells (iPSCs) were developed that were homozygous for the variant to examine cellular and transcriptional consequences in neuronal cells (Laverde-Paz, et al, 2021). Quantitative PCR and western blot analysis demonstrated that *TTC3* levels were decreased in edited compared to unedited iPSCs, as well as differentiated neurons. Transcriptome analysis of day 70 neurons identified 979 genes that were differentially expressed (FDR<0.05). This included known AD genes (*BACE1*) and genes in AD GWAS loci (*ADAMTS1*, *MAF*, *NCK2*). Furthermore, there was an enrichment for genes involved in axon guidance, regulation of actin cytoskeleton, and GABAergic synapse. In growing neuronal precursor cells (NPCs), cells with the *TTC3* variant recovered more quickly from a scratch wound. Since there is evidence that modulation of *TTC3* affects neurite growth, morphological measures of axon formation were assessed using the Incucyte Zoom. Studies demonstrate an increase in neurite outgrowth, which phenotypically corresponds with previous studies of a decrease in *TTC3* function. This phenotype was tempered by treatment with Cytochalasin D, an inhibitor of actin polymerization. Additionally, TTC3 ubiquitinates phosphorylated AKT and regulates AKT signaling. The edited cells were found to have an increase in phosphorylated AKT (pAKT) relative to total AKT. Combined, these results suggest that the *TTC3* p.S1038C variant causes a loss of function. Utilizing a CRISPR genome edited iPSC carrying a homozygous alteration in *TTC3*, we were able to identify potential mechanisms by which *TTC3* may contribute to LOAD risk.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1344 An analytical framework to identify novel disease genes including common and rare variants

Authors:

C. Salvoro[1], Q. Wang[2], S. Das[1], K. Megy[1], M. Fabre[1,3], S. Petrovski[1,4,5], K. Carss[1]; [1]Ctr. for Genomics Res., Discovery Sci., R&D, AstraZeneca, Cambridge, United Kingdom, [2]Ctr. for Genomics Res., Discovery Sci., R&D, AstraZeneca, Gaithersburg, MD, [3]Dept. of Haematology, Cambridge Univ. Hosp. NHS Trust, Cambridge, United Kingdom, [4]Dept. of Med., Univ. of Melbourne, Austin Hlth., Melbourne, Australia, [5]Epilepsy Res. Ctr., Univ. of Melbourne, Austin Hlth., Melbourne, Australia

Complex diseases arise with the contribution of both common and rare inherited genetic variants. Traditionally, these two types of variants have been investigated separately, however, recent studies have demonstrated that their combination and interaction is critical. Polygenic Risk Scores (PRS), which capture the effect of common variants on a particular phenotype, have been shown to modify the penetrance of rare variants, and subjects who deviate from their phenotypic prediction based on common variants are enriched in potentially causal rare alleles. Consistent with this, including PRS in disease regression models improves rare variant identification. Here we hypothesise that incorporating previous knowledge of both common and rare variants in disease models could increase power to reveal novel implicated genes, and improve our understanding of the contribution of these two classes of variant to complex diseases. Using 470,048 exome sequences from UK Biobank (UKB), we built two frameworks, to investigate binary disease outcome and time to disease occurrence, both including a rare-variant-carrier status and a PRS percentile for each subject. As a proof of concept, we considered 28 diseases for which PRS were released by UKB. To determine whether a subject was a rare variant carrier for a given disease, we used pathogenic or likely pathogenic annotations in ClinVar. We modelled binary disease outcome using Firth's logistic regression, and regressing outcome on gene-level qualifying rare variant carrier status. Rare-pathogenic variant-carrier status and PRS percentile were included as covariates, together with sex, age, sex*age, and the first 4 ancestry principal components. Analogously, we modelled time to disease with the same gene-level genotypes and covariates, using Cox's proportional hazards regression with Firth's penalized likelihood. Time to disease was measured in years considering UKB recruitment as baseline. Preliminary findings identified 6 significant associations ($p<1x10^{-8}$), including positive controls such as *CFH* in age-related macular degeneration ($p=4.0x10^{-11}$, OR=2.1) and *BRCA1* in epithelial ovarian cancer ($p=3.2x10^{-14}$, OR=7.2). PRS were consistently associated with disease outcome in all models, while rare-pathogenic variant-carrier status showed variable contributions, indicating still incomplete knowledge for rare variants. We believe this exploratory study could set the basis for a new approach to identify disease genes with high clinical relevance. With the increasing availability of PRS and rare pathogenic variant annotations, our frameworks could be applied to a growing number of different diseases.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1345 An EBV-associated atypical B cell signature in clinically isolated syndromeis implicated in progression of multiple sclerosis

Authors:

S. Gregory[1], E. SoRelle[2], H. Ellora[1], J. Vaibhav[1], J. Giarraputo[3], K. Abramson[1], E. Hocke[1], L. Cooney[4], K. Harris[5], S. Zamvil[6], M. Luftig[2]; [1]Duke Molecular Physiology Inst., Durham, NC, [2]Duke Dept. of Molecular Genetics and Microbiol., Durham, NC, [3]Yale Sch. of Med., New Haven, CT, [4]Immune Tolerance Network, Bethesda, MD, [5]Immune Tolerence Network, Bethesda, MD, [6]Dept. of Neurology, San Francisco, CA

Introduction: Expansion and pathogenicity of CD19+/CD20+/CD11c+/T-bet+ atypical B cells (ABCs) are hallmarks of numerous autoimmune disorders and chronic infections. In many such cases Epstein-Barr virus (EBV) is an associated or etiologic factor, though EBV involvement in these diseases remains poorly understood. Notably, expansion of pro-inflammatory ABCs and a putative causal role for EBV each have been identified independently in multiple sclerosis (MS). A common precipitating event in MS onset is Clinically Isolated Syndrome (CIS), a neuroinflammatory demyelinating condition in which 60-80% of cases progress to relapsing-remitting MS (RRMS).
Methods: We used single-cell gene and surface protein expression (scRNA/CITE-seq) in peripheral B cells collected longitudinally from patients with CIS during the Immune Tolerance Network STAyCIS Trial. We focused on the transcriptomic signatures of ABCs from this cohort, publicly available scRNA-seq datasets from five other autoimmune and chronic infectious diseases, healthy donors, and in vitro EBV infection. Ongoing analyses include the incorporation of MS GWAS loci into our differentially expressed scRNA-Seq ABC pathways.
Results: Patients with CIS exhibited significantly higher frequencies of ABCs than healthy adults (5.4% vs 3.0% of peripheral B cells), establishing ABC compartment expansion as a clinical feature. An EBV-associated transcriptome signature in ABCs from CIS, distinct from healthy controls and other disease contexts, was also identified although direct evidence of EBV infection was not obtained. ABCs from patients with CIS exhibited significantly elevated expression of inflammatory cytokines (CXCL8 / IL8, IL18, VEGFA), some of which (IL-8, VEGF) were secreted by B cells upon de novo EBV infection.
Conclusion: Outcome stratification of CIS samples revealed a rare yet distinctive inflammatory ABC subset that was significantly underrepresented in long-term non-progressors (LTNP) versus cases with primary endpoint RRMS activity (~7x difference). Thus, this study provides evidence for ABC dysregulation - possibly arising from responses to EBV infection - preceding MS onset.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1346 Analysis of caudate nucleus transcriptome in individuals with schizophrenia identifies new risk genes and implicates pre-synaptic DRD2 autoreceptor dysregulation as a genetic risk factor for schizophrenia

Authors:

A. Paquola[1], K. J. Benjamin[1], Q. Chen[2], A. Jaffe[3], J. Stolz[3], L. Collado-Torres[1], L. Huuki[3], E. Burke[3], R. Arora[3], A. Feltrin[3], A. Barbosa[3], E. Radulescu[3], G. Pergola[3], J. Shin[4], W. Ulrich[3], A. Deep-Soboslay[3], R. Tao[3], BrainSeq Consortium, T. Hyde[3], J. Kleinman[3], J. Erwin[1], D. Weinberger[3]; [1]Lieber Inst. for Brain Dev. / Johns Hopkins Med., Baltimore, MD, [2]LIBD, Baltimore, MD, [3]Lieber Inst. for Brain Dev., Baltimore, MD, [4]Lieber Inst. for Brain Devel, Baltimore, MD

Most studies of gene expression in the brains of individuals with schizophrenia have focused on cortical regions, but subcortical nuclei such as the striatum are prominently implicated in the disease, and current antipsychotic drugs target the striatum's dense dopaminergic innervation. Here, we performed a comprehensive analysis of the genetic and transcriptional landscape of schizophrenia in the postmortem caudate nucleus of the striatum of 443 individuals (245 neurotypical individuals, 154 individuals with schizophrenia and 44 individuals with bipolar disorder), 210 from African and 233 from European ancestries. Integrating expression quantitative trait loci analysis, Mendelian randomization with the latest schizophrenia genome-wide association study, transcriptome-wide association study and differential expression analysis, we identified many genes associated with schizophrenia risk. We found that reduced expression of the pre-synaptic DRD2 autoreceptor is associated with increased genetic risk for schizophrenia. We also found that antipsychotic medication has an extensive influence on caudate gene expression. Using a deep learning approach based on variational autoencoders, we constructed gene expression networks that highlight interactions involving schizophrenia risk. These analyses provide a resource for the study of schizophrenia and insights into risk mechanisms and potential therapeutic targets.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1347 † Analysis of de novo variants in trios with orofacial clefts identifies GRHL2 as a novel gene

Authors:

C. Yang[1], S. W. Curtis[1], J. C. Murray[2], C. D. Padilla[3], T. H. Beaty[4], L. M. Moreno[5], S. E. Daack-Hirsch[6], E. Feingold[7], S. M. Weinberg[7], M. L. Marazita[8], M. P. Epstein[9], D. J. Cutler[10], H. Brand[11], E. J. Leslie[1]; [1]Emory Univ., Atlanta, GA, [2]U of Iowa, Iowa City, IA, [3]Univ. of the Philippines Manila, Manila, Philippines, [4]Johns Hopkins Univ, Sch PubHlth, Baltimore, MD, [5]Univ Iowa, Iowa City, IA, [6]Univ Iowa Hosps & Clinics, Iowa City, IA, [7]Univ of Pittsburgh, Pittsburgh, PA, [8]Univ Pittsburgh, Ctr. for Craniofacial and Dental Genetics, Pittsburgh, PA, [9]EMORY Univ., Atlanta, GA, [10]Emory Univ. Sch. of Med., Decatur, GA, [11]MGH, Boston, MA

Orofacial clefts (OFCs), including cleft lip and cleft palate, are among the most common birth defects, occurring in approximately 1 in 700 live births worldwide. The prevalence of OFCs varies around the world with many populations at much higher risk than those of European ancestry. However, genetic studies in individuals of non-European ancestry are more limited and require further exploration. OFCs have a strong genetic basis, with genetic variants of all types and frequencies increasing risk for an OFC. Here, we focused our analysis on de novo variants (DNVs) found in whole genome sequence (WGS) data from 419 child-parent trios of Filipino ancestry with OFCs from the Gabriella Miller Kids First (GMKF) Research Program, in order to determine the contributions of coding DNVs to OFC risk. After DNVs were called and rigorously filtered for quality, we identified 438 coding DNVs. There was a significant enrichment of loss-of-function ($p = 0.00708$) and protein-altering DNVs ($p = 0.000143$), with 45 loss-of-function and 313 protein-altering mutations identified. Two loss-of-function DNVs were identified in GRHL2 ($p = 4.74E-06$), a gene intolerant to loss-of-function variants (LOEUF = 0.27, pLI = 1). We then investigated DNVs from three other GMKF cohorts with OFCs of Colombian, Taiwanese, and European ancestries, totaling 1,114 trios. In these samples, we identified a third DNV, a structural variant deleting exons 2-9 of GRHL2. GRHL2 is one of a family of conserved transcription factors important for embryonic development. Mice deficient for Grhl2 or Grhl3 are known to have craniofacial anomalies, including facial and palatal clefts. Although heterozygous GRHL3 variants are known to cause Van der Woude syndrome and isolated cleft palate in humans, GRHL2 was previously only implicated in craniofacial anomalies by virtue of being one of many genes deleted in microdeletions of the 8p22. Truncating GRHL2 mutations are associated with autosomal dominant progressive hearing loss and missense variants are associated with autosomal recessive ectodermal dysplasia. The three loss-of-function DNVs identified here conclusively implicate GRHL2 in orofacial clefting and suggest that these families may also be at risk for developing hearing loss. We also identified a de novo truncating variant in GRHL1 and a de novo missense variant in GRHL3, which suggests that this family of transcription factors is important to craniofacial development in humans. These findings support a need for additional work on the GRHL gene family to determine their association with OFCs and other birth defects, plus their role in comorbidities such as deafness.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1348 Analysis of gut microbiome in a low and high endotoxin house dust mite induced airway hyperresponsiveness mouse model

Authors:

K. Mulholland, N. Fuentes, P. Warrener, C. Caceres, B. Mindt, B. Sellman, C. Tkaczyk, V. Gopalakrishnan; Astrazeneca, Gaithersbug, MD

Asthma affects nearly 339 million people worldwide; the gut microbiome has been growing in importance as a key mediator of asthma. Although therapies exist, there is an unmet need for patients with a neutrophilic phenotype. We characterized the impact of endotoxin abundance in house dust-mite (HDM) on gut microbiome dynamics. C57BL/6J male mice were challenged with low (LE-HDM) or high (HE-HDM) endotoxin abundance or PBS over 21 days. The gut microbiome was characterized by 16S rRNA sequencing on fecal samples collected before challenge, on day 14 and day 21. Inflammatory markers in BAL were characterized by flow cytometry. Firmicutes, Bacteroidota and Actinobacteriota were the most prevalent phyla in all groups. Alpha diversity was higher in LE-HDM, but there were no appreciable differences in microbiome composition at any timepoint. Differential abundance comparisons revealed unique microbial signatures with enrichment of Firmicutes and *Intestinimonas* in LE-HDM and Bifidobacteriaceae and *Bifidobacterium* in HE-HDM. Differential taxa between HE-HDM and LE-HDM significantly associated with levels of Th1 and Th17 or Th2 cytokines in BAL at day 21, respectively.
These findings reveal that differences in specific taxa in gut microbiomes are associated with endotoxin abundances in HDM and are also correlated with inflammatory responses in asthma.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1349 Analysis of host genetics using genome sequencing data of 1,220 SARS-CoV-2 infected individuals

Authors:

A. Schmidt[1], N. Casadei[2], F. Brand[1], G. Demidov[2], E. Vojgani[3], R. Aldisi[1], G. Butler-Laporte[4,5], B. Richards[5,6,7], A. Heimbach[1], K. Becker[3], S. Ossowski[2], J. Schultze[8,9], P. Nürnberg[3], M. M. Nöthen[1], S. Motameny[3], M. Nothnagel[3], O. Riess[2], E. Schulte[1,10,11], **K. U. Ludwig**[1], DeCOI Host Genetics Group; [1]Univ. Hosp. Bonn, Bonn, Germany, [2]Univ. of Tuebingen, Tuebingen, Germany, [3]Univ. of Cologne, Cologne, Germany, [4]Univ. of Oxford, Oxford, United Kingdom, [5]McGill Univ., Montreal, QC, Canada, [6]King's Coll. London, London, United Kingdom, [7]5 Prime Sci. Inc, Montreal, QC, Canada, [8]Deutsches Zentrum für Neurodegenerative Erkrankungen (DZNE), eV, Bonn, Germany, [9]Univ. of Bonn, Bonn, Germany, [10]Technical Univ. Munich/Helmholtz Munich, Munich, Germany, [11]Univ. of Munich, Munich, Germany

SARS-CoV-2 infections can result in variable courses of disease, with substantial contribution by the genetic make-up of the host. While large-scale GWAS have identified common variants associated with different SARS-CoV-2-related traits, analyses of extreme phenotypes have identified rare variants causing monogenic predispositions for severe COVID-19, such as *TLR7*-deficiency. Genome sequencing covers both common and rare genetic variation, but so far has rarely been performed in the context of SARS-CoV-2 infections. We here describe the analysis of genome sequencing data of 1,220 individuals with mild, moderate (hospitalization required) and severe (mechanical ventilation required) COVID-19. We did not observe any individual with *TLR7* deficiency, but identified homozygous variants known to be causative of autosomal recessive disease (genes: *BBS1*, *AGXT*, *SERPIN1C*, *AIRE*) in four young men (mean age: 33 years) with moderate (n=2) or severe (n=2) COVID-19. Gene-set based collapsing analyses provided evidence for nominally significant enrichment of rare predicted loss of function and missense variants in genes of the innate immune system, in individuals with severe COVID-19, with a stronger effect in males. Finally, GWAS on the genome sequencing data confirmed the association of common variants at 12 previously associated loci, and polygenic risk scores were highest in young and severely affected individuals. Overall, this analysis comprehensively analyzes COVID-19 host genetics and generates valuable hypotheses that can be tested by future studies.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1350 Analysis of variants of uncertain significance in familial dyslipidemia in the Puerto Rican population: Implications for risk stratification and early intervention.

Authors:

F. Benítez Ríos[1], M. Gómez Ocasio[1], A. de Jesús Barbosa[2], S. Flórez Marqués[2], A. Ortega Vega[2], S. Carlo[1], J. García Mateo[1], M. Mangual García[3], A. González Bóssolo[3], E. Albino[4]; [1]Ponce Hlth.Sci. Univ., Ponce, PR, [2]Univ. of Puerto Rico, San Juan, PR, [3]San Juan City Hosp., San Juan, PR, [4]Univ. of Puerto Rico - MSC, San Juan, PR

Introduction: Dyslipidemia, characterized by abnormal lipid levels, is a major risk factor for atherosclerotic cardiovascular disease (ASCVD). Familial dyslipidemias, including familial hypercholesterolemia (FH), familial chylomicronemia syndrome (FCS), and familial dysbetaloproteinemia, often exhibit a high prevalence of variants of uncertain significance (VUS) in genetic tests. This study aims to catalog all genetic variants associated with selected familial dyslipidemias in the Puerto Rican population and establish a phenotype/genotype correlation.
Methods: Demographic and clinical data were retrospectively collected from electronic medical records of Puerto Rican patients with a clinical diagnosis of familial dyslipidemia who underwent genetic testing from February 2023 through June 2023. Vital clinical data, including age, sex, body mass index, lipid tests, VUS results, comorbidities, treatment and familial history, were analyzed. Statistical analysis was performed to determine the diagnostic yield.
Results: Genetic testing reports from 24 Puerto Ricans with suspected familial dyslipidemia were analyzed. Among them, 2 (8.33%) tested positive, 4 (16.67%) had uncertain results, and 18 (75%) tested negative. VUS were meticulously cataloged, providing valuable information for reporting and further investigation.
Conclusion: The diverse genetic makeup of Puerto Rican population necessitates the establishment of comprehensive clinical and genetic profiles of familial dyslipidemia, along with the associated risk of ASCVD. This knowledge is crucial for implementing preventative treatments and conducting cascade screening to identify individuals at increased risk of ASCVD due to familial dyslipidemias. The cataloging of VUS at a larger scale can pave the way for early interventions and improved patient outcomes.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1351 Analyzing large-scale Tourette Syndrome whole-exome sequencing data reveals a significant contribution of de novo mutations.

Authors:

L. Zhan[1], D. Yu[2], L. Domenech-Salgado[3], F. Ivankovic[4], P. Giusti-Rodriguez[5], Tourette Association of America International Consortium for Genetics (TAAICG), M. Niarchou[6], L. K. Davis[7], C. A. Mathews[8], J. M. Scharf[9], R. A. Ophoff[10]; [1]UCLA, Los Angeles, CA, [2]Massachusetts Gen. Hosp., Hopkinton, MA, [3]Massachusetts Gen. Hosp., Boston, MA, [4]Massachusetts Gen. Hosp., Arlington, MA, [5]Univ. of Florida Coll. of Med., Gainesville, FL, [6]Vanderbilt Univ. Med. Ctr., Nashville, TN, [7]Vanderbilt Univ Med Ctr., Nashville, TN, [8]Univ. of Florida, Gainesville, FL, [9]Massachusetts Gen Hosp, Newton Center, MA, [10]Univ California Los Angeles, Los Angeles, CA

Tourette syndrome (TS) is an early-onset neurodevelopmental disorder (NDD) characterized by vocal and motor tics. TS is highly heritable (60-80%) and has a complex genetic architecture with both rare and common variants contributing to the genetic etiology. However, the contribution of *de novo* variants has not been widely studied due to the lack of large-scale high-quality family-structured sequencing datasets and insufficient statistical power. In this study, we generated whole-exome sequencing (WES) data for over 1,300 TS trio families and jointly called the data with a selected subset of over 6,600 families from SSC and SPARK datasets containing at least one Autism Spectrum Disorder (ASD) proband. This produced a high-quality unified dataset with 30,803 individuals and 4,425,974 variants. Our principal component analysis showed a diverse ancestry background with all major population groups present in the ASD cohorts, while only Europeans and Admixture Americans were present in the TS cohort. Analyzing samples from all three cohorts demonstrated a trend of increased burden of *de novo* pathogenic singletons ($FC_{TS/Control}$=1.23; n.s.), defined as protein-truncating variants (PTVs), in TS probands compared to healthy controls, while confirming a significant enrichment of *de novo* mutations in ASD probands compared to healthy controls ($FC_{ASD/Control}$=1.05; $P_{ASD,Control}$=4.70*10^{-2}). Considering the phenotypic heterogeneity previously reported for the SPARK cohort, we further investigated a subset with only TS and SSC cohorts ($N_{Proband+Control}$=3,655) and revealed a significant stepwise enrichment of *de novo* pathogenic variants across healthy controls, TS probands, and ASD probands ($FC_{TS/Control}$=1.04; $P_{Control,TS}$=2.59*10^{-2}; $FC_{ASD/TS}$=1.11; $P_{TS,ASD}$=4.62*10^{-7}), recapitulating the phenotypic spectrum of NDD. In brief, our study includes one of the largest TS-ASD WES family datasets that will facilitate future genetic and clinical studies on TS and comorbidities with improved power. Moreover, our work provides evidence of the contribution of *de novo* mutations to TS etiology, as well as the shared genetic architecture between TS and ASD. Further studies will focus on locating TS-specific and TS-ASD shared risks through gene-specific and gene-set analysis, bridging rare variant effects to polygenic scores to boost modeling accuracy, and dissecting the heterogeneity in the SPARK cohort for improved statistical power.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1352 Analyzing the impact of genetic loci that decouple obesity from abdominal fat distribution through in vitro adipocyte perturbation and differentiation

Authors:

J. Belanich[1,2], M. Urena[1], T. Batista[2], T. Kilpeläinen[1]; [1]Univ. of Copenhagen, Copenhagen, Denmark, [2]Broad Inst., Cambridge, MA

Obesity is often linked to abdominal fat distribution which can negatively influence cardiometabolic health. However, the relationship between overall and abdominal fat can vary greatly among individuals due to genetic differences. The biological mechanisms that regulate this relationship remain incompletely understood. Intriguingly, some genetic loci have been found to show an inverse relationship between higher body mass index (BMI) and abdominal fat distribution, indicated by a higher waist-hip ratio (WHR). We aimed to systematically screen for genetic loci that invert the relationship between higher BMI and WHR and characterize their function in human adipocytes. Leveraging summary results from the largest published GWAS for BMI and WHR, which included ~700,000 participants each, we identified >100 independent loci associated with higher BMI but lower WHR. Several of these loci were also associated with higher insulin sensitivity and a lower risk of type 2 diabetes.To understand the biological mechanisms underlying these findings in adipose tissue, we studied the impact of two selected novel loci, *ADAMTS9* and *EMILIN2*, in mouse and human preadipocytes. In vitro knockdown of these genes using siRNA in mouse and human white preadipocytes led to decreased differentiation and function of the cells, affecting cellular metabolism as well as adipogenic marker genes. Perturbed cells experienced reduction in lipid storage, as well as glucose uptake, without affecting proliferation, or cellular size. We are consequently following up on these two genes through additional CRISPRi perturbation studies and the LipocyteProfiler platform, which utilizes a combination of staining, sequencing, and other cellular profiling methods, to detect distinct phenotypic and cellular signatures. Our studies will provide new insights into how genetic variants may uncouple obesity from abdominal fat distribution and could identify new targets for therapeutics to increase insulin sensitivity and improve cardiometabolic health. Grants from the Novo Nordisk Foundation (NNF18CC0034900, NNF20OC0063707).

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1353 Ancestry Specific Polygenic Risk Score, Dietary Patterns and Cardiovascular Disease

Authors:

D. Hardy[1], J. Garvin[2], T. Mersha[3]; [1]Morehouse Sch. of Med., Atlanta, GA, [2]Walden Univ., Minneapolis, MN, [3]Cincinnati Children S Hosp. Med. Ctr., Cincinnati, OH

Background: Healthy dietary patterns can lower risk for atherosclerotic cardiovascular disease (ASCVD). A gap in knowledge is whether the impact of healthy dietary patterns in different ancestries is polygenic risk score (PRS) dependent. **Objective:** Our main aim was to examine the associations between a risk-raising functional PRS, dietary patterns, and high physical activity with ASCVD in European Americans and African Americans. A secondary aim examined the biological pathways associated with PRS mapped genes and their relationship with dietary intake. **Methods**: Our cross-sectional study utilized de-identified data from 7-National Heart, Lung, and Blood Institute Candidate Gene Association Resource studies obtained from the Database of Genotypes and Phenotypes for European Americans (n=6,575) and African Americans (n=1,606). Data collection years spanned from 1987 to 2010. **Results:** In the highest PRS tertile, ASCVD risk increased by 59% (Risk Ratio=1.59; 95% confidence Interval:1.16-2.17) in African Americans and by 18% (RR=1.18; 1.04-1.35) in European Americans compared to the lowest tertile. The Dietary Approaches to Stop High Blood Pressure (DASH) and Mediterranean healthy dietary patterns lowered ASCVD risks in the highest PRS tertiles by 5% (RR=0.95; 0.91-0.98) to 6% (RR=0.94; 0.91-0.98) in European Americans but did not have a significant effect in African Americans. The harmful Southern diet increased ASCVD risk in African Americans by 11% (RR=1.11; 1.02-1.21) to 12% (RR=1.12; 1.03-1.23). However, within the highest PRS tertiles, high physical activity together with the DASH, or Mediterranean, or Southern diet reduced ASCVD risks by 9% (RR=0.91; 0.85-0.96) to 15% (RR=0.85; 0.80-0.90) in European Americans. Among African Americans, high physical activity with the DASH diet lowered ASCVD risk by 13% (RR=0.87; 0.78-0.97) and the Mediterranean diet by 18% (RR=0.82; 0.72-0.95). Top biological pathways included fructose metabolism and catabolism linked to obesity, insulin resistance, and type 2 diabetes present in both ancestries. Additional pathways present for African Americans were for Vitamin D which is linked to depression; aging acceleration; and death signaling associated with cancer. Conclusions: The effects of healthy dietary patterns and high physical activity can counterbalance the effects of a risk-raising functional PRS. This can give more health benefits especially in African Americans to reverse metabolic responses regardless of a high PRS burden in associated pathways.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1354 Ancestry-informed regression identifies ancestry-specific effects for multiple sclerosis risk in Hispanic / Latino and African American populations.

Authors:

A. Beecham, Alliance for Research in Hispanic MS (ARHMS) Consortium; Univ Miami, Coral Gables, FL

Despite the rising incidence of MS in African Americans and Hispanics, these communities remain underrepresented in genetic research. The genetic admixture and unique linkage disequilibrium (LD) structure inherent to these populations affords the opportunity to identify novel susceptibility variants which may be ancestry-specific. In total, 3131 African Americans (1625 MS cases, 1506 controls) and 4160 Hispanics (2046 MS cases, 2114 controls) ascertained from more than 10 sites across the United States were genotyped on a customized Illumina genome-wide array and imputed with TopMed. Local ancestry was computed with RFMixV2. Genome-wide association was performed using Tractor to obtain ancestry-specific effect size estimates. We used an inverse variance-weighted fixed-effects meta-analysis of summary statistics from the European (EUR), African (AFR), and Native American (NAM) deconvolved tracts to estimate the combined effect. We identified three loci outside of the Major Histocompatibility Complex with genome-wide significance in the meta-analysis; all within 500KB of a previously identified European MS susceptibility variants (*LCK* in 1p35.2, *CD58* in 1p13.1, and *ETV7* in 6p21.31). The most strongly associated variant in the *LCK* gene region (rs145088108, OR = 1.18, P = 1.2E-13) is a missense mutation that is common on the NAM background (MAF= 0.38) but rare otherwise (MAF= 0.00017 and 0.00044 for AFR and EUR, respectively) and represents a novel NAM-specific signal at this locus. Several additional variants within 500KB of the European identified risk variant (rs79979643) for the *LCK* gene region achieved genome-wide significance in the subset of NAM alleles but were not significant for EUR or AFR alleles, including three coding mutations in proximal genes *FAM167B, TMEM39B*, and *IQCC*. The previously identified European risk variant demonstrated association with MS in only EUR alleles (OR = 1.07, p = 3.78E-06) but not in AFR or NAM alleles, and no LD is observed between the previously identified European risk allele and any of the novel NAM alleles (R2 less than 0.2 in all three ancestral populations). In alleles residing on an AFR haplotype, we identified a novel genome-wide significant signal for MS on chromosome 13q14.2; a non-coding transcript variant (rs3803245) with a frequency of 0.09 in AFR alleles but approximately 0 in EUR and NAM alleles. We have demonstrated that novel ancestry-specific risk alleles are present both within known susceptibility loci and in novel loci. These findings highlight the utility of a trans-ethnic approach to variant discovery in a way that will facilitate treatment and prevention in diverse populations.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1355 Application of scDRS in autism spectrum disorder characterizes cell type and single cell level disease association.

Authors:

X. Li, K. Hou, K. W. Eyring, C. Liu, K. Abuhanna, Y. Zhang, D. H. Geschwind, C. Luo, B. Pasanuic; UCLA, Los Angeles, CA

Identification of cells exhibiting disease related transcriptomic changes in autism spectrum disorder (ASD) remain one of the main hurdles to better understanding ASD disease mechanism. Statistical tools such as scDRS (Zhang et al. 2022 Nature Genetics) that quantifies enrichment in gene expression toward disease associated genes and assign single cell level disease risk scores would allow interrogation of within cell type heterogeneity of disease associated transcriptomic profile.

To identify such patterns, we first searched previously published gene sets that are manually curated such as SFARI gene list (Abrahams et al. 2013 Molecular autism). We retained 931 genes with high confidence (gene score 1, 2) and used number of publications as weight. We also derived 1000 genes computed with MAGMA (de Leeuw et al. 2015 PLoS computational biology) on ASD GWAS (Grove et al. 2019 Nature genetics).

Next, we deployed these disease associated gene sets onto the publicly available transcriptomics atlas (TMS) (The Tabula Muris Consortium 2020 Nature) with scDRS and identified 123 and 56 cells with significant disease risk score (fdr < 0.1) when using SFARI gene set and ASD GWAS gene set respectively. Across cells with significant risk score, the majority of them are neuronal cells at 87.8% and 85.7% respectively when computed using SFARI gene set and ASD GWAS gene set.

To further quantify how the signal could manifest in a case and control setting, we computed scDRS on ASD scRNA dataset with case and control individuals. This approach revealed 13 cells with significant scDRS score (FDR < 0.1) when using SFARI gene set but retained little signal when gene sets computed from ASD GWAS were deployed.

In this study, we report scDRS score for ASD in the TMS atlas using SFARI gene set as well as gene set computed from ASD GWAS using MAGMA. We also followed up by applying scDRS with those two gene sets on scRNA data with case and control. We noticed SFARI gene set provided stronger signal in both settings compared to the gene set derived from ASD GWAS. We attribute this result to the small sample size of ASD GWAS and warrants additional research on including more individuals in ASD GWAS to allow for higher statistical power in detecting disease associated genes and how they manifest in disease.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1356 Arginine, Glycine, and Creatine Supplementation Improves Symptoms in a Female with *SLC6A8* Creatine Transporter Deficiency

Authors:

C. Theile, K. Tauer, J. Owens, K. Cecil, A. Shillington; Cincinnati Children's Hosp. Med. Ctr., Cincinnati, OH

X-linked Creatine Transporter Deficiency (CTD) is caused by pathogenic variants in *SLC6A8* and is characterized by intellectual disability, behavioral / developmental disorders, and seizures. The *SLC6A8* protein has critical roles in synthesis and distribution of creatine throughout the body. The central nervous system is especially reliant on *SLC6A8* to allow creatine to cross the blood brain barrier. Cerebral creatine deficiency disorders are found in up to 2.6% of males with intellectual disability, but few female cases have been reported. Small cohorts have suggested that supplementation of creatine, arginine, and glycine can stop disease progression in males and improve symptoms in females with CTD, though only four females on this therapy have been reported in the medical literature. We offer a case of a female with a de novo pathogenic *SLC6A8* variant who significantly improved with supplementation.

At time of diagnosis, the reported patient had weight loss, mild intellectual disability, attention deficit hyperactivity disorder, generalized anxiety disorder, and intermittent explosive disorder. Magnetic resonance spectroscopy (MRS) of the brain showed reduced creatine on all acquired spectra. Based on literature review of previously treated female patients, the reported patient was started on creatine-monohydrate at 200 mg/kg/day, L-arginine at 400 mg/kg/day, and L-glycine at 150 mg/kg/day. Within three weeks of starting supplementation, her caregiver reported that she was "doing better in almost every way." She gained 5 kg and symptom improvement was also seen through her mood: she was sleeping through the night, was more alert during the day, and was showing more appropriate behaviors. After 8 months of supplementation, MRS showed improved creatine concentrations with normalizing semi-quantitative ratios with other brain metabolites.

Previous small cohort studies of male patients with *SLC6A8* deficiency demonstrated neither improvement nor decline on treatment. One female supplemented only with creatine showed increased aggression and worsening of self-injurious behaviors. However, two female patients in the literature supplemented with dosages of creatine, arginine, and glycine demonstrated improvement in neuropsychiatric symptoms. The reported patient's experience in conjunction with similar results in the literature, provides evidence to clinicians trialing creatine, arginine, and glycine supplements for female patients with CTD.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1357 Assessment of the Impacts of Blood Collection Tube Types on Blood Acylcarnitine Determinations.

Authors:

D. Shao, L. Zhai, R. Giles, K. Pap, M. Natowicz; The Cleveland Clinic Fndn., Cleveland, OH

Background:
Blood acylcarnitine profile analysis is a powerful tool to diagnose numerous inherited metabolic disorders, including many mitochondrial fatty acid oxidation disorders and organic acidemias. It is used for follow-up testing of screen positive results from newborn screening programs and in the evaluation of children and adults suspected of having a fatty acid or organic acid disorder. Serum or plasma samples, the latter obtained with a variety of anticoagulants, are normally accepted for acylcarnitine profile analysis. In view of the diverse types of blood collection tubes used in acylcarnitine analyses, it is important to evaluate possible matrix effects on the measurement of the many acylcarnitines that are assessed in blood acylcarnitine assays.
Methods:
52 acylcarnitines and related analytes were measured by liquid chromatography-tandem mass spectrometry using plasma obtained from sodium heparin (green top tube), lithium heparin (green top tube), and EDTA (lavender top tube) anticoagulated tubes, as well as serum (red top tube). Samples from each blood tube type were obtained at the same time from three healthy individuals. Each sample is assayed in triplicates.
Results/Conclusions:
At least three of 50+ measured analytes showed significant (>25-30%) matrix effects. Butyrobetaine was ~30-75% higher in heparinized plasma than in serum or EDTA anticoagulated plasma. Decanoylcarnitine (C10) was ~25-30% higher in heparinized plasma or serum compared to EDTA anticoagulated plasma. Propionylcarnitine (C3) was about 25-35% lower when measured in serum than the other tube types. When such differences straddle medical decision cut-offs, diagnostic decisions can be impacted; this may be particularly relevant for mild and late-onset forms of inherited metabolic disorders. These data suggest the utility of using matrix-specific reference intervals for those analytes that differ significantly between tube types.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1358 Association between *APOL1* risk variants and progression from infection to sepsis

Authors:

L. Jiang[1], G. Liu[1], A. Oeser[1], A. Ihegword[1], A. Dickson[1], L. Daniel[2,3], A. Hung[1,4], N. Cox[1], C. Chung[2,3], W-Q. Wei[1], M. Stein[1], Q. Feng[1]; [1]Vanderbilt Univ. Med. Ctr., Nashville, TN, [2]Univ. of Miami, Coral Gables, FL, [3]Miami VA Hosp., Coral Gables, FL, [4]Tennessee Valley Hlth.care System, Nashville, TN

Abstract Importance: Two risk variants in the apolipoprotein L1 gene (*APOL1*) have been associated with increased susceptibility to sepsis in Black patients. However, it remains unclear whether *APOL1* high-risk genotypes are associated with either progression from infection to sepsis or sepsis-related phenotypes, independent of their association with severe renal disease.**Objective:** To examine the association between *APOL1* high-risk genotypes and the risk of progression from infection to sepsis and sepsis-related phenotypes.**Design, setting, and participants:** A retrospective cohort study of 2,242 Black patients hospitalized with an infection.**Exposure:** *APOL1* high-risk genotypes.**Main outcomes and measures:** The primary outcome was sepsis; secondary outcomes were in-hospital death and organ failure related to sepsis.**Results:** Of 2,242 Black patients hospitalized with infections, 565 developed sepsis. Patients with *APOL1* high-risk genotypes had a significantly increased risk of sepsis (odds ratio [OR]=1.29 [95% CI, 1.00-1.67; p=0.047]); however, this association was not significant after adjustment for pre-existing severe renal disease (OR=1.14 [95% CI, 0.88-1.48; p=0.33]), nor after exclusion of those patients with pre-existing severe renal disease (OR=0.99 [95% CI, 0.70-1.39; p=0.95]). *APOL1* high-risk genotypes were significantly associated with the renal dysfunction component of the Sepsis-3 criteria (OR=1.64 [95% CI, 1.21-2.22; p=0.001], but not with other sepsis-related organ dysfunction or death. The association between high-risk *APOL1* genotypes and sepsis-related renal dysfunction was markedly attenuated by adjusting for pre-existing severe renal disease (OR=1.36 [95% CI, 1.00-1.86; p=0.05]) and was no longer significant after exclusion of patients with pre-existing severe renal disease (OR=1.16 [95% CI, 0.74-1.81; p=0.52]). A restricted PheWAS, designed to assess significant outcomes from a previous PheWAS among Million Veteran Program (MVP) patients, yielded similar results; we replicated the association between high-risk *APOL1* genotypes and sepsis-related phenotypes previously reported in MVP, and additionally found that those associations were attenuated or no longer present after accounting for pre-existing renal disease.**Conclusion and relevance:** *APOL1* high-risk genotypes were associated with an increased risk of sepsis; however, this increased risk was attributable predominantly to pre-existing renal disease.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1361 Association of OCT-1 gene polymorphism in type II diabetes mellitus patients from North Indian population

Authors:

H. Parveen, S. kumari; King George Med. Univ., Lucknow, Lucknow, India

Background: Diabetes mellitus has become a most important health problem worldwide in recent times. **Aims and objective:** To investigate the possible association of organic cation transporter- 1(OCT1) gene polymorphism (A-G) in type II Diabetes mellitus. **Methodology**: 100 diagnosed T2DM patients were recruited for the study and the genotypes for OCT-1 gene polymorphism using polymerase chain reaction (PCR) followed by restriction fragment length polymorphism (RFLP) were done. **Result:** We did not find any significant association between GG homozygous alleles ($p<0.079$) and AG heterozygous alleles ($p=0.209$) in type 2 diabetes mellitus patients compared to the control. It this case-control study, the frequency of the G allele of OCT-1 was found significant in alcoholic type 2 diabetes mellitus patients ($p=0.026$*). **In conclusion,** OCT-1 gene G allelic polymorphism is associated with type 2 diabetes mellitus patients in the north Indian population.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1362 Association of rare functionally deficient *SMPD1* mutations with Parkinson's disease

Authors:

E. Chew[1], Z. Liu[2,3], Z. Li[3], S. Chung[4], M. Lian, M[1], M. Tandiono[1], E. Y. Ng[5], L. C. Tan[5], W. Chng[2], T. Tan[3], E. K. Peh[6], Y. Ho[6], X. Chen[3], E. Y. Lim[3], C. Chang[1], J. J. Leong[1], Y. Heng[7], T. Peh[3], L-L. Chan[8], Y. Chao[5], W-L. Au[5], K. M. Prakash[5], A. Ahmad Annuar[9], S-Y. Lim[9], A-H. Tan[9], J. Lim[9], Y. Tay[9], V. Mok[10], A. Y. Chan[11], J-J. Lin[12], B. S. Jeon[13], K. Song[14], L. Robak[15,16], R. Alcalay[17], Z. Gan-Or[18], M. Makarious[19,20], C. Blauwendraat[19], M. Nalls[19,21], C. C. Tham[22], C. Pang[22], J. Ahn[23,13], K. Park[13], J. Wiggs[24], T. Aung[25,2], C-C. Khor[3,2], E-K. Tan[5,2], Z. Wang[2,3], J. Foo[1,3]; [1]Nanyang Technological Univ. Singapore, Singapore, Singapore, [2]Duke-Natl. Univ. of Singapore Med. Sch., Singapore, Singapore, [3]Genome Inst. of Singapore, Singapore, Singapore, [4]Asan Med. Ctr., Seoul, Korea, Republic of, [5]Natl. NeuroSci. Inst., Singapore, Singapore, [6]Bioprocessing Technology Inst., Singapore, Singapore, [7]Nanyang Technological Univ., Singapore, Singapore, [8]Singapore Gen. Hosp., Singapore, Singapore, [9]Univ. of Malaya, Kuala Lumpur, Malaysia, [10]Margaret K.L. Cheung Res. Ctr. for Management of Parkinsonism, Gerald Choa NeuroSci. Inst., Lui Che Woo Inst. of Innovative Med., Li Ka Shing Inst. of Hlth.Sci., Div. of Neurology, Dept. of Med. and Therapeutics, Hong Kong, China, [11]Margaret K.L. Cheung Res. Ctr. for Management of Parkinsonism, Lui Che Woo Inst. of Innovative Med., Div. of Neurology, Dept. of Med. and Therapeutics, Prince of Wales Hosp., The Chinese Univ. of Hong Kong, Hong Kong, China, [12]Chushang Show-Chwan Hosp., Nantou, Taiwan, [13]Seoul Natl. Univ. Hosp., Seoul, Korea, Republic of, [14]Univ. of Ulsan Coll. of Med., Seoul, Korea, Republic of, [15]Baylor Coll. of Med., Houston, TX, [16]Texas Children's Hosp., Houston, TX, [17]Tel Aviv Sourasky Med. Ctr., Tel Aviv, Israel, [18]Montreal Neurological Inst., McGill Univ., Montreal, QC, Canada, [19]NIH, Bethesda, MD, [20]Univ. Coll. London, London, United Kingdom, [21]Data Tecnica Intl. LLC, Glen Echo, MD, [22]The Chinese Univ. of Hong Kong, Hong Kong, China, [23]SMG-SNU Boramae Med. Ctr., Seoul, Korea, Republic of, [24]Harvard Med Sch, MEEI, Boston, MA, [25]Singapore Eye Res. Inst., Singapore, Singapore

Parkinson's disease is a progressive neurodegenerative disease that adversely affects balance, muscle control and movement. Current understanding of Parkinson's disease pathogenesis mechanisms is poor and can be improved by gleaning mechanistic insights from rare protein-altering genetic variants associated with the disease.
We performed whole-exome sequencing on a total of 9,810 individuals (Discovery cohort: 4,298 Parkinson's disease patients and 5,512 unaffected controls) of East Asian ancestry from Singapore, Malaysia, Hong Kong, South Korea and Taiwan. After conducting stringent filtering for sample quality, variant quality and variant pathogenicity prediction, we tested for gene-based association of rare protein-altering variants with Parkinson's disease status and observed two genes (*GBA1* and *SMPD1*) which surpassed exome-wide significance ($P<2.5\times10^{-6}$). Exome-wide significant association of both *GBA1* and *SMPD1* was replicated in another 11,227 individuals (Replication cohort: 5,585 Parkinson's disease patients and 5,642 unaffected controls) of Asian and European ancestry.
The association of rare protein-altering *GBA1* variants with Parkinson's disease has been well established, while reports of rare protein-altering *SMPD1* association in Parkinson's disease in the literature is inconsistent and implicated *SMPD1* variants of unclear pathogenicity. As *SMPD1* encodes for the enzyme acid sphingomyelinase, we conducted functional assay for enzyme activity level of rare protein-altering *SMPD1* variants observed in the Discovery and Replication samples. We established that rare protein-altering *SMPD1* variants with less than 44% of normal acid sphingomyelinase activity were associated with Parkinson's disease risk in both Discovery (odds ratio (OR) = 2.37, 95% CI = 1.68 - 3.35, $P = 4.35\times10^{-7}$) and Replication cohorts (OR = 2.18, 95% CI = 1.69 - 2.81, $P = 4.80\times10^{-10}$), thus resulting in highly significant association observed when all cohorts were meta-analyzed (OR = 2.24, 95% CI = 1.83 - 2.76, $P = 1.25\times10^{-15}$).
Our findings show that rare functionally deficient *SMPD1* variants are associated with increased risk of Parkinson's disease and emphasize the significance of sphingomyelin metabolism in the pathobiology of Parkinson's disease. Further research is warranted to delve into the precise mechanisms underlying impaired sphingolipid and ceramide metabolism in the development of Parkinson's disease. Our approach also underscored the usefulness of functional assays to delineate pathogenic variants in exome sequencing investigations.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1363 Association of angiotensinogen rs699 SNP with pre-eclampsia in Lagos, Nigeria: Case-control study

Authors:

T. Fagbayi[1], A. O. Somade[2], A. K. Adefemi,[3], O. G. Shonde[1]; [1]Univ. of Lagos, AKOKA, YABA,, Nigeria, [2]Univ. of Lagos, Akoka, Nigeria, [3]Lagos State Univ. Teaching Hosp. (LASUTH), Lagos,, Nigeria

Pre-eclampsia is a pregnancy-specific disorder that is influenced by genetics and socio-medical factors. Angiotensinogen rs699 (AGT rs699) polymorphism affects spiral arteries in the uterus and controls blood pressure. This SNP has been associated with pre-eclampsia and there have been conflicting results in certain populations of negative associations. This study aimed to investigate the association of AGT rs699 and ascertain the socio-medical factors with pre-eclampsia in Lagos, Nigeria. Ethical approval and informed consent were obtained from Health Research and Ethical committees. Thirty (30) preeclamptic (P) and 39 normotensive (N) pregnant women from a selected maternity hospital were recruited for this study. Demographic characteristics, medical history, and present medical conditions were assessed using questionnaires. Blood samples were collected and AGT rs699 was genotyped. Descriptive statistics, independent sample t-test, chi-square, and multiple regression analysis were used to analyze the data obtained ($p < 0.05$). The mean age and weight of women with pre-eclampsia were higher in preeclamptic women with no significant difference (P = 32.43 ± 5.76, N = 30.97 ± 4.54, and P =79.04 ± 10.28, N = 76.82 ± 15.35 respectively). Diastolic blood pressure (DBP), Systolic blood pressure (SBP), and Proteinuria were significantly higher in preeclamptic patients. Surprisingly, the incidence of AGT rs699 was significantly higher in normotensive (82%) women than in preeclamptic women (17%) ($X2 = 29.15$, $p=<0.05$). A significant correlation was observed between AGT rs699 and SBP in normotensive patients ($r2 = 0.201$, $p=0.003$) and a negative correlation in preeclamptic patients. In contrast, AGT rs699 was negatively related to DBP with no statistical difference in both the normotensive and preeclamptic groups. In addition, no correlation was observed between AGT rs699 and weight nor was there any significant difference. In conclusion, this study revealed a negative significant association between pre-eclampsia and AGT rs699 among cases and the control population. Invariably, this will serve as baseline information and strengthen population-specific biomarkers for pre-eclampsia.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1364 Associations between Microbiome-Associated Variants and Diseases: Insights into the Role of the Gut Microbiome in Deep Vein Thrombosisin diverse populations.

Authors:

T. Cherlin[1], C. Morse[1], S. Bordenstein[2], A. Verma[1], S. Verma[1]; [1]Univ. of Pennsylvania, Philadelphia, PA, [2]Penn State Univ., University Park, PA

Recent efforts have been made to understand the role of gut microbiome in disease etiology. Using high throughput sequencing, studies have investigated the microbiome's association with diseases and genetic variants Markowitz et al. discovered 33 significant microbiome-associated variants (MAVs) linked to diseases in a European population, but none identified in African population. We aim to identify MAVs in diverse populations by including more MAVs from various populations and conducting phenome-wide association studies (PheWAS) in large cohorts like the Penn Medicine Biobank (PMBB). We extended the initial 908 MAVs from 11 microbiome GWASs to 1,003 unique MAVs from 13 GWAstudies including the two newer non-European population studies: the Multi-ethnic cohort (MEC) and Helius. We performed PheWASs for these 1,003 variants and 1,227 phenotypes in an African population (n = 10,995) and European populations (n = 28,853) in the PMBB dataset. Subsequently, a meta-analysis was conducted. Notably, one MAV (rs8176645) showed Bonferroni significant association (p-value<$4.06x10^{-08}$[SS1]) with deep vein thrombosis (DVT), venous embolism, replicating the findings of Markowitz et al. Our findings demonstrate significant associations between microbiome-associated variants (MAVs) from microbiome GWASs and diseases in individuals of African populations. Particularly, we identified a MAV (rs8176645) associated with DVT, indicating the potential role of the gut microbiome in DVT development. These results suggest that specific microbiome compositions or MAVs may contribute to the susceptibility or progression of DVT. Understanding the microbiome's involvement in DVT has implications for novel preventive and therapeutic approaches. Further investigation is needed to elucidate the underlying mechanisms of the MAV-DVT association in diverse populations. Future studies can explore the functional implications of these MAVs, examine the interplay between genetic factors and the gut microbiome, and evaluate microbiome-based interventions for DVT management and prevention. In summary, our findings provide valuable insights into the associations between MAVs from microbiome GWASs and diseases, particularly highlighting the potential significance of the gut microbiome in DVT. This research sets the stage for further investigations aiming to uncover the mechanisms and clinical implications of microbiome-DVT associations, potentially advancing personalized medicine and precision therapeutics for DVT patients.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1365 Assortative mating and parental genetic similarity drive pathogenicity of variably expressive variants

Authors:

C. Smolen[1], M. Jensen[2], D. Banerjee[3], L. Pizzo[4], 16p12.1 Deletion Consortium, S. Girirajan[5]; [1]Pennsylvania State Univ., State College, PA, [2]Yale Unversity, New Haven, CT, [3]Penn State, State College, PA, [4]Univ. of Utah, Salt Lake City, UT, [5]Pennsylvania State Univ., University Park, PA

Many disease-associated rare variants are variably expressive, conferring a range of neurodevelopmental features at varying levels of penetrance. Understanding the mechanisms underlying this variable expressivity is critical in predicting the phenotypic trajectory of individuals carrying these variants. Given the well-known role of family history in predicting phenotypes in affected individuals, we examined phenotypic and genetic patterns in parents associated with neurodevelopmental disease risk in children by analyzing more than 38,000 spouse pairs form the UK Biobank and parental pairs from four neurodevelopmental disease cohorts. We identified correlations between neurological and psychiatric features in parents and children, including clinical diagnoses, such as obsessive-compulsive disorder ($R=0.31-0.49$, $p<0.001$), and sub-clinical autism features in parents affecting autism severity in children, such as bi-parental mean Social Responsiveness Scale (SRS) scores affecting proband SRS scores (regression coefficient=0.11, $p=0.003$). We also observed patterns of both within-and cross-disorder correlations for seven neurological and psychiatric phenotypes among spouse pairs across cohorts, including a within-disorder correlation for depression ($R=0.25-0.72$, $p<0.001$) and a cross-disorder correlation between schizophrenia and personality disorder ($R=0.20-0.57$, $p<0.001$). Further, these spouses with similar phenotypes were also significantly correlated for rare variant burden ($R=0.07-0.57$, $p<0.0001$). We propose that assortative mating on these neurological and psychiatric features drives the increases in genetic risk over generations and “anticipation” associated with many variably expressive variants. We further identified parental relatedness as a risk factor for neurodevelopmental disorders through its inverse correlations with burden and pathogenicity of rare variants and propose that parental relatedness drives disease risk by increasing genome-wide homozygosity in children ($R=0.09-0.30$, $p<0.001$). Our results highlight the utility of assessing parent phenotypes and genotypes in predicting features in children carrying variably expressive variants and counseling families carrying these variants and provide insights into the phenotypic trajectory of families carrying these variants.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1366 Bilateral hearing loss & mild dysmorphism associated to a DSPP gene variant.

Authors:

A. Rodriguez Rivera[1], A. Cornier[2], E. Albino[3], S. Carlo[4]; [1]Univ. of Puerto Rico Bayamon Campus, Bayamon, PR, [2]San Jorge Children s & Women s Hosp., San Juan, PR, [3]Ponce Sch. of Med., Ponce, PR, [4]Ponce Hlth.Sci. Univ., Cabo Rojo, PR

Non-syndromic sensorineural hearing loss can represent a diagnostic challenge from a molecular point of view. We are presenting the case of a 10-year-old boy who was initially referred to Genetics due to bilateral sensorineural hearing loss. A comprehensive hearing loss gene panel demonstrated genetic variants of DSPP and OTOf genes. DSPP c1786_1788del & OTOF c.788C>T were identified. DSSP is associated with autosomal dominant disease and is predicted to result in an in-frame deletion in the DSPP gene. This variant has been associated with a clinical condition in the Human Gene Mutation Database (HGMD). Although the exact function (s) of DSPP-derived proteins are unknown, these appear to be essential for the development and mineralization of the tooth. The DSPP gene is also presently active in the inner ear with an unclear role in the hearing process. Hearing loss was initially documented at age 5 years when language & learning difficulties were identified. Hearing test & ABR studies identified bilateral sensorineural mild to moderate hearing loss at high frequencies of 3 -8K Hz., that has remained stable for the last 5 years. Clinical description of the patient includes synophrys of the eyebrows, epicanthal folds, brad nasal bridge, low set, posteriorly rotated & overfolded ears. Flat philtrum & thin upper lip were also documented upon examination. He presents mild dentinogenesis imperfecta with history of frequent and multiple dental caries. Chromosomal microarray studies were essentially normal: arr[GRch37] 14q21.3q23.1 x2 hmz. Interestingly, the maternal grandfather & a grand uncle & aunt presented moderate hearing loss throughout their lives, but his mother does not, suggesting a reduced penetrance of this genetic variant (gene). As mentioned before the patient also carries a heterozygous genetic change on the OTOF gene predicted to be pathogenic. This gene is associated with autosomal recessive hearing loss. We do not know if this haploinsufficiency has any potential adding effect on this patient hearing loss.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1367 Biological Insight into Alzheimer's Disease Sex Dimorphism Leveraging Predictive Clinical Models, Knowledge Graphs, and Colocalization Analysis

Authors:

A. Tang[1], K. P. Rankin[1], G. Cerono[1], S. Miramontes[1], H. Mills[1], J. Roger[1], B. Zeng[1], C. Nelson[1], K. Soman[1], S. Woldemariam[1], Y. Li[1], A. Lee[1], R. Bove[1], M. Glymour[1], T. Oskotsky[1], Z. Miller[1], I. E. Allen[1], S. J. Sanders[2], S. Baranzini[1], M. Sirota[1]; [1]UCSF, San Francisco, CA, [2]Univ. of Oxford, Oxford, United Kingdom

Predicting Alzheimer's Disease (AD) risk can allow for insights or interventions before irreversible disease progression. Electronic health records (EHRs) are emerging as a rich source of longitudinal data that can in combination with knowledge graph databases like SPOKE (Scalable Precision Medicine Open Knowledge Engine) and genetic colocalization analysis elicit biological meaning from complex constellation of phenotypic data. We trained random forest (RF) models on the UCSF EHR data (conditions, drugs, abnormal measurements) to predict AD onset, with index time as the first dementia diagnosis or drug prescription among the AD cohort and one year before last visit for controls. Models were also trained with sex stratification, and with matching, at various time points before index time, and evaluated with AUROC/AUPRC on 30% held-out sample. Top clinical conditions were mapped to the SPOKE knowledge graph for high order biological insight. Select clinical predictors were validated in an independent EHR across the UC health system with survival analysis, and genetic validation with colocalization analyses and evidence from UK Biobank and Open Targets Genetics. From over 5 million patients in the UCSF EHR, 2,996 AD patients were identified. After quality filtering, 749 AD (468 F, 281 M) and 250,545 controls (139,548 F, 110,829 M) were selected. RF models predict AD onset with mean AUROC/AUPRC of at most .86/.07 (.85/.07 F, .82/.08 M) at -7 years to .89/.18 (.87/.13 F, .85/.16 M) at -1 day. Top features of models trained on matched cohorts identified predictors among all patients (e.g., hyperlipidemia, dizziness), female-enriched (e.g. osteoporosis), or male-enriched (e.g. prostate hyperplasia). Survival analysis of osteoporosis exposure in external EHRs supports a significantly increased risk of AD diagnosis (HR 1.81 (95% CI 1.70-1.92) p<.005). SPOKE network highlight high-order shared genetic relationships between osteoporosis and AD including ALB, IL6, TNF, INS, and HFE. Genetic colocalization of osteoporosis and AD identifies a shared locus near MS4A6A with stronger association in females, which is further supported by MS4A6A eQTL relationship with AD and female heel bone mineral density and literature association between MS4A6A and AD and bone density. This study demonstrates opportunity to integrate molecular and clinical data to derive biological insight, leading to identification of MS4A6A relationship for both osteoporosis predictor and AD with possible sex specificity. Future extensions of this work include AD early identification and personalized insights to inform hypotheses between biology, sex, and AD.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1368 Can Metformin inhibit abdominal aortic aneurysm progression? A Mendelian randomisation study.

Authors:

K. Saxby[1], C. Gellatly[1,2], S. Stoma[1,2], A. Wanhainen[3], J. S. Lindholt[4], J. Golledge[5], M. Sabater-Lleal[6,7], R. Lee[8], E. Strauss[9], F. Dudbridge[10,2], M. J. Bown[1,2], C. P. Nelson[1,2]; [1]Dept. of Cardiovascular Sci., Univ. of Leicester, Leicester, United Kingdom, [2]NIHR Leicester BioMed. Res. Ctr., Univ. of Leicester, Leicester, United Kingdom, [3]Dept. of Surgical Sci., Vascular Surgery, Uppsala Univ., Uppsala, Sweden, [4]Dept. of Cardiothoracic and Vascular Surgery, Odense Univ. Hosp., Denmark, Denmark, [5]Queensland Res. Ctr. for Peripheral Vascular Disease, James Cook Univ. and Townsville Univ. Hosp., Townsville, Australia, [6]Genomics of Complex Disease Unit, Inst. d'Investigació Biomèdica Sant Pau (IIB Sant Pau), Barcelona, Spain, [7]Cardiovascular Med. Unit, Dept. of Med., Karolinska Inst.t, Ctr. for Molecular Med., Stockholm, Sweden, [8]Nuffield Dept. of Surgical Sci., Univ. of Oxford, John Radcliffe Hosp., Oxford, United Kingdom, [9]Inst. of Human Genetics, Polish Academy of Sci., Poznan, Poland, [10]Dept. of Population Hlth.Sci., Univ. of Leicester, Leicester, United Kingdom

Abdominal aortic aneurysm (AAA) is a dilatation of the body's main artery estimated to affect 0.92% of adults (aged 30-79) worldwide. The risk of rupture increases as an aneurysm grows, which is fatal without surgical intervention. There is no treatment to slow aneurysm growth aside from lifestyle recommendations including smoking cessation. We have previously used Mendelian Randomisation (MR) to identify evidence of a causal link between Metformin, a drug prescribed to treat type 2 diabetes, and a reduction in AAA risk. We now use MR to test for evidence of a causal effect between several Metformin pathways and AAA progression. We conducted two sample MR analyses using previously published genetic instruments for several Metformin drug pathways associated with both gene expression and decreased glycated haemoglobin (HbA1c) levels. Effect sizes were obtained from within UK Biobank for HbA1c, and from the largest existing AAA progression genomic dataset, consisting of 6,760 patients of European ancestry. We found evidence of a causal association of the AMPK Metformin pathway with AAA progression where a reduction in HbA1c is associated with slower AAA progression. We found no evidence of a causal association between a combined multiple pathway Metformin instrument and AAA progression. Clinical trials currently underway will assess the efficacy of Metformin in reducing AAA growth.
Grant references: This work was supported by the Wellcome Trust and British Heart Foundation grants.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1369 Candidate and genome wide pathway analyses of healthy aging in 'Super Seniors'

Authors:

R. Hoque[1], A. Brooks-Wilson[2]; [1]BC Cancer Genome Sci. Ctr., Vancouver, BC, Canada, [2]BritishColumbia Cancer Res Ctr, Vancouver, BC, Canada

The Super Seniors Study aims to identify genetic factors that influence healthy aging. 'Super Seniors' are individuals 85 or older who have never been diagnosed with cancer, cardiovascular or major pulmonary disease, diabetes or dementia. We conducted candidate and genome-wide pathway analyses (GWPA) of the Super Seniors' phenotype to identify biological pathways for which genetic variation may contribute to healthy aging. The analysis was performed using 6.6 M imputed variants, in individuals of European ancestry. Cases were 541 Super Seniors; controls were 373 population-based mid-life individuals ascertained without regard to health status. Using MAGMA-1.08b, gene-based association tests were performed, then the outputs of the gene-based association were used to test for pathway-based associations that compare genes in a specific pathway with rest of the genome. The KEGG database was used to define gene sets for pathway analyses. Initially, analysis of 3 candidate pathways identified that one pathway, insulin signaling (P = 0.008) remained significant after Bonferroni correction for 3 tests, whereas two did not: the mammalian (mechanistic) target of rapamycin (mTOR) signaling (P = 0.58), and AMP-activated protein kinase (AMPK) signaling (P = 0.10). For further exploration, GWPA was performed for 186 gene sets representing all KEGG canonical pathways. No pathway remained significant after Bonferroni adjustment for 186 tests. The pathways were ranked based on their P values. Top ranked pathways included several known to be relevant to healthy aging or longevity, including Alzheimer disease (P = 0.003), P53 signaling (P = 0.004), Insulin signaling (P = 0.008), Type 1 diabetes mellitus (P = 0.01), Huntington disease (0.02), Lysosome (P = 0.02), nicotinate and nicotinamide metabolism (P = 0.02), phosphatidylinositol signaling (P = 0.03), glycosaminoglycan degradation (P = 0.03), TGF Beta signaling (P = 0.03), mitogen-activated protein kinase (MAPK) signaling (P = 0.04), adipocytokine signaling (P = 0.04), and apoptosis (P = 0.04). Though the results are not genome wide significant, the results of candidate pathway analyses and the observation of pathways of known relevance among the lowest p-values supports that multiple genetic variants in these pathways may contribute to healthy aging in this unique cohort of Super Seniors.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1370 Candidate genetic modifier of Alagille Syndrome, *THROMBOSPONDIN2*, worsens the liver phenotype in a transgenic Alagille Syndrome zebrafish model

Authors:

E. Keefer-Jacques, N. Valente, G. Matwijec, J. Nassur, B. Wilkins, K. Loomes, N. Spinner, M. Gilbert; The Children's Hosp. of Philadelphia, Philadelphia, PA

Alagille syndrome is a multi-organ, autosomal dominant disease caused by loss-of-function mutations in the Notch signaling genes *Jagged1* (*JAG1*) and *NOTCH2*. Variable expressivity with the associated liver disease is a hallmark feature of Alagille syndrome, however no genotype-phenotype correlation has been established. We previously identified the matricellular protein, Thrombospondin2 *(THBS2*), as a candidate genetic modifier of liver disease severity and showed that it interacts with NOTCH2 and is expressed in the bile ducts. These results indicated that *THBS2* overexpression reduced Notch signaling and could be associated with a worsened liver phenotype. We are now investigating the functional effects of THBS2 overexpression using an Alagille syndrome zebrafish model. We created transgenic fish overexpressing human THBS2 from a liver-specific promoter and crossed this to a second line containing a loss-of-function mutation in the *JAG1* zebrafish homolog, *jag2b*. These loss-of-function/transgenic fish (*jag2b*+/-; *THBS2*+) had lower survival rates than fish without THBS2 overexpression, with some exhibiting signs of heart edema and curved spines. The *jag2b*+/-; *THBS2*+ fish that survived to adulthood were smaller than those without THBS2 overexpression. To measure liver function, we used a fluorescent lipid, BODIPY FL C5, and measured the accumulation of fluorescence in the gallbladder as a readout of bile flow. We found that *jag2b*-/-; *THBS2*+ fish had reduced gallbladder fluorescence than *jag2b*-/-; *THBS2*- fish, indicating a more severe liver defect in fish with more THBS2. Additionally, gallbladders were larger in *jag2b*+/+ and *jag2b*+/- fish than in *jag2b*-/-fish but did not differ in size between *THBS2*- and *THBS2*+ expression. This shows that while *THBS2* influences bile flow, it is not the driver of gallbladder size. Further characterization, such as immunohistochemistry to study liver morphology and gene expression analysis of downstream Notch signaling genes is planned to better elucidate the effect of THBS2 overexpression. We expect that these studies will help to explain the phenotypic variability of Alagille Syndrome and lead to improvements in diagnostics and the ability to better predict disease course.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1371 Cannabis and psychosis: Genetic pathway analysis.

Authors:

I. Austin-Zimmerman[1], E. Johnson[2], E. Vassos[1], G. Trotta[1], E. Spinazzola[1], R. Murray[1], M. Di Forti[1]; [1]King's Coll. London, London, United Kingdom, [2]Washington Univ. Sch. of Med., St. Louis, MO

Cannabis use has been associated with an increased risk of psychosis and psychotic like experiences. Previous studies have demonstrated that this is a dose-dependent relationship, where more frequent use or the use of high potency cannabis is associated with increased risk. Both psychosis and cannabis use are complex and heritable traits. Previous GWAS have identified evidence of shared genetic liability for these traits. However, the specific genes and pathway underlying the association have not been extensively studied. We have built on the existing literature by conducting localised genetic correlation and pathway analyses to identify the shared pathways and pleiotropic variants, and to define potential differences in the genetic architecture of psychosis that develops with and without cannabis use.

We have meta-analysed the most recent GWAS of schizophrenia and bipolar disorder from the Psychiatric Genetics Consortium (PGC) to create a broad psychosis phenotype, as well as data from the Genomic Psychiatry Cohort (GPC) to enable cross-population analyses. We use these summary statistics and a recent GWAS of cannabis use disorder to the genome-wide genetic correlation between the traits, within and across ancestry. We conduct localised genetic correlation analysis using SUPERGNOVA to fine map the specific regions that contribute to the overall genetic correlation. We used PASCAL to conduct genetic pathway analysis in psychosis and CUD. A meta-analysis of schizophrenia and bipolar disorder GWAS results identified 413 independent genomic risk loci, and 1,479 mapped genes. The genetic correlation between the primary studies (PGC schizophrenia and PGC bipolar) was r^2=0.7. MAGMA analysis reveals enrichment of gene sets in the brain, including the frontal cortex, basal ganglia, and hippocampus, as well as the pituitary. SNP-based heritability (h^2) for our broad psychosis phenotype is estimated to be 0.14±0.004, p=$6.3x10^{-25}$. SNP-based h^2 = CUD is 0.12±0.01, p=$3.6x10^{-33}$. The global genetic correlation between CUD and the broad psychosis phenotype was 35% (p=$3.2x10^{-30}$). We identified three regions of local genetic correlation, after multiple testing correction, on chromosomes 1, 2, and 11. Many genes within these regions have been previously linked to a range of neuropsychiatric traits and substance use. Genetic pathway analysis identified nine pathways of potential shared importance between the two traits. These pathways played a role in synaptic transmission, neurodevelopment, signal transduction, and metabolism.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1372 Capturing Personal Health History and Family Health History in Genomic Risk Assessment Framework

Authors:

J. Cortopassi, J. Cimino, B. Davis, C. Dillon, B. Etheridge, E. Fields, A. Garrett, B. Goff, R. Johnson, B. Korf, J. Lucky, M. Wyatt, N. Limdi; Univ. of Alabama at Birmingham, Birmingham, AL

Genomic risk assessments are increasingly being used to predict risk of common complex diseases. These assessments present monogenic and polygenic risks while considering (or integrating) clinical factors, personal health history (PHHx) and family health history (FHHx).

To enable providers to contextualize the genomic risk(s) for the individual patient, accurate documentation of PHHx and FHHx is essential. However, reliance solely on patient self-report can introduce inaccuracy and/or missingness and may result in inadequate contextualization of genomic risk.

The Electronic Medical Records & Genomics (eMERGE) study aims to integrate genomic, clinical, family history and lifestyle/behavioral data into a genome informed risk assessment for several common chronic conditions. As part of the eMERGE network, the University of Alabama at Birmingham (UAB) has a focus on recruiting underrepresented and medically underserved patients. Here, we share our experience in obtaining PHHx and FHHx and identify discrepancies between actual and perceived risk of disease. We present with a focus on coronary heart disease and breast cancer, as these are conditions that combine genomic, clinical, PHHx, and FHHx data to generate an integrated risk score in the eMERGE study. Methods to minimize inaccuracies in PHHx and FHHx are considered.

Of the 2405 adults enrolled at UAB, as of June 2023, 72% are underrepresented or medically underserved. Of the 268 patients with results, 22.0% were found to be at high polygenic risk for one or more condition, 2.2% had one high monogenic risk, 2.2% had a high integrated risk for breast cancer and 31.7% had a high FHHx for at least one condition. We evaluated self-reported PHHx and FHHx and assessed the influence of missingness of data elements that trigger high risk. Missingness was also assessed by social determinants of health and under-represented status.

As genomic medicine efforts expand to include more under-represented and medically underserved populations, accurate recalling and reporting of FHHx and PHHx is needed to minimize exacerbating disparities for medically underserved groups. This is vital to ensuring that all individuals benefit from medical advances.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1373 Cardiovascular measures from abdominal MRI provide insights into abdominal vessel genetic architecture

Authors:

E. Sorokin[1], N. Basty[2], M. Thanaj[2], B. Whitcher[2], Y. Liu[1], J. Bell[2], E. Thomas[2], M. Cule[1]; [1]Calico Life Sci. LLC, South San Francisco, CA, [2]Univ. of Westminster, Res. Ctr. for Optimal Hlth., London, United Kingdom

Population-scale cardiac imaging studies have yielded many insights into the genetic basis of cardiac structure and function, by applying deep-learning-based segmentation models to tens of thousands of images of genotyped participants. In this study, we extend these principles to abdominal magnetic resonance imaging (MRI) to explore the genetic basis of quantitative traits both within and beyond the thoracic region typically studied with cardiac MRI or ultrasound.

We developed a deep learning model to segment the heart, aorta, and vena cava from abdominal MRI scans of 44,541 UK Biobank participants. From these, we generated six image-derived phenotypes (IDPs): heart volume, and four aortic and one vena cava cross-sectional areas (CSA). We performed genome- and phenome-wide association studies, and constructed a polygenic risk score for each phenotype.

Traits derived from abdominal MRI are highly correlated with previously published traits derived from cardiac MRI. We replicated previous findings related to age-related changes in heart and vessel dimensions. Infrarenal descending aorta CSA was associated with increased risk of abdominal aortic aneurysm, and heart volume was associated with several cardiovascular disorders. We demonstrated substantial genetic correlation with cardiovascular traits including aneurysms, varicose veins, dysrhythmia, and cardiac failure. Heritability enrichment analysis confirmed vascular tissue in the heritability of these traits. Genetic association study revealed 72 associations at 59 independent loci, 15 of which are novel. Colocalization with eQTL data implicated *MRC2*, which plays a major role in collagen degradation, in infrarenal aortic diameter. We derived a polygenic risk score for each trait and demonstrated an association with thoracic aortic aneurysm, pointing to a potential screening method for individuals at elevated risk for this condition.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1374 Categorization of Alzheimer's Disease Risk Variants Identifies Tissue Shared and Tissue Specific Genetic Regulatory Effects

Authors:

R. Sale[1], M. Betti[1], G. Kass[1], E. Gamazon[1,2]; [1]Vanderbilt Univ., Nashville, TN, [2]Claire Hall, Cambridge Univ., Cambridge, United Kingdom

Genome Wide Association Studies (GWAS) of Alzheimer’s Disease (AD) have identified over 70 associated loci. Many of these associations lie in intronic and intergenic regions and are assumed to regulate gene transcription. Identifying a causal link between genetic variation and gene expression may identify novel therapeutic targets for AD. Furthermore, AD presentation is heterogeneous and by partitioning AD risk variants into functional and tissue specific categories, one can identify context-specific risk factors for personalized medicine approaches. To understand the tissue and cell type specific genetic regulation of gene expression, we utilized a functional prediction framework developed in the Gamazon lab, CoRE-BED. CoRE-BED leverages cell types and various functional assays from the EpiMap compendium to train a decision tree that categorizes individual variants into cell type specific functional categories, such as promoter or enhancer. These functional annotations are then utilized to estimate cell type contributions to AD heritability and risk by generating partitioned heritability and cell type specific Polygenic Risk Scores (PRS) in each cell type. Finally, to determine causal relationships between cell-type dependent gene expression and AD, we utilized Mendelian Randomization with Joint Tissue Imputation (MR-JTI). MR-JTI leverages instrumental variables (Z) to estimate the causal effect of a gene (G) on the trait (Y). This causal framework will provide mechanistic insights into AD genetic predisposition.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1375 Causal association between celiac disease & autism spectrum disorder: A twosample Mendelian randomization approach.

Authors:

A. Fatoba, C. Simpson; Univ. of Tennessee Hlth.Sci. Ctr., Memphis, TN

The association of celiac disease (CD) and autism spectrum disorder (ASD) remains inconclusive. Reports from different observational studies have become controversial, necessitating exploring the causal relationship between CD and ASD. This study used two-sample Mendelian randomization analysis to determine the causal link between CD and ASD. Summary-level data from a genome-wide association study (GWAS) of the European population was used to screen instrument variables (IVs) at genome-wide significance ($P < 5 \times 10\text{-}8$). The strength of instrument variables was also evaluated with F-statistics. Inverse variance weighted method (IVW) was the primary MR analysis, followed by sensitivity analyses such as pleiotropy, heterogeneity, and leave-one-out analysis. Our results showed no association between CD and ASD (OR, 0.997; 95% CI, 0.974-1.019; P = 0.770). There was also no evidence of horizontal pleiotropy (MR-Egger intercept = -0.001; P-value = 0.813) and heterogeneity (Q =16.187; P-value= 0.239). These results were also complemented by the leave-one-out analyses and scatterplot, which showed that none of the SNPs influenced the result. In conclusion, our MR study showed no causal relationship between CD and ASD.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1376 Causal effects of the plasma proteome on MRI-quantified liver fat within the UK Biobank.

Authors:

R. Smit[1,2,3], V. Diez-Obrero[1], S. D'Silva[1,4], R. Loos[1,2,3]; [1]Novo Nordisk Fndn. Ctr. for Basic Metabolic Res., Faculty of Hlth.and Med., Univ. of Copenhagen, Copenhagen, Denmark, [2]Charles Bronfman Inst. for Personalized Med., Icahn Sch. of Med. at Mount Sinai, New York, NY, [3]Dept. of Environmental Med. and Publ. Hlth., Icahn Sch. of Med. at Mount Sinai, New York, NY, [4]Ctr. for Cardiovascular Sci., Univ. of Edinburgh, Edinburgh, United Kingdom

Introduction: Excess liver fat is increasingly recognized as a sensitive proxy of cardiometabolic disease. By leveraging results from proteo-genomic studies through Mendelian randomization (MR), we aimed to gain insight into the potential druggability of liver fat. **Methods:** We employed a two-pronged approach for instrumental variable selection. First, we selected independent (LD r2<0.001) cis- and trans-pQTLs for 1,436 OLINK-measured proteins that attained genome-wide significance (p<5x10-8) in the discovery phase pQTL analyses (n=35,571 EUR) of the UK Biobank Pharma Proteomics Project (UKB-PPP). Separately, we restricted analyses to cis-pQTLs for 1,161 proteins with ≥1 primary cis-pQTL at a more stringent, protein-wide, genome-wide significance threshold (p<3.4x10-11) in discovery with nominal significance and directional consistency in replication (n=18,181). SNP-outcome weights were obtained by performing a GWAS on magnetic resonance imaging (MRI)-derived liver fat % in up to 34,308 UKB participants of White-British descent. To exclude confounding by LD, co-localization analyses were undertaken for proteins with cis-pQTL evidence of a causal effect. **Results:** In our cis/trans-pQTL combined analyses, the median per-protein number of independent instruments was 17, whereas only one protein had >1 instrument in our cis-pQTL only approach. Sufficiently strong instruments were available for both approaches (median F-statistic of 47 and 866, respectively). After FDR-correction, our combined cis/trans-pQTL approach identified Calsyntenin-1 (CLSTN1) and Heat Shock Protein beta-6 (HSPB6) as having a potential causal effect on liver fat % (random-effects IVW; CLSTN1: -0.48 (-0.69, -0.27) SD difference in liver fat % per doubling in protein concentration; HSPB6: -0.20 (-0.30, -0.11)). Pleiotropy-robust sensitivity analyses showed consistent results for HSPB6, but were not possible for CLSTN1 due to its small number of instruments (n=2). Separately, the cis-pQTL approach identified Apolipoprotein H (APOH), Neurocan (NCAN), and TNF receptor superfamily member 6B (TNFRSF6B) as having potential causal effects (Wald ratio; APOH: -0.07 (-0.10, -0.04); NCAN: -0.65 (-0.75, -0.56); TNFRSF6B: -0.12 (-0.17, -0.07)). However, co-localization analyses showed strong evidence of confounding by LD for NCAN (posterior probability for distinct causal variants near 1). **Conclusion:** With previous NAFLD-GWAS literature showing evidence for several of the loci underlying the currently identified putatively causal proteins, our results provide proteo-genomic evidence in favor of four potential drug targets aimed at modifying liver fat content.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1377 Characterization of telomere length in Alzheimer's disease in the Midwestern Amish.

Authors:

Y. Liu[1,2,3], A. Lynn[1,2], Y. Song[1,2], K. Miskimen[1], P. Wang[1,2], M. Prough[4,5], D. Dorfsman[4,5], A. Fernandez[4,5], R. Laux[1], S. Fuzzell[1], S. Hochstetler[1], A. Zaman[4], L. Adams[4], L. Caywood[4], J. Clouse[4], S. Herington[4], P. Whitehead[4], L. Main[6,2], P. Ogrocki[7,8], A. Lerner[7,8], J. Vance[4,5], M. Cuccaro[4,5], W. Scott[4,5], M. Pericak-Vance[4,5], J. Haines[1,2]; [1]Dept. of Population and Quantitative Hlth.Sci., Case Western Reserve Univ. Sch. of Med., Cleveland, OH, [2]Cleveland Inst. for Computational Biology, Case Western Reserve Univ., Cleveland, OH, [3]Dept. of Nutrition, Case Western Reserve Univ. Sch. of Med., Cleveland, OH, [4]John P. Hussman Inst. for Human Genomics, Univ. of Miami Miller Sch. of Med., Miami, FL, [5]The Dr. John T. Macdonald Fndn. Dept. of Human Genetics, Univ. of Miami Miller Sch. of Med., Miami, FL, [6]Dept. of Genetics and Genome Sci., Case Western Reserve Univ. Sch. of Med., Cleveland, OH, [7]Dept. of Neurology, Univ. Hosp. Cleveland Med. Ctr., Cleveland, OH, [8]Dept. of Neurology, Case Western Reserve Univ. Sch. of Med., Cleveland, OH

Telomere length (TL) is a candidate biomarker for aging and many aging-related diseases. To date, mixed results have been reported regarding the relationship between TL and Alzheimer's Disease (AD). As a result, the precise role of TL in AD remains uncertain. The Midwestern Amish population is a founder population demonstrating less genetic variation compared to the general population. There have been limited efforts devoted into characterizing TL in the Amish, in particular in relationship to age-related disorders such as AD. In this study, we aimed to characterize the TL from whole genome sequencing (WGS) and investigate the association between TL and both aging and cognitive function in the Amish. The study sample comprises 1,055 Amish individuals over age 60 residing in Indiana and Ohio, and their cognitive status was determined based on both adjudication from clinical experts and education-adjusted mini-mental status exam results (3MS) scores at the time of blood draw. Individuals were classified as cognitively unimpaired (CU, N=713) either due to adjudication or had a 3MS score ≥ 87. Individuals were classified as cognitively impaired (CI, N=228) either due to adjudication as mild cognitively impaired (MCI, N=87), AD Dementia (AD, N=102), or cognitively impaired but not AD (CINAD, N=25), or had a 3MS score < 87 (N=14). Blood WGS data was obtained via Illumina NovaSeq 6000 platform. TLs were calculated from WGS using Telseq (Ding et al., 2014). Linear regression models were used to study the effects of age, sex, *ApoE* genotype, smoking history, and medical history on TL. Association between TL and cognitive function was assessed using multivariate linear regression models, adjusting for age only. As expected, age showed a significant impact on TL (β = -0.02, p=4.21x10^{-08}), whereas no statistically significant effects were observed for other factors (sex: p=0.88, *ApoE*ε4: p=0.26, smoking: p=0.70, cancer: p=0.51, AMD: p=0.81, anxiety: p=0.95). Association between TL and education-adjusted 3MS score was not observed both before (β = 0.00, p=0.86) and after adjusting for age (β = -0.02, p=0.16). The distribution of TL between CI and CU groups was not significantly different both before (β = -0.02, p=0.60) and after adjusting for age (β = -0.08, p=0.09). A similar result was also seen between the more restrictive AD and CU comparison (β = -0.01, p=0.89, age-adjusted: β =-0.09, p=0.17). These data confirm the association of shorter TL with increasing age and provide the basis for additional studies of TL with multiple phenotypes in the Amish.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1378 Characterization of the genomic landscape of a rat model of Alcohol Use Disorder: Initial findings

Authors:

J. Calderon[1,2], G. Garrido[1], J. Ugalde[3], V. Jaramillo[1,2], F. Ezquer[1,2], Y. Israel[1,2]; [1]Univ. del Desarrollo, Santiago, Rm, Chile, [2]Res. Ctr. for the Dev. of Novel Therapeutic Alternatives for Alcohol Use Disorders, Santiago, Chile, [3]Univ. Natl. Andrés Bello, Santiago, Rm, Chile

Alcohol Use Disorder (AUD) is one of the most widespread causes of preventable diseases in the world and accounts for 4.2% of all disease burden measured in disability-adjusted life years. AUD has been associated primarily to social determinants but genetic associations to this trait have never been undisputedly identified and described. In an attempt to identify potential genes and variants underlying AUD we performed Whole Genome Sequencing (WGS) to two rat models: a set of low-alcohol drinking Wistar rats (UChA) and a set of high-drinking rats (UChB) that have been selectively bred for over 90 generations via brother-sister mating and subsequent phenotypic selection upon exposure to ethanol.
Genomic DNA (gDNA) from 3 high-drinking animals and 2 low-drinking animals were sequenced at BGI with DNBseq PE150 platform and approximately 38Gb of clean data were obtained per sample. We aligned these data to mRatBN7.2 reference genome and generated variant call files (VCFs) using GATK best practices workflow for variant discovery. The genomic landscape of high- and low-drinking animals has been characterized considering several different approaches. First, we investigated whether there were variants of significant functional impact on a subset of approximately 70 genes that have been previously associated to AUD. Secondly, we decided to inquiry whether there were other genes that carried variants that were not previously associated to AUD. We also performed Gene Ontology (GO) analysis to understand whether these animals had a non-random distribution of genetic variants in regard to any particular cellular process or structure. Our results show that UChA and UChB rats have significant differences in frequency and impact of variants associated with oxidoreductase and aldehyde catabolism activity, implying that an impaired metabolism of alcohol derived products could be product of selection over generations in our model. These results suggest that further characterization of the genomic architecture of our animal model of AUD could shed light on novel variants and/or genomic regions potentially implicated in the onset of this devastating phenotype

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1379 Characterizing BIN1K358R SNP rs138047593 effects in 5xFAD plaque pathology of Alzheimer's disease using snRNA-seq

Authors:

Z. Shi, L. Garcia-Agudo, S. Kawauchi, H. Liang, C. Da Cunha, G. Milinkeviciute, K. Tran, V. Swarup, A. J. Tenner, F. M. LaFerla, A. Mortazavi, G. R. MacGregor, K. Green; Univ. of California Irvine, Irvine, CA

Background: Recent genome-wide association studies (GWAS) in AD have identified more than 70 risk loci, including bridging integrator 1 (BIN1). Previous studies elucidated BIN1's basic role as an important regulator of endocytosis and membrane dynamics. But as an important late-onset Alzheimer's disease genetic risk locus with multiple isoforms expressed in the brain, determining the pivotal role of BIN1 in pathogenesis has been challenging due to conflicting findings and its presence in both neurons and glial cells. The presence of a mutated form (K358R) in BIN1 has been associated with heightened susceptibility to Late-Onset Alzheimer's Disease. Our study introduced this variation into mice to comprehend the effects of BIN1K358R risk variants on brain biology, used snRNA-seq to characterize its status in different celltypes, and assessed any resulting effects on plaque formation or subsequent impairments.
Method: The study involved the breeding of BIN1K358R mice with 5xFAD mice, generating distinct groups, WT, BIN1K358R homozygous, 5xFAD and 5xFAD/BIN1K358R mice. After aging all test subjects to 12 months old, snRNA-seq were performed on the prefrontal cortex to analyze BIN1-related celltype specific changes. We investigated on celltype subpopulation status, and inferred the subpopulation changes through pseudotime trajectory. Differential module eigengene analysis followed by gene ontology analysis was performed across genotypes to understand celltype changes with the presence of BIN1K358R in mice.
Result: Our results in functional enrichment analysis on the differentially expressed genes showed that the 5xFAD/BIN1K358R mice had enriched inflammatory response in microglia and attenuated activity of neurons and synaptic structure in oligodendrocytes at 12 months. Our disease module and network analysis further revealed both 5xFAD and 5xFAD/BIN1K358R mice had induced response of active glial cells but also differential co-expressed module eigengenes.
Conclusion: The presence of LOAD risk variant-carrying mice, coupled with comprehensive phenotyping and transcriptomic information across distinct age groups and pertinent AD pathologies, will significantly augment our knowledge of the fundamental biology of LOAD. This can potentially contribute to the creation of novel therapeutic approaches.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1380 Characterizing pleiotropic effects of schizophrenia and insomnia.

Authors:

M. Ewald[1], W. A. Houssain[1], O. Ponnuru[1], E. Young[1], O. Veatch[2]; [1]Kansas Univ. Med. Ctr., Kansas City, KS, [2]Univ. of Kansas Med. Ctr., Kansas City, KS

Schizophrenia is a complex neurodevelopmental condition defined by positive (e.g., delusions, hallucinations) and negative (e.g., apathy, lethargy) symptoms. Schizophrenia has no cure, requiring lifelong treatment. Many individuals affected by schizophrenia experience co-occurring sleep disturbances, like insomnia. Often, sleep issues precede positive and negative symptoms of schizophrenia. Sleep problems are associated with increased severity of daytime schizophrenia symptoms. It is possible that targeting and treating sleep problems could help reduce severity of other symptoms present in individuals with schizophrenia. Several schizophrenia risk genes overlap with genes implicated in sleep problems, suggesting pleiotropic effects. To search for central driving genes and factors with a role in co-occurring sleep problems in individuals with schizophrenia, more comprehensive investigations using evidence from multiple sources are necessary to elucidate these complex mechanisms, specifically pleiotropic effects. By characterizing these pleiotropic effects, we can develop more targeted treatment options instead of merely managing symptoms. To do this, we developed a bioinformatics pipeline to prioritize high confidence pleiotropic genes more likely to be functionally relevant. Briefly, the GWAS Catalog v1.0 was queried using the application program interface, API, to identify single nucleotide variants (SNVs) associated at a genome-wide significance ($p<5x10-8$) with both schizophrenia and insomnia. This search resulted in 15 SNVs which were then assigned to 12 unique genes based on positional, eQTL and 3D chromatin interaction mapping using FUMA GWAS v1.5.4. We then mapped human genes to mouse orthologs using DIOPT v9.0 and determined whether any potentially pleiotropic genes were associated with schizophrenia-relevant phenotypes, i.e., increased exploration and hyperactivity, in the International Mouse Phenotype Consortium database v1.5.0. There were two protein-coding genes, *Cnksr2* (connector enhancer of kinase suppressor of Ras 2) and *Gtpbp2* (GTP binding protein 2), where single gene knock out mice exhibited a schizophrenia-like phenotype. Both encoded proteins have been implicated in the altered development characteristic of epilepsy and neurodevelopment disorders. Upon querying the Pharos database v3.16.1 both proteins are currently categorized as TBio, meaning these proteins do not have any known drug target properties. Follow-up studies examining how these two genes interact together or independently are warranted.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1381 Characterizing the microbiomes of diseased and adjacent non-diseased ileum of patients with varying Crohn's disease severity.

Authors:

E. Ryu[1], W. Koltun[2], G. Yochum[2], E. Davenport[1]; [1]Pennsylvania State Univ., University Park, PA, [2]Pennsylvania State Coll. of Med., Hershey, PA

Crohn's disease (CD) is a chronic disease of inflammation in the gastrointestinal tract. Characterized by its patchy distribution, CD remains incurable due to limited understanding of its complicated pathogenesis involving both genetic and environmental components. One known contributing factor is the human gut microbiome, which has been demonstrated to play a vital role in CD. Prior research has primarily focused on characterizing the microbial signature of CD by comparing healthy to diseased individuals, identifying key microbial contributors driving the onset of CD, and determining important microbes that discern ulcerative colitis and CD. We still, however, lack a complete understanding of how the gut microbiome varies with severity of CD, and what microbes might contribute to the development of more severe CD. Furthermore, studies have often utilized fecal samples. While convenient, fecal samples represent an amalgamation of the microbes residing in the gut. CD generally manifests as patchy areas of inflammation, so it is even more important to specifically target the microbiome in an inflamed area vs. nearby non-inflamed tissue. To address these gaps, I will investigate the gut microbiomes of patients with either moderate (surgery > 6 years post-diagnosis) or severe (surgery < 6 years post-diagnosis) Crohn's disease. I will specifically be comparing the microbial communities located on inflamed and proximal non-inflamed areas of the mucosal lining of the ileum within each patient (n = 23 * 2 for moderate CD and n = 23 * 2 for severe CD; for a total n = 92). I will extract the microbiome from mucosal scrapings acquired from the Inflammatory Bowel Disease Tissue Biobank at the Penn State Milton S. Hershey Medical Center. This study will be one of the first to investigate the microbiomes of inflamed and proximal non-inflamed ilea as they vary with CD severity. As genetic factors have also been demonstrated to contribute to CD pathogenesis, the results will also allow us to subsequently pair microbiome findings with human gene expression data and better understand how host genes and microbes physiologically interact to influence Crohn's disease pathogenesis.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1382 Chronic Kidney Disease and Kidney Function Genome-Wide Association Study Reveals Population-Specific Associations.

Authors:

M. D'Antonio[1], W-Y. Park[1], W. Gonzalez Rivera[1], K. Frazer[2], M. Gymrek[3]; [1]Univ. of California, San Diego, La Jolla, CA, [2]UC San Diego, San Diego, CA, [3]Univ California San Diego, La Jolla, CA

The prevalence of chronic kidney disease (CKD) is increasing worldwide due to the rising rates of chronic conditions such as diabetes and hypertension, as well as the aging population. It is estimated that more than 15% of US adults have CKD. Prevalence of CKD is almost twice as high in African Americans (AFR) than individuals of European descent (EUR) that have hypertension as a comorbidity, suggesting the presence of ancestry-specific effects for this disease. Genome-wide association studies (GWAS) have been performed mostly on EUR individuals on CKD and kidney function, measured as estimated glomerular filtration rate (eGFR), and have identified hundreds of genome-wide significant loci. However, differences in the effects of genetic variation between populations on kidney function are currently unknown, rendering polygenic risk scores poorly transferrable across populations.To address this issue, we performed a GWAS on CKD and eGFR on >475,000 UK Biobank (UKBB) individuals divided into four populations: AFR, EUR, East Asians (EAS) and South Asians (SAS). Overall, we identified 417 loci, including 112 (26.9%) novel loci not previously in CKD and eGFR GWAS.We observed 14 loci in non-EUR populations, including 11 unique to non-EUR populations. There are two possible explanations as to why these loci were not found in the EUR population, which accounts for >90% of UKBB individuals: 1) the causal variant is rare in the EUR population; or 2) LD structure in these regions vary between populations, suggesting possible evolutionary differences between populations. Indeed, for three of the 11 signals, the non-EUR causal variant had an allele frequency <1%, and for seven of the remaining eight signals the LD structure was significantly different between populations.The most significant locus, which overlapped GATM, a gene well known to be associated with kidney function and disease, was shared across AFR, EUR and SAS, but each population had a different lead variant, only in moderate LD with each other. Fine mapping using Susie and simulations with haptools showed that, while there exists a certain overlap between the GWAS signals across the three populations, at least partial population-specific GWAS signals occur in the GATM locus.Our study shows the presence of multiple population-specific GWAS loci associated with kidney function, indicating a possible reason for the differences in CKD prevalence between EUR and AFR. Further studies on larger non-EUR populations are needed to improve the statistical power of GWAS and better characterize differences between populations, in order to improve the accuracy of polygenic risk scores in diverse and admixed individuals.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1383 Circulating metabolite levels associated with risks of psychiatric disorders: a Mendelian randomization study for biomarker and drug target identification.

Authors:

T. Lu[1], Y. Chen[2], S. Yoshiji[3], Y. Ilboudo[4], V. Forgetta[5], C. Greenwood[4], S. Zhou[2], B. Richards[3]; [1]Univ. of Toronto, Toronto, ON, Canada, [2]McGill Univ., Montréal, QC, Canada, [3]McGill Univ., Montreal, QC, Canada, [4]Jewish Gen. Hosp., Montreal, QC, Canada, [5]5 Prime Sci., Montreal, QC, Canada

Preventive measures and treatments for psychiatric disorders are limited. Circulating metabolites are potential candidates for biomarker and therapeutic target identification, because they play essential roles in biological processes, and their levels are measurable. Leveraging large-scale genome-wide association studies (GWASs) of psychiatric disorders, we performed Mendelian randomization (MR) analyses to assess the associations between circulating metabolite levels and the risks of bipolar disorder, major depressive disorder, and schizophrenia.
We conducted GWASs for circulating metabolite levels in the Canadian Longitudinal Study of Aging (N = 8,299 European ancestry individuals). To select genetic instruments while reducing the risk of horizontal pleiotropy, we combined gene expression profiles in GTEx and evidence in literature to identify variants which affect metabolite levels by altering expression of genes that are central to the metabolic pathways. MR analyses were performed, followed by multiple sensitivity analyses, colocalization analyses, and replication based on the UK Biobank, INTERVAL, and EPIC-Norfolk studies.
Low-pleiotropy genetic instruments were selected for 94 metabolites. After validating MR assumptions and colocalization evidence, we found that genetically increased circulating levels of n-3 unsaturated fatty acids, such as eicosapentaenoate (EPA) and docosapentaenoate (n-3 DPA), were associated with a decreased risk of bipolar disorder (maximum false discovery rate, FDR = 8.1×10^{-10}). Findings in replication cohorts confirmed the protective effects of increased n-3 unsaturated fatty acids level and n-3-to-total fatty acids ratio, as well as a decreased n-6-to-n-3 ratio. Genetically predicted circulating level of N-alpha-acetylornithine was associated with increased risk of schizophrenia with an odds ratio (OR) of 1.31 (95% CI: 1.18-1.44; FDR = 6.3×10^{-6}) per one standard deviation increase. Furthermore, a one standard deviation increase in genetically predicted circulating level of hypotaurine, a key intermediate in taurine biosynthesis, was associated with an OR of 0.85 (95% CI: 0.78-0.93; FDR = 5.5×10^{-3}) for major depressive disorder.
In conclusion, through a metabologenomics-guided MR, we have validated n-3 unsaturated fatty acids as a protective risk factor against bipolar disorder. We have identified N-alpha-acetylornithine as a candidate biomarker for schizophrenia, and hypotaurine as a candidate biomarker and possible therapeutic target for major depressive disorder. These findings encourage further explorations of these metabolites in the context of psychiatric disorders.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1384 Classification of Variants in Systemic Autoinflammatory Diseases using 3D Structures and Protein-protein Interactions

Authors:

A. Bülbül[1], E. Timucin[1], A. C. Timucin[2], **U. Sezerman**[1], E. Tahir Turanli[2]; [1]Acibadem Univ., Graduate Sch. of Hlth.Sci., Dept. of Biostatistics and Bioinformatics, Istanbul, Turkey, [2]Acibadem Univ., Faculty of Engineering and Natural Sci., Dept. of Molecular Biology and Genetics, Istanbul, Turkey

Systemic Autoinflammatory diseases (SAIDs) are a group of diseases, mainly of innate immunity, that have high heterogeneity at both clinical and genetic level. Depending on the inheritance of pattern of the disorder (monogenic or polygenic) approximately 20-50% of the time. Moreover, 420 of identified variants fall into the category of VUS (variants of uncertain significance).The efficacy of in silico variant prediction tools varies depending on the specific type of genetic variation and the functionality of the protein in question. The aim of our study is to utilize protein-protein interactions (PPI), 3D analysis, and to implement machine learning techniques for pathogenicity prediction of variants of SAIDs genes. We analyzed 702 missense variants from 34 genes such as *ADA2*, *IL10*, *IL10RA*, *MEFV*, and *MVK* implicated in SAIDs, as sourced from the Infevers database (Available at: https://infevers.umai-montpellier.fr. Accessed on 20.04.2022). Among these variants, 572 were pathogenic, 130 were benign.We computed ddG values. ZDOCK and SPRINT scores from 3D protein docking and sequence based PPI networks in STRING and IntAct databases. Scores were normalized using HGPEC rank scores. Additional features of PSIC, ASA, BLOSUM62 scores as well as volume and alpha helix propensity changes of variants were also considered. To overcome the sample imbalance issue, we utilized the SMOTE algorithm. For predictive modeling, we employed the Random Forest algorithm, optimized using Grid Search.
According to the model, which is trained on 80% splitted data, results obtained from an average of 20 split cross validations, we achieved a ROC AUC value of 984%, an F1 value of 0.955, precision of 0.968, recall of 0.948, accuracy of 0.956, and balanced accuracy of 0.955.. Our prediction method exhibited superior performance compared to commonly used mutation prediction methods, namely SIFT, PolyPhen, and CADD. Evaluation on a small test set of 59 variants found in the exons 2 and 10 of the *MEFV* gene showed that our model outperformed all of the predictors achieving an accuracy of 0.864 while the performance of the other predictors ranged in between 0.694 to 0.813. The prediction results for an additional set of 420 VUS missense variants is available at https://github.com/(autor)/Variant_Prediction_Tool. The implemented pathogenicity prediction strategy is particularly valuable in monogenic diseases and proves promising for complex diseases when genes are considered according to disease subgroups.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1385 Clinical application of polygenic risk scores in obesity

Authors:

N. Vos, J. van Uhm, E. Bulthuis, L. kleinendorst, M. Mannens, P. Jansen, M. van Haelst; Amsterdam UMC, Amsterdam, Netherlands

Introduction: In recent years, significant progress has been made in unravelling both polygenic and monogenic factors in obesity. Developments in polygenic risk score (PRS) analyses have emerged as a promising strategy in predicting obesity risk and understanding the variable penetrance of monogenic mutations. Despite this, translation and implementation of PRS into clinical settings remain limited.
Methods: Our current study aims to evaluate the feasibility, utility and predictive value of PRS in genetic diagnostics in real-world settings of a clinical obesity cohort. We include patients referred to the clinical genetics department (Amsterdam UMC) for suspected genetic obesity (n=83), and those with chromosome 16p11.2 deletions (n=10), of which polygenic susceptibility are compared with the general population (GO-NL cohort, n=498). PRS are extracted from clinical SNP arrays and calculated based on the largest available BMI GWAS (N= 681,275). We tested whether: 1) our clinical cohort differs in polygenic susceptibility for obesity compared to the general population; 2) PRS can aid in identifying patients at the highest risk of carrying a monogenic mutation for obesity.
Results: We show high feasibility to extract PRS from clinical SNP arrays in a high throughput manner. PRS of obesity was significantly higher in the obesity cohort ($P<0.001$), but not 16p11.2 deletion cohort, compared to the general population, suggesting a strong polygenic contribution to clinically referred patients. No such effect was observed for other non-obesity PRS. We found significantly lower PRS in patients with obesity and a mutation in an obesity gene, (suspected) monogenic obesity, compared to those in which no mutation is observed ($P=0.015$-0.049, depending on threshold), with PRS showing an area under the curve (AUC) of 0.73 for predicting mutation status.
Conclusions: We show that PRS explains the obesity phenotype in a considerable fraction of patients suspected of monogenic obesity. Our real-world results show that PRS aids in identifying patients with the highest probability of finding a monogenic form of obesity, which has important implications for the clinical use of PRS in genetic obesity.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1386 Clinical utility of exome sequencing in pediatric patients with extracorporeal membrane oxygenation for refractory lung disease

Authors:

J-H. Wu[1,2], E-T. Wu[1], C-C. Wang[1], F. Lu[1], H-W. Chou[3], C. Yih-Sharng[3], S-C. Huang[3], N-C. Lee[4]; [1]Dept. of Pediatrics, Natl. Taiwan Univ. Hosp., Taipei, Taiwan, [2]Dept. of Med., Natl. Taiwan Univ. Hosp. Jinshan branch, New Taipei City, Taiwan, [3]Dept. of Surgery, Natl. Taiwan Univ. Hosp. and Coll. of Med., Taipei, Taiwan, [4]Dept. of Pediatrics and Med. Genetics, Natl. Taiwan Univ. Hosp., Taipei, Taiwan

Purpose: Pediatric refractory lung disease is rare, has a broad differential diagnosis, and results in high morbidity and mortality even under extracorporeal membrane oxygenation (ECMO) support. The etiologies for these rare pediatric end-stage pulmonary conditions include lethal alveolar capillary dysplasia with misalignment of the pulmonary veins (ACD-MPV) and childhood interstitial lung disease which progress rapidly or chronically and require ECMO for bridge to lung transplants. In contrast to pathological diagnosis by invasive procedure, exome sequencing has reported timely diagnostic yields about 52.5% for pediatric patients in intensive care units. We aimed to characterize the clinical utility of exome sequencing in pediatric patients with extracorporeal membrane oxygenation for refractory lung disease. Methods: Children (age below 18 years) who required ECMO support for refractory lung disease at a tertiary children's hospital were recruited from 2017 to 2022. Prematurity and patients with complex congenital heart diseases were excluded. Demographic and diagnostic information was obtained from medical records. A standard whole-exome sequencing platform with text mining-assisted variant prioritization was applied to selected patients for genetic diagnosis. Results: A total of 30 (19 neonates and 11 children) patients under ECMO for refractory lung disease during 2017 to 2022 were recruited with mortality rate of 27% (8/30) and the diagnoses as follows: ten congenital diaphragmatic hernia, nine persistent pulmonary hypertension (PPHN), six lung infection, three pulmonary fibrosis, one asthma, and one trauma. Three children with pulmonary fibrosis and one neonate with refractory PPHN received exome sequencing and a molecular diagnosis was identified in 3 patients (67%) including a de novo heterozygous mutation in *FOXF1*, a hemizygous mutation in *DKC1* and a homozygous mutation in *SFTPC.* After diagnosis, all three patients were prepared for lung transplant but died in the prolonged waiting time due to limited organ donation for these young children. On the other hand, the remaining young teenager without genetic diagnosis received living-donor lung transplantation and recovered smoothly. Conclusions: Given the broad differential diagnosis and critical status of pediatric patients with ECMO for refractory lung disease, exome sequencing provides timely diagnoses which helps clinicians in shared decision-making with parents. However, hospice care for patients with lethal genetic disease, especially young children with limited organ donation, is an another choice.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1387 Clustered Mendelian randomization analyses identify distinct biological mechanisms that link adiposity to type 2 diabetes: towards precision medicine.

Authors:

H. Yaghootkar[1], M. Cule[2], M. Thanaj[3], N. Basty[3], A. Hashemloo[4], E. Sorokin[5], B. Whitcher[3], S. Burgess[6], J. Bell[3], N. Sattar[7], E. Thomas[1], A. Abraham[1]; [1]Univ. of Lincoln, Lincoln, United Kingdom, [2]Calico Life Sci., San Francisco, CA, [3]Univ. of Westminster, London, United Kingdom, [4]Brunel Univ., London, United Kingdom, [5]Calico Life Sci. LLC, South San Francisco, CA, [6]Univ. of Cambridge, Cambridge, United Kingdom, [7]Univ. of Glasgow, Glasgow, United Kingdom

Understanding the mechanisms that link adiposity to type 2 diabetes is challenging due to their complex nature and the heterogeneity of the conditions and individuals affected by them. Individuals with similar adiposity levels can exhibit different risks for cardiometabolic diseases. This study aims to identify genetic subtypes of adiposity and characterize their metabolic signature, specifically focusing on their impact on type 2 diabetes and its complications. We utilized 388 genetic variants associated with body fat percentage ($p<5E-8$) and assessed their effect on type 2 diabetes to identify distinct clusters representing different causal pathways using MR-clust, a statistical method for clustering variants based on their associations with an outcome trait. We then characterized the metabolic signature of each cluster based on various markers using published genome-wide association studies. We further analyzed MRI scans from the UK Biobank (in 40,000 individuals) to compare the impact of these clusters on measures of fat distribution and muscle quality. Finally, using FinnGen, we estimated the causal effect of each cluster on type 2 diabetes and its complications. Our findings revealed the identification of three clusters of alleles associated with both higher adiposity and higher type 2 diabetes risk (odds ratio [95% confidence intervals] per 1-standard deviation higher body fat percentage 11.20 [6.90,18.21]; 4.42 [3.72,5.25]; 1.41 [1.07,1.86]). Additionally, two clusters were associated with higher adiposity but lower type 2 diabetes risk (0.29 [0.18,0.48]; 0.05 [0.03,0.08]). All clusters exhibited associations with higher childhood obesity, body mass index, leptin, C-reactive protein, subcutaneous fat (abdominal and thigh), pancreas fat, and muscle fat. However, the associations with liver fat, lipids, insulin resistance, liver enzymes, adiponectin, and pro- and anti-inflammatory cytokines demonstrated opposing directions, with type 2 diabetes-decreasing clusters showing a favorable effect. Higher adiposity through type 2 diabetes-increasing clusters was causally associated with higher risk of all complications, while higher adiposity through type 2 diabetes-decreasing clusters was causally associated with lower risk of cardiometabolic complications but still higher risk of thrombotic events and osteoarthritis. These findings provide further support for adiposity-increasing mechanisms that may serve as protection against type 2 diabetes and its complications, particularly when accompanied by decreased ectopic liver fat. This may aid future optimisation of prediction, prevention, and treatment of type 2 diabetes.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1388 Coding and non-coding sequence variation across keratoconic cone surface

Authors:

M. Gajecka[1,2], M. Maleszka-Kurpiel[3,4], M. Kabza[1], J. Karolak[1], K. Jaskiewicz[2]; [1]Chair and Dept. of Genetics and Pharmaceutical Microbiol., Poznan Univ. of Med. Sci., Poznan, Poland, [2]Inst. of Human Genetics, Polish Academy of Sci., Poznan, Poland, [3]Chair of Ophthalmology and Optometry, Poznan Univ. of Med. Sci., Poznan, Poland, [4]Optegra Eye Hlth.Care Clinic, Poznan, Poland

Background: Keratoconus (KTCN) has been originally described by ophthalmologists as a non-inflammatory eye disease with multifaceted etiology. Although reported transcriptomic and proteomic findings suggested inflammatory aspects in KTCN, genetic features relating to immune system function in KTCN have not yet been demonstrated. Here, coding and non-coding sequence variation in the corneal epithelium (CE) across the keratoconic cone surface in patients with KTCN and its relevance in the functioning of the immune system were assessed.

Materials and methods: We performed WGS in an experimental model consisting of three CE topographic regions (TRs), *central*, *middle*, and *peripheral*, derived from four unrelated adolescent patients with KTCN and two adequate controls (Σ=18 experimental samples). The RNA/DNA/protein samples were extracted simultaneously (RNA/DNA/Protein Purification Plus MicroKit, Norgen Biotek), and after library preparation (TruSeq Nano DNA HT LibraryPrep Kit, Illumina), were sequenced with 30X coverage. SNP and indel calling was performed using Platypus, and detected variants were annotated with Ensembl Variant Effect Predictor software. Next, the recognized variants were evaluated together with the previously reported transcriptomic outcomes for the same CE samples, and the full-thickness corneas.

Results: Pathway enrichment analysis of genes with identified coding variants pointed to 'Antigen presentation' and 'Interferon alpha/beta signaling'. Both coding and non-coding sequence variants were found in genes (or in their close proximity) linked to the previously revealed KTCN-specific cellular components as 'Actin cytoskeleton', Extracellular matrix', and 'Collagen-containing extracellular matrix', and pathways as 'Focal adhesion', 'Hippo signaling pathway', and 'Wnt singling'. No genomic heterogeneity across the corneal surface was found comparing the assessed TRs. 35 chromosomal regions enriched in KTCN-specific sequence variants were revealed, with a most representative 5q locus, previously recognized as involved in KTCN.

Conclusions: Identified variants in non-coding regions of the genome further emphasized the genetic heterogeneity in KTCN and complemented the KTCN-specific sequence variation previously identified in ES. Moreover, the identified genomic features pointed to the involvement of innate and adaptive immune system responses in KTCN pathogenesis.

Support: The National Science Centre grant no. 2018/31/B/NZ5/03280.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1389 Co-expression of subgenual anterior cingulate cortex transcriptomehighlights translational modifications of synaptic signaling during early development in Schizophrenia.

Authors:

R. Smith[1,2], A. Mihalik[2], N. Akula[1], A. Schulmann[1], S. Marenco[3], P. K. Auluck[3], E. Bullmore[2], A. Raznahan[1], P. Vertes[2], F. J. McMahon[1]; [1]Human Genetics Branch, Natl. Inst. of Mental Hlth., NIH, Bethesda, MD, [2]Brain Mapping Unit, Dept. of Psychiatry, Univ. of Cambridge, Cambridge, United Kingdom, [3]Clinical Brain Disorders Branch, Intramural Res. Program, Natl. Inst. of Mental Hlth., Bethesda, MD

Weighted gene co-expression network analysis (WCGNA) applied to postmortem brain tissue is useful in identifying functional gene modules that are differentially regulated in neuropsychiatric disorders. However, current analysis techniques linking modules to diagnosis are limited by inability to 1) account for full variation in gene expression data and 2) quantify the influence of individual features and covariates. To address these issues, we performed gene co-expression followed by multivariate canonical correlation analysis. Total RNA was extracted from the sgACC of 185 human postmortem brain samples (55 controls, 44 schizophrenia (SCZ), 35 bipolar disorder (BD), and 51 major depressive disorder (MDD)) and sequenced. Gene co-expresssion was perfomed using WGCNA, and each module was functionally annotated. To establish relationships between modules and diagnostic groups at the univariate level, linear regression was run on each module, with the outcome variable as the first principal component (module eigengene) and the dependent variable as diagnostic group with and without drug covariates. To quantify the module-diagnosis-drug relationships and the importance of specific genes, Group Regularized Canonical Correlation Analysis (GRCCA) was run on the full expression dataset. At the gene level, 23 functionally significant co-expression modules were identified. Topic modeling of these modules highlights synaptic signaling, immune process, cellular organization and transport among others. Linear regression (without covariates) identified a link between SCZ and modules related to immune and inflammatory processes. However, post-hoc analyses with the inclusion of drug covariates suggested a strong influence of drug use. GRCCA run with drug covariates in the model matrix identified genes in modules related to translation, synaptic signaling, inflammatory response, and mitochondrial activity as robust drivers of a relationship with SCZ. No robust link was identified between BD and MDD and gene expression. The functional annotation of co-expression modules followed by GRCCA suggests a strong role of transcription and translation processes related to synaptic signaling in early development of SCZ. This key result corroborates what has been reported in the literature but provides an additional dimension by ensuring the effect is due to SCZ and not drug use. In conclusion, multivariate results are more specific than results obtained using traditional univariate methods applied to WGCNA, suggesting that the multivariate GRCCA is the key in identifying functional links between psychiatric disorders and gene co-expression networks.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1390 Colocalization and Mendelian randomization unveil interactions between eGFR, lifestyle, molecular pathways, and diseases for improved CKD management

Authors:

M. Masala, M. Pala; Consiglio Nazionale delle Ricerche, Monserrato, Italy

Chronic Kidney Disease (CKD) impacts more than 10% of adults in developed countries, constituting a substantial global health concern. CKD is linked to various health issues such as hypertension, diabetes, cardiovascular disease, and bone disorders. However, the molecular causes of CKD, as well as whether CKD is an effect or a cause of these related health conditions, remain unclear. A key indicator of CKD is a reduction in the glomerular filtration rate (GFR), estimable via serum creatinine or cystatin levels, which serve as widely accepted markers of kidney function.Recent genome-wide association studies (GWAS) on eGFR have encompassed over a million individuals, revealing over 400 associated loci. However, comprehensive characterization of these loci necessitates post-GWAS analyses, such as colocalization and Mendelian Randomization, which have thus far been limited to a restricted number of traits and diseases.To gain a deeper understanding of the complex interplay between eGFR, lifestyle factors, molecular pathways, and disease susceptibilities, we propose creating a platform that includes standardized GWAS and QTLs summary statistics for over 40,000 diseases and traits. This platform would enable the rapid evaluation of millions of GWAS combinations for standard post-GWAS analysis. The summary statistics include molecular traits (such as RNA, plasma proteins, and surface receptor proteins), biochemical measurements, immunophenotypes, lifestyle information, and diseases.Our initial findings have identified colocalization of eGFR (with a posterior probability H4 > 0.8) at 192 loci with 158 diverse quantitative traits or diseases. These include anthropometric measurements, physiological markers, lifestyle factors, immunophenotypes, and various diseases, such as Alzheimer's, type 2 diabetes, diverse cancers, cardiovascular diseases, and autoimmune conditions.Our systematic post-GWAS analysis will provide a comprehensive catalog of traits and diseases that have a genetic correlation with eGFR. It will offer scores for causal inference and could be used to suggest potential therapeutic targets, inform therapeutic strategies (for example, which pathways should be inhibited), and identify potential side effects of therapeutic intervention.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1391 Combined effects of genetics and oral contraceptive use on the risk of venous thromboembolism.

Authors:

V. Lo Faro[1], T. Johansson[2], Å. Johansson[1]; [1]Uppsala Univ., Uppsala, Sweden, [2]Immunology, genetics and pathology, Uppsala, Sweden

Background. Over 150 million women worldwide use oral contraceptives, and the risk of venous thromboembolism (VTE) is increased in oral contraceptive users. Women with inherited thrombophilia and carriers of certain thrombophilia gene mutations, such as factor V Leiden (FVL) and prothrombin mutation (PTM), are at increased risk of VTE, especially in combination with oral contraceptive use. VTE is a complex disorder involving many genetic risk factors and recently polygenic risk scores (PRS) have been proposed to capture a significant proportion of the genetic risk of VTE. The aim of this study is to estimate the risk of VTE at initiation and during continued use of oral contraceptives in women with a high genetic liability.

Methods. We used a prospective study design in which 244,420 participants from the UK Biobank were followed from birth. The effect of oral contraceptive use initiation and continuation on VTE risk was estimated using Cox regression with a time-dependent exposure variable. Women were stratified according to their PRS and whether they were carriers of FVL and/or PTM.

Results. When genetic risk was not considered, an increased risk of VTE was observed during the first two years of oral contraceptive use (HR=3.09; 95% CI = 3.00 - 3.20), but not during continued use (HR=0.92;95% CI, 0.80 - 1.05). However, when genetic risk was considered, women with the highest PRS risk category had a more pronounced increased risk of VTE associated with initiation of oral contraceptive use (HR=7.27; 95% CI, 6.61 - 7.99), and a similar risk was also seen in FVL (HR, 8.29 [95% CI, 7.40 - 9.28]) and PTM carriers (HR, 11.61 [95% CI, 9.76 - 13.81]). A high PRS in combination with FVL and PTM carriers resulted in the highest risk of VTE at the initiation of oral contraceptive use (HR, 38.14 [95% CI, 22.7333 - 64.01]). Women with a high genetic liability also had an increased risk during continued use, with the highest risk in women with a high PRS in combination with carriers of both FVL and PTM (HR, 7.14 [95% CI, 4.33 - 11.76]).

Conclusions. Polygenic risk can capture additional VTE risk that is not included in the commonly investigated genes for inherited thrombophilia. Our results indicate that oral contraceptive use is associated with an increased risk of VTE, particularly in women with a high genetic predisposition. Finally, our results also show that oral contraceptive use dramatically increases the risk of VTE just before starting use, but that this risk gradually decreases with continued use.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1392 † Combining GWAS and tissue-specific multiomic data to identify candidate human height genes.

Authors:

C. Fadil[1,2], A. U. Eliasen[3,4,5,6], E. Bartell[3,4,7], S. Sakaue[4,8,9], L. Yengo[10], S. Vedantam[3,4], E. Marouli[11], A. Craft[12,13,14], T. Capellini[15,16], Y. Okada[17], A. R. Wood[18], J. N. Hirschhorn[3,4,7], GIANT consortium height working group; [1]Div. of Med. Sci., Harvard Univeristy, Boston, MA, [2]Graduate Sch. of Arts and Sci., Harvard Univ., Boston, MA, [3]Dept. of Pediatrics, Div. of Endocrinology, Boston Children's Hosp. and Harvard Med. Sch., Boston, MA, [4]Program in Med. and Population Genetics, Broad Inst. of MIT and Harvard, Cambridge, MA, [5]COPSAC, Copenhagen Prospective Studies on Asthma in Childhood, Herlev and Gentofte Hosp., Univ. of Copenhagen, Copenhagen, Denmark, [6]Dept. of Bio and Hlth.Informatics, Technical Univ. of Denmark, Kongens Lyngby, Denmark, [7]Dept. of Genetics, Harvard Med. Sch., Boston, MA, [8]Ctr. for Data Sci., Brigham and Women's Hosp., Harvard Med. Sch., Boston, MA, [9]Div.s of Genetics and Rheumatology, Dept. of Med., Brigham and Women's Hosp., Harvard Med. Sch., Boston, MA, [10]Inst. for Molecular BioSci., The Univ. of Queensland, Brisbane, Australia, [11]William Harvey Res. Inst., Barts and the London Sch. of Med. and Dentistry, Queen Mary Univ. of London, London, United Kingdom, [12]Dept. of Orthopedic Res., Boston Children's Hosp., Boston, MA, [13]Dept. of Orthopedic Surgery, Harvard Med. Sch., Boston, MA, [14]Harvard Stem Cell Inst., Cambridge, MA, [15]Dept. of Human Evolutionary Biology, Harvard Univ., Cambridge, MA, [16]Broad Inst. of MIT and Harvard, Cambridge, MA, [17]Dept. of Statistical Genetics, Osaka Univ. Graduate Sch. of Med., Osaka, Japan, [18]Genetics of Complex Traits, Coll. of Med. and Hlth., Univ. of Exeter, Exeter, United Kingdom

The largest genome-wide association study (GWAS) on adult height conducted by the GIANT consortium in 5.4 million individuals of diverse ancestries identified 12,111 independent genetic variants associated with height (Yengo et al., 2022). These span 21% of the human genome and together explain over 90% of SNP-based heritability. Despite this significant coverage, the specific mechanisms through which these regions affect human height remain largely unknown. The aim of this study is to gain deeper insights into the role of these genetic variants in the biology of skeletal growth by leveraging multiple independent lines of genetic and genomic evidence.
We applied the newly-developed method SCENT (Single-Cell ENhancer Target gene mapping) to predict enhancer-gene pairs from multimodal single-cell data of 8 cell types (Sakaue et al., 2022). Of the 12,111 lead COJO variants reported by the GIANT study, 181 SNPs mapped to the regulatory regions of 267 genes.
We performed fine-mapping with SuSIE to identify putative causal variants associated with height in UK Biobank individuals of White British ancestry. Among the 2881 finemapped credible sets, 302 overlapped with predicted gene regulatory regions, implicating a total of 518 genes. Of these, 31 genes were implicated by credible sets containing only one SNP. Only 24 of the 518 genes were previously known to be functionally implicated in height (as part of the OMIM gene set, Yengo et al. 2022). As a complementary line of evidence, we used a combination of gene prioritization methods (PoPS, MAGMA, and DEPICT) to generate a shortlist of high-confidence height genes. Collectively, we identified 11 genes that were implicated by both enhancer-gene mapping and at least one gene prioritization method. Notably, 3 genes (STAT6, CTDSP2 and ARID3A) were implicated by all methods. We are extending this pipeline to human fetal distal femur tissues, in order to infer enhancer-gene pairs implicated in human skeletal development. Preliminary analysis of single-cell ATAC-seq data across 9 cell types shows significant enrichment of accessible chromatin regions for height heritability compared to other traits.
Using a combination of computational approaches combined with tissue-specific multiomic data we were able to implicate through converging lines of evidence a set of high-confidence height genes for further functional follow-up.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1393 Combining proteomics and metabolomics to identify signatures protective of neurological consequences of post-acute SARS-CoV-2 infection.

Authors:

C-Y. Su, S. Zhou; McGill Univ., Montréal, QC, Canada

Introduction: Post-acute sequelae SARS-CoV-2 (PASC) is an emerging public health concern with heterogeneous manifestations. Among PASC symptoms, long-term neurological manifestations are particularly debilitating. Their molecular underpinning remains poorly understood. Here, we combined proteomics and metabolomics to address this issue. **Methods:** Proteomic and metabolomic data were generated for 1,234 individuals from the Biobanque Québécoise de la COVID-19 (BQC19), which followed COVID-19 patients over two years. After quality control, 4,984 proteins and 943 metabolites were retained for analysis. Using 689 clinical entries from the BQC19, we defined broad neurological PASC (BNP) as any neurological manifestations, whereas refined neurological PASC (RNP) was defined as neurological complications that either recently appeared or deteriorated due to COVID-19. We then fitted single biomarker logistic regression models while adjusting for age at diagnosis and sex to determine the association of each protein or metabolite with BNP or RNP. **Results:** A total of 381 and 65 cases with both proteomics and metabolomics data met our definition of BNP and RNP and were compared with 535 and 713 controls, respectively. A total of 1,416 and 734 proteins were found significantly associated with BNP and RNP, respectively, with 544 overlapping proteins. In addition, 158 and 130 metabolites were found significantly associated with BNP and RNP, respectively, with 53 overlapping metabolites. In particular, we identified phosphatidylethanolamine-binding protein 1 (PEBP1) as being protective against BNP (OR (95% CI) = 0.76 (0.66-0.87), FDR p = 0.002) and RNP (OR (95% CI) = 0.61 (0.46-0.81), FDR p = 0.014). Interestingly enough, previous studies suggested downregulation of PEBP1 may lead to Alzheimer's disease. Moreover, PEBP1 functions as an enzyme for 1-palmitoyl-2-docosahexaenoyl-GPE, a phosphatidylethanolamine (PE) whose impairment may lead to neurodegenerative disorders, which is concordant with our finding that increased circulating PE is associated with decreased neurological PASC risk (BNP: OR (95% CI) = 0.78 (0.68-0.90), FDR p = 0.01; RNP: OR (95% CI) = 0.56 (0.41-0.75), FDR p = 0.006). **Conclusion:** Here, we used an integrative bi-omics approach combining proteomics and metabolomics to provide new insight into the pathophysiological mechanisms underlying neurological PASC risk which, possibly, could point to new ways to treat this condition.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1394 Combining research in humans and flies to uncover the genetic basis of ageing: A Transcriptome-wide Association Study identifies putative ageing genes in humans with experimental validation in flies.

Authors:

G. Navoly[1], A. Alizan[2], L. Partridge[2], N. Alic[2], K. Kuchenbaecker[1,3]; [1]UCL, Genetics Inst., London, United Kingdom, [2]Inst. of Hlth.y Ageing, UCL, London, United Kingdom, [3]Div. of Psychiatry, UCL, London, United Kingdom

Ageing and age-related diseases represent a major challenge for public health systems worldwide. While we have a solid grasp of the genetic architecture of ageing in model organisms, our understanding of human ageing is limited to a few candidate genes. To better connect research in model organisms and humans, we first performed a multi-tissue Transcriptome-wide association study (TWAS) on human lifespan, using >1 million parental lifespans from the UK Biobank and the LifeGen Consortium; followed by validation TWAS on two different ageing outcomes: Longevity and Healthspan. We also performed TWAS fine-mapping, to identify putative causal genes. We tested the role in ageing of the fly orthologues of four of the candidate genes: *COASY, MAD2L, SH3PXD2A* and *TOMM40*, by modulating their expression in specific tissues via RNA interference or over-expression. Lifespan assays were then carried out to assess the effects the genes have on *Drosophila l*ongevity. TWAS uncovered 563 significant gene associations, of which 139 genes were replicating in Longevity or Healthspan outcomes (replication threshold p<0.05). *TOMM40*, a gene that has an important role in maintaining mitochondrial function, had the strongest association with Parental lifespan (p=2.79E-66) and Longevity (p=1.30E-45) at the *APOE-TOMM40* region. *TOMM40* was successfully fine-mapped as the putatively causal lifespan gene, yielding a higher posterior inclusion probability (PIP) than *APOE* ($TOMM40^{PIP}$=0.99 vs. $APOE^{PIP}$=0.2). Using TWAS, we identified 9 novel genes associated with all the three ageing outcomes tested, including *TOMM40, HTR3B* and *COASY*. *COASY* plays an important role in synthetic and degradative metabolic pathways. Mutations in this gene have been associated with cognitive decline and neurodegenerative diseases such as Alzheimer's disease and Parkinson's disease, which are strongly associated with ageing. Knocking down the fly orthologue of *COASY, Ppat-dpck*, and *SH3PXD2A, Cindr*, resulted in significant lifespan extension in flies, validating their role in ageing across organisms. In conclusion, TWAS uncovered significantly more associations with Parental lifespan than the GWAS. Using multiple ageing outcomes may yield more robust results and highlight genes involved in fundamental ageing processes that are conserved across species. Our results, highlighting the role of fly orthologue of *COASY, Ppat-dpck*, and of *SH3PXD2A, Cindr* in ageing in humans and flies show that TWAS can be an important tool in investigating the transferability of findings from model organisms to humans, as it identifies putative genes that can be validated experimentally.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1395 Common and rare variant contributions to genomic architecture of atrial fibrillation using array-based genotyping and whole-genome sequencing data.

Authors:

K. Miyazawa[1], K. Ito[1], Y. Takanori[2], R. Kurosawa[1], H. Yoshida[1,3], N. Enzan[1,4], F. Takechi[1,5], M. Fukuyama[1,3], R. Osako[1,2], M. Koido[3,1], Y. Momozawa[1], S. Nomura[3], H. Akazawa[3], N. Koichi[2], Y. Kamatani[3,1], C. Terao[1], I. Komuro[3,6]; [1]RIKEN Ctr. for Integrative Med. Sci., Yokohama, Japan, [2]Saga Univ., Saga, Japan, [3]The Univ. of Tokyo, Tokyo, Japan, [4]Kyusyu Univ., Fukuoka, Japan, [5]Chiba Univ., Chiba, Japan, [6]Intl. Univ. of Hlth.and Welfare, Tokyo, Japan

Genome-wide association studies (GWAS) have successfully identified common-variant associations with atrial fibrillation (AF), while the proportion of phenotypic variation captured by common variants is substantially less than the estimates from pedigree-based analyses. To address this remaining heritability, whole-exome and whole-genome sequencing studies have examined rare-variant associations, but the contribution of rare genetic variation to AF risk is not well characterized. Furthermore, the relationship in genomic architecture of common and rare variants remains unexplored. Here, we performed GWAS using imputed array-based genotype data and whole-genome sequence data, respectively, and assessed the contributions of common variants and rare variants to AF genetics. First, a cross-ancestry meta-analysis of large-scale AF-GWAS comprising 116,532 AF cases and 1,338,446 controls identified 76 novel genetic loci among 223 genome-wide significant associated loci. We then conducted transcriptome-wide and proteome-wide association studies leveraging the models of expression and protein quantitative trait loci, respectively. These studies implicated 172 genes and 10 proteins associated with AF, and in particular we found that IL6R revealed a significant association in both studies. Next, we analyzed the whole-genome sequence data of two datasets from 547 patients with early-onset AF (less than 46 years of age) extracted from BioBank Japan and 315 patients who underwent catheter ablation for AF, and explored the consequences of rare variants using the LOFTEE plug-in implemented in the Variant Effect Predictor. Some genes with rare loss-of-function (LOF) variants were identified in both datasets including *TTN*, whereas other genes with rare LOF variants were detected only in individual dataset. Then, we performed enrichment analysis of LOF variants to identify genes associated with early-onset AF. Moreover, we conducted a gene-based burden analysis using rare LOF variants in a case-control study including control samples extracted from whole-genome sequencing data of 7,825 samples. Finally, we applied a polygenic risk score derived from a previous cross-ancestry meta-analysis of AF-GWAS to individuals who had whole-genome sequencing data available, and assessed the distribution of the polygenic risk score from common-variant associations in rare LOF variant carriers and non-carriers. Our results provide the contribution of common-variant and rare-variant associations to AF pathophysiology and revealed the relationship between these associations in genomic architecture of AF.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1396 Common and unique genetic effects contribute to development of alcohol related liver disease.

Authors:

A. Oliveri, K. C. Cushing, E. Speliotes; Univ. of Michigan, Ann Arbor, MI

Background: Alcoholic-related liver diseases (ALD) affects about 5% of individuals worldwide. Alcoholic liver diseases are heritable or genetically influenced. Identifying genetic predictors of disease can help us identify causal factors to inform risk stratification efforts and development of novel therapeutics. Here we carried out analyses to identify which alcohol consumption risk loci had effects on ALD and determine their effects on alcoholic hepatitis (AH) and alcoholic cirrhosis (AC).
Methods: We carried out a genome wide association study of alcohol consumption in the UK Biobank cohort. Genome-wide significant variants were then tested for association with ALD, AH, and AC in the UK Biobank using established ICD-10 codes (ALD: K70.1 and/or K70.3, AH: K70.1, AC: K70.3). Significantly ($p < 0.05$) associated variants and implicated genes were then evaluated for having expression QTL (eQTL) and gene expression effects in liver in the GTEx Portal.
Results: A total of 23 variants significantly associated with alcohol consumption. This included rs1229984-C (*ADH1B*, p= 3.7e-154), rs28712821-A (*KLB*, p= 2.4E-34), and rs1260326-C (*GCKR*, p= 6.5E-30). Of the 23 alcohol intake associated variants, 2 associated with ALD (*ADH1B*-rs1229984-C [p=0.000419], *MLXIPL*-rs7805504-C [p=0.0453]), 4 associated with AH (*MLXIPL*-rs7805504-C [p=0.0104]; *ADH1B*-rs1229984-C [p= 0.017]; *BRD3OS*-rs109536-C [p=0.0223]; *CYP1A1*-rs2470893-T [p=0.0255]), and 1 associated with AC (*ADH1B*-rs1229984-C [p=0.00208]). Several implicated genes were found to be expressed in the liver (*ADH1B*, *MLXIPL*, *CYP1A1*) and several genes were found to serve as variant-specific eQTLs in the liver (*LINC00094*, *ULK3*, *MPI*).
Discussion: These results show overlapping (*ADH1B*) and unique risk (*MLXIPL*, *BRD3OS*, and *CYP1A1*) of developing ALD, AH, and AC. These results suggest there are multiple ways to curb development of alcoholic liver disease. Further studies will help us to understand the functional effects of these genes on the risk of AH.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1397 Common genetic risk factors for scleroderma renal crisis in African Americans

Authors:

E. Roberson[1], G. Pravitt[2], J. Atkinson[1], J. Varga[3]; [1]Washington Univ., Saint Louis, MO, [2]NIH, Bethesda, MD, [3]Univ. of Michigan, Ann Arbor, MI

Systemic sclerosis (**SSc**) is a systemic disease with autoantibodies, increased connective tissue deposition, immune dysregulation, and vasculopathy. Some people with SSc develop scleroderma renal crisis (**SRC**), which has sudden onset hypertension, rapid renal failure, thrombocytopenia, and microangiopathic hemolysis. Pathologically, the kidney and endothelial damage in SRC is similar to that in thrombotic microangiopathies, some of which are caused by loss-of-function in inhibitors complement activation. *These pathological similarities made us question whether SRC risk was at least partially driven by common variants in complement regulators leading to an increased likelihood of complement activation leading to endothelial damage.*

Testing for common variant modifiers a sub-phenotype of an orphan disease is a challenge. We collaborated with the Genome Research in African Scleroderma Patients (**GRASP**) project to obtain genome-wide SNP data from African American disease controls with SSc only (44) and disease cases with SSc and SRC (50). The samples were typed on the Illumina Multiethic Global Array and were imputed using the 1000G phase 3 v5 reference after QC. We tested each SNP for SRC risk as an additive, dominant, or recessive trait using logistic regression, taking the first 5 principal components into account.

No variant passed a genome-wide significance threshold ($5x10^{-8}$). Despite the small sample size available, 91 variants passed a suggestive threshold ($1x10^{-5}$) under at least one model. One of these suggestive SNP regions was within 200kb of *C8A* and *C8B*, genes that are terminal complement effectors. MAGMA analysis similarly didn't have Bonferroni correction significant hits, but nominally significant hits in alpha/beta T cell activation and the activation/regulation of Th17 responses. It is striking that even using this small sample size that identified SNPs near complement effector genes. Further work is required to determine if these common variants are true risk factors for SRC and to decipher their molecular mechanism.

Given that we have a rare phenotype, it may be more fruitful to look for rare variant enrichment of target complement genes in people with SRC versus the expected population distribution. To this end, we have an ongoing orthogonal project with 300 SRC exomes to generate additional biological insights.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1398 Common variants are associated with congenital diaphragmatic hernia, a rare birth defect

Authors:

L. Qiao[1], C. L. Welch[1], R. Hernan[2], C. Farkouh-Karoleski[1], T. Crombleholme[3], F-Y. Lim[4], H. Needelman[5], R. A. Cusick[5], G. B. Mychaliska[6], B. W. Warner[7], A. Wagner[8], M. Danko[9], D. Chung[9], D. Potoka[10], P. Kosiński[11], D. McCulley[12], M. Li[13], J. C. Fisher[14], M. Elfiky[15], K. Azarow[16], E. Fialkowski[16], D. Schindel[17], S. Z. Soffer[18], C. Maj[19], F. High[20], X. Sun[21], P. Donahoe[22], C. Bendixen[23], E. Brosens[24], Y. Shen[25], W. Chung[25]; [1]Columbia Univ. Irving Med. Ctr., New York, NY, [2]CUIMC, New York, NY, [3]Med. City Children's Hosp., Dallas, TX, [4]Cincinnati Children's Hosp. Med. Ctr., Cincinnati, OH, [5]Univ. of Nebraska Med. Ctr. Coll. of Med., Omaha, NE, [6]Univ. of Michigan Hlth.System, Ann Arbor, MI, [7]Washington Univ. Sch. of Med., St. Louis, MO, [8]Children's Hosp. of Wisconsin, Med. Coll. of Wisconsin, Milwaukee, WI, [9]Monroe Carell Jr. Children's Hosp. at Vanderbilt, Nashville, TN, [10]Univ. of Pittsburgh, Pittsburgh, PA, [11]Med. Univ. of Warsaw, Warsaw, Poland, [12]Dept. of Pediatrics, Univ. of California, San Diego, CA, [13]Rush Univ. Med. Ctr., Chicago, IL, [14]New York Univ. Grossman Sch. of Med., New York, NY, [15]Cairo Univ., Cairo, Egypt, [16]Oregon Hlth.& Sci. Univ., Portland, OR, [17]Med. City Dallas Children's Hosp., Dallas, TX, [18]Northwell Hlth., New York, NY, [19]Inst. for Genomic Statistics and Bioinformatics, Univ. of Bonn, Bonn, Germany, [20]Massachusetts Gen. Hosp., Boston, MA, [21]Dept. of Pediatrics, Univ. of California, San Diego Med. Sch., San Diego, CA, [22]Pediatric Surgical Res. Lab., Massachusetts Gen. Hosp., Boston, MA, [23]Unit of Pediatric Surgery, Dept. of Gen., Visceral, Vascular and Thoracic Surgery, Univ. Hosp. Bonn, Bonn, Germany, [24]ErasmusMC, Breda, Netherlands, [25]Columbia Univ., New York, NY

Background - Congenital diaphragmatic hernia (CDH) is a severe congenital anomaly that is often accompanied by other anomalies and/or neurodevelopmental manifestations. Previous studies showed that rare *de novo* protein-coding variants and copy number variations contribute to CDH. However, the majority of CDH cases cannot be explained by rare *de novo* changes.**Methods** - To investigate the role of common variants in CDH risk, we utilized a cohort of 1,439 cases from the multicenter Diaphragmatic Hernia Research & Exploration; Advancing Molecular Science (DHREAMS) study and Boston cohorts. The cases were comprised of approximately 66% isolated and 34% complex cases, with complex including 52% cardiac defects, 19.5% neurodevelopmental manifestations, and 6.5% pulmonary defects (excluding pulmonary hypoplasia and hypertension). We performed a genome-wide association study (GWAS) using 1,005 unrelated European cases and 4,565 ancestry-matched, unaffected parents from the Simons Powering Autism Research for Knowledge (SPARK) study as controls. For replication, we used an independent cohort of 389 unrelated Dutch/German CDH cases (82% isolated, 18% complex) and 4,815 ancestry-matched controls, and then performed a meta-analysis.**Results** - We identified two loci with genome-wide significance in the discovery cohort - chromosome 3p14.3/lead SNP rs55705711 and chromosome 7q36.3/lead SNP rs7777647 - with replication in the Dutch/German cohort. Meta-analysis confirmed the associations and directionalities: rs55705711 ($p=1.410^{-16}$, OR=1.54) and rs7777647 ($p=3.410^{-11}$, OR=1.26). Allele frequencies of top SNPs at both loci were similar in isolated and complex CDH cases and between cases with and without likely pathogenic *de novo* variants. The lead SNPs explained 1.31% (rs55705711) and 0.7% (rs7777647) of the variance in susceptibility to CDH. The chromosome 3p14.3 locus resides in an intron of *ERC2* (ELK2/RAB6-interacting/CAST family member 2) and adjacent to *WNT5A* (Wnt family member 5A), a developmental patterning gene. Epigenomic data (ATAC-seq, Hi-C, and capture Hi-C) suggest the causal variant is likely in an enhancer region regulating *WNT5A* expression. The 7q36.3 locus resides in a topologically associated domain containing five protein coding genes and a long-range *cis*-acting regulatory domain upstream of sonic hedgehog (*SHH*).**Conclusions** - This first genome-wide association study of CDH identifies new CDH risk loci with potential enhancer functions for known developmental genes and supports a polygenic model as part of the CDH genomic architecture.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1399 Comparing the effect profile of CETP in individuals of East Asian and European ancestries

Authors:

D. Dunca[1], S. Chopade[1], M. Gordillo-Maranon[2], A. Hingorani[2], K. Kuchenbaecker[3], C. Finan[2], A. Schmidt[4]; [1]UCL, London, United Kingdom, [2]Univ. Coll. London, London, United Kingdom, [3]Univ Coll. London, London, United Kingdom, [4]Amsterdam UMC Heart Ctr., Amsterdam, Netherlands

Cholesteryl ester transfer protein (CETP) is a lipid drug target under development for coronary heart disease (CHD) in both European and East Asian populations. Previous drug target Mendelian randomization (MR) studies conducted in East Asians failed to show a CHD effect, which has been interpreted as lack of effectiveness of CETP inhibition for CHD prevention in this population. In this study, we inferred the effect of CETP inhibition in individuals of European and East Asian ancestries using drug target Mendelian randomization.

We leveraged genetic associations of *CETP* variants with major blood lipid fractions for individuals of European (n=1,320,016) and East Asian (n=146,492) ancestries. Colocalization was employed to identify potential cross-ancestry signals of CETP variants for plasma concentrations of LDL-C or HDL-C. Drug target MR was used to estimate ancestry-specific effects of on-target CETP inhibition on cardiometabolic biomarkers, cardiovascular disease endpoints, and potential safety outcomes. Differences between ancestries were evaluated using interaction tests, applying a multiplicity corrected alpha of 1.9×10^{-3} based on the 26 considered traits.

There was strong support (posterior probability=1.00) of a shared causal *CETP* variant affecting HDL-C in both populations, which was not observed for LDL-C. Employing drug target MR scaled to a standard deviation increase in HDL-C, we found that lower CETP was associated with lower LDL-C, Lp[a], systolic blood pressure and pulse pressure in both groups, but the effects were more pronounced in European individuals (interaction p-values < 1.9×10^{-3}). Lower CETP was protective against CHD, angina, intracerebral haemorrhage and heart failure in both ancestries, for example for CHD in East Asians (OR 0.89, 95%CI 0.84;0.94) compared to Europeans (OR 0.95, 95%CI 0.92;0.99, interaction p-value=0.05).

In conclusion, on-target inhibition of CETP is anticipated to decrease cardiovascular disease in individuals of both European and East Asian ancestries.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1400 Comprehensive analysis of genetic factors and subjective response to alcohol in Japanese youths.

Authors:

K. Hikino[1], I. Otsuka[2], A. Hishimoto[2,3], K. Ozaki[4], X. Liu[1], Y. Ishikawa[5], T. Mushiroda[1], S. Matsushita[6], C. Terao[1,7,8]; [1]RIKEN Ctr. for Integrative Med. Sci., Yokohama, Japan, [2]Kobe Univ. Graduate Sch. of Med., Kobe, Japan, [3]Yokohama City Univ. Graduate Sch. of Med., Yokohama, Japan, [4]Natl. Ctr. for Geriatrics and Gerontology, Obu, Aichi, Japan, [5]Riken, Ctr. for Integrative Med. Sci., Yokohama, Japan, [6]Natl. Hosp. Organization Kurihama Med. and Addiction Ctr., Kanagawa, Japan, [7]Shizuoka Gen. Hosp., Shizuoka, Japan, [8]Univ. of Shizuoka, Shizuoka, Japan

Background: Alcohol metabolism is known to be influenced by both environmental and genetic factors. Numerous studies have utilized the subjective response to alcohol (SR) as an intermediate phenotype to examine its significant variability, heritability, and predictive potential for alcohol-related disorders. However, comprehensive analyses linking alcohol reactivity indicators to genetic factors using a substantial number of study participants are still lacking. Methods: We conducted the first comprehensive study to assess the relationship between pharmacodynamics (subjective response to alcohol, SR) using various scales, including the Bodily Sensation Scale (BSS), the Biphasic Alcohol Effects Scale (BAES), and the Subjective High Assessment Scale (SHAS), with a sample of 429 healthy Japanese youths. This study used the intravenous clamp technique and analyzed key genes related to alcohol metabolism, namely rs671 in *ALDH2* and rs1229984 in *ADH1B*. We also examined three loci associated with alcohol behaviors (rs1260326 in *GCKR*, rs3043 in *ALDH1B1*, and rs8187929 in *ALDH1A1*), in conjunction with whole-genome genotyping and imputed data. Results: Our findings showed that (1) SR, as measured by BSS, BAES, and SHAS, were divided into three categories by clustering and principal component analysis. (2) SR may be more significantly influenced by five alcohol-related loci—*ALDH2*, *ADH1B*, *ALDH1B1*, *ALDH1A1*, and *GCKR*—in that specific order. It appears that higher acetaldehyde concentrations have a more profound impact on SR scores. Importantly, these genetic influences also tend to segregate into three distinct clusters. (3) All five loci contributed to the SR at different proportions and stages of alcohol metabolism, the most significant being rs671 in *ALDH2* with 30% of variance explained. The subscales within cluster 1 exhibited a larger explained variance compared to clusters 2 and 3. (4) ADH1B functioned at later stages when ALDH2 function decreased. (5) *ALDH2* and *ADH1B* affected the SR independently. (6) None of the genes significantly contributed to any time points on the BAES (stimulant) scale. (7) Regions unrelated to the five loci were correlated with SR at the polygenic level. Conclusions: This study reveals a more detailed understanding of the involvement of alcohol and related genes, suggesting a risk for alcohol-related diseases based on specific types of sensation. These findings could lead to better preventive measures through preemptive genetic testing against health hazards.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1401 Comprehensive serological profiling in juvenile idiopathic arthritis identifies novel viral and gene-viral interactions.

Authors:

O. Arege[1], M. Sudman[2], M. C. Marion[1], T. D. Howard[3], D. A. Ornelles[1], P. A. Nigrovic[4], S. D. Thompson[5], C. D. Langefeld[6]; [1]Wake Forest Univ., Winston-salem, NC, [2]Cincinnati Children's Hosp. Med. Ctr., Cincinnati, OH, [3]Wake Forest Sch. of Med., Winston-salem, NC, [4]Havard Med. Sch., Boston, MA, [5]Cincinnati Children's Hosp., Cincinnati, OH, [6]Wake Forest Univ. Sch. of Med., Winston-salem, NC

Juvenile idiopathic arthritis (JIA) is the most common childhood chronic rheumatic disease. Patients with JIA tend to have stiffness, swelling, and pain in the joints that can lead to long-term complications (joint and bone damage). JIA disproportionately affects girls, particularly at earlier ages, and has a strong genetic contribution to risk from both HLA and non-HLA genes. Since non-genetic factors explain a majority of the risk of JIA, we hypothesize that viral infection history, in combination with specific HLA alleles, may contribute to the risk for JIA. To test this hypothesis, we performed serological comparison of the viral infection history of 83 children with predominately oligoarticular or polyarticular JIA to 83 healthy age-, sex-, and geographic region-matched controls (37 pairs from Boston Children's Hospital; 46 pairs from Cincinnati Children's Hospital) using high-throughput phage immunoprecipitation sequencing assay (PhIP-Seq, VirScan) covering the complete reference protein sequences of the human virome. We tested seropositivity and antiviral antibody breadth (number of antigens) for each virus using McNemar's test, Wilcoxon signed-rank test, and a paired t-test. We jointly tested antiviral antibody breadth in taxonomic groups of viruses (Baltimore classification) using a permutation-based MANOVA. Finally, we tested five established HLA-*DRB1* four-digit JIA risk alleles in 23 pairs of subjects to determine if a virus-JIA association was modified by the presence of an HLA risk allele. **Results**: JIA patients (76% female, mean ± SD 7.07 ± 2.83 years) and matched healthy controls (77% female; 5.77 ± 2.95 years) exhibited differences ($p<0.05$) in antiviral antibody profiles for seropositivity (HPV-10: OR=5.06; Norwalk virus: OR=0.44) and antibody breadth (HPV-10: JIA 1.94 ± 3.77 vs control 0.74 ± 1.71; BKV: JIA 1.79 ± 2.59 vs control 1.07 ± 1.66; Human parechovirus: JIA 0.91 ± 1.55 vs control 0.50 ± 1.31). Importantly, *DRB1*08*01* modified the viral breadth association with JIA for Adenovirus ($p=0.0476$) and Human papillomavirus ($p=0.057$). **Conclusion:** We tested the paradigm that in JIA, an important childhood autoimmune disease, viral exposures may contribute to risk for JIA and that exposure-induced risk is modified by genetic background. We observed serological differences in viral exposures and antiviral antibody breadth between JIA patients compared to healthy controls, and for some viruses, that relationship was modified by a known JIA-risk HLA allele. This pilot study is limited by the modest sample size and requires replication but puts forth an important gene-environment paradigm in autoimmune disease.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1402 Constructing and Evaluating Sex-Specific Obesity Predictive Genetic Risk Scores Using Taiwan Biobank Data.

Authors:

K-M. Chiang, W-H. Pan; Academia Sinica, Taipei, Taiwan

Background: Obesity has emerged as a significant global public health issue. Over the past three decades, there has been a noticeable increase in the prevalence of obesity and morbid obesity in Taiwan. Previous research has highlighted the important role of genetic factors in the development of obesity. Predicting individualized genetic risk for obesity can enable early interventions to prevent obesity and reduce the risk of related diseases and associated medical costs. Polygenic risk scores (PRS) have emerged as a promising tool for predicting individual clinical phenotypes. However, most studies investigating obesity-related single nucleotide polymorphisms (SNPs) have focused on European and American populations, limiting their applicability in predicting obesity risk in Asian and Chinese populations. Additionally, the influence of gender-specific obesity-related SNPs needs to be considered. **Methods:** In this study, we collected 6,056 obesity-related SNPs from 284 relevant publications. Using biological data from 68,960 Chinese individuals (20,693 males and 46,158 females) from the Taiwan Biobank, we selected 191 SNPs for males and 378 SNPs for females after removing SNPs with high linkage disequilibrium (R^2 > 0.8) and those lacking statistical significance (p-value >0.01) in the Chinese population. These SNPs were utilized in the calculation of gender-specific PRS. The PRS were further divided into deciles to analyze their associations with obesity. Additionally, we further validated the efficacy of these PRS using an independent sample set comprising 27,701 participants (13,806 males and 13,895 females). **Results:** The highest PRS group showed a significant association with obesity (BMI ≥ 24) in both males (OR = 4.35, p-value <0.0001) and females (OR = 5.22, p-value <0.0001). For severe obesity (BMI ≥ 35), the OR was even higher, reaching 14.69 (p-value <0.0001) in males and 26.04 (p-value <0.0001) in females. In the independent sample, we successfully validated the association between PRS and obesity, with an OR of 2.26 (p-value <0.0001) in males and 2.5 (p-value <0.0001) in females for obesity and 2.86 (p-value=0.0048) in males and 5.29 (p-value=0.0007) in females for severe obesity. **Conclusion:** In this study, we have developed and validated gender-specific obesity predictive PRS specifically for Chinese males and females. These risk scores can facilitate early intervention strategies targeting high-risk individuals, thereby mitigating the occurrence of obesity and reducing the socioeconomic burden.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1403 † Contribution of common and rare variants to schizophrenia risk in East and South Asian ancestries.

Authors:

S. Kim[1,2], A. Hassan[3], C. Huai[4], Z. Guo[1,2], J. Vincent[5], M. Mattheisen[6], B. Neale[1,2], M. Ayub[7], S. Qin[4], J. Knowles[8], H. Huang[1,2]; [1]Broad Inst. of MIT and Harvard, Boston, MA, [2]Massachusetts Gen. Hosp., Boston, MA, [3]Univ. of Peshawar, Peshawar, Pakistan, [4]Bio-X Inst.s, Shanghai Jiaotong Univ., Shanghai, China, [5]Ctr. for Addiction and Mental Hlth., Toronto, ON, Canada, [6]Dept. of Community Hlth.and Epidemiology, Dalhousie Univ., Halifax, NS, Canada, [7]Dept. of Psychiatry, Univ. Coll. London, London, United Kingdom, [8]Human Genetics Inst. of New Jersey (HGINJ), Rutgers Univ., Piscataway, NJ

Schizophrenia genetic studies have largely focused on European ancestries, leaving genetic variants in other populations underexplored and potentially increasing health disparities. Recent studies have somewhat rectified this for common variants in East Asian ancestries, but rare variants remain largely uncharacterized in non-European populations. We present the first large-scale schizophrenia sequencing study that examines both common and rare variants in East and South Asian populations, using an innovative, cost-effective sequencing technology.
We developed Blended Genome Exome (BGE), a novel sequencing method merging low-pass whole genome and deep whole exome sequencing. Using the PCR-free library spiked into TWIST exome capture, BGE sequences the exome content at a deep depth (~40x) and the genome content at a low depth (2-4x). This cost-effective technology accomplishes both sequencing tasks in a single experiment, outperforming arrays in non-European ancestries by removing the need to pre-select ancestry-biased variants of interest.
We performed BGE sequencing on large samples of both Chinese (n=17,392) and Pakistan (n=5,181) ancestry, equally split between cases and controls. We performed rigorous processing and quality control followed by association analyses to find schizophrenia-associated genetic variants. We analyzed the exome regions (WES) following the strategy described in the Schizophrenia Exome Sequencing Meta-analysis (SCHEMA) study. Due to its high missing rate, low-pass whole-genome sequencing (WGS) requires special considerations. We used GLIMPSE, a newly developed tool to refine and impute genotype likelihoods for low-pass WGS efficiently. We found BGE outperformed all genotyping arrays for common variants, even dense ones such as the Omni2.5, and has sufficient sensitivity to capture ultra-rare coding variants that were shown to play a key role in schizophrenia. Combining with existing schizophrenia common and rare variant resources, we expect to identify novel genetic loci associated with schizophrenia in East and South Asian ancestries.
This study provides insight into the contribution of common and rare variants to schizophrenia risk in East and South Asia ancestries through a new sequencing technology and a large-scale sample.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1404 † Contribution of the X chromosome to complex traits genetics and sex differences and evidence for escape from X chromosome inactivation in height genetics

Authors:

Y. Fu[1], S. Ruotsalainen[1], L. Lehtonen[1], M. Pirinen[2,1], T. Tukiainen[1]; [1]Inst. for Molecular Med. Finland, FIMM, Helsinki, Finland, [2]Univ Helsinki, Helsinki, Finland

In humans, random X chromosome inactivation (XCI) in females ensures one active chrX per female and male cell. However, gene expression studies showed ~25% of genes escape XCI, expressing at a reduced level from the inactivated X. These phenomena complicate the analysis and interpretation of genetic data on chrX, leading to a lack of consensus on the impact of chrX on complex traits and sex differences. Here, by leveraging the UK Biobank data of 184,583 females and 159,112 males across 48 quantitative traits, we show that 1) in population chrX contributes to complex trait heritability comparably to any autosome, yet with clear sex difference introduced by the difference in chrX copies; 2) escape from XCI can be reflected in the genetic architecture of a trait with subtle impacts on allelic effects and heritability.The hemizygosity of chrX in males could potentially lead to sex difference in chrX contribution (XC) to genetics. By quantifying XC as the ratio of chrX to autosomal heritabilities, we indeed found that XCs in males (median 4.4%) exceeded the expected proportion based on the number of independent chrX SNPs (3.5%). In females, XCs (median 1.7%) were much lower than the expectation, with 67% of the traits significantly depleted of XCs. In the overall population, these sex differences were, however, balanced out, averaging closely to the expectation (median 3.0% vs 3.5%), with height exceptionally enriched with XCs in both sexes (female 3.8%, male 5.9%). While random XCI equalizes chrX dosage between sexes, local escape from XCI potentially introduces a subtle increase in female allelic effects, thus, in theory, further balancing out the sex differences in XCs. To examine if escape is seen on the chromosome-wide level, we compared XCs between sexes per trait. For most traits, we lacked statistical power to distinguish if the observed sex differences in XCs are in line with full XCI or partial escape from XCI as only 1.1-fold difference is expected between the two scenarios. For height, a trait highly heritable on chrX, we observed an increase in female XC consistent with a partial escape from XCI. To assess escape at a finer scale, we quantified the proportion of SNPs with larger effect sizes in females. We observed an excess of female bias in chrX compared to autosomes (75% vs 28%) in height and further pinpointed regions that potentially escape XCI.In conclusion, our study shows chrX as an important source of missing heritability and sex differences of complex trait genetics. We demonstrate that the effect of escape from XCI is reflected in genetic associations, although this effect seems small enough that considerably larger sample sizes are required for traits other than height.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1405 Convergent evidence from common and rare variant analyses implicating the immune system and late infancy cerebellar development in ADHD using a Hong Kong sample

Authors:

Y. Zhong[1], B. Larry[1], T. Justin[1], Y. Rui[1], H. Se-Fong[2], T. Chun-Pan[3], H. Ting-Pong[2], L. Chi-Chiu[3], S. Pak C[1], L. Patrick W. L.[2]; [1]The Univ. of Hong Kong, Hong Kong, Hong Kong, [2]The Chinese Univ. of Hong Kong, Hong Kong, Hong Kong, [3]Kwai Chung Hosp., Hong Kong, Hong Kong

Attention-deficit hyperactivity disorder (ADHD) is a common neuropsychiatric disorder with a significant genetic component, characterized by persistent symptoms of inattention, hyperactivity and/or impulsivity. The latest genome-wide association study (GWAS) meta-analysis of ADHD identified 27 whole-genome significant risk loci in the European population. However, genetic risk factors for ADHD are less well-characterized in the Asian population, especially for rare variants. Here, we present an analysis of common and rare variant contributions to ADHD in a Hong Kong sample comprising 279 cases and 432 controls, who were genotyped using the Illumina Infinium Global Screening Array. We identified 41 potential genomic risk loci with a suggestive association ($p < 1e-4$), pointing to 111 candidate risk genes, which were enriched for genes differentially expressed in the cerebellum. Furthermore, brain developmental stage enrichment analysis indicated the late infancy stage, a landmark for human nervous system development especially the cerebellum, to be a critical window for ADHD vulnerability. At the polygenic level, we also discovered a strong genetic correlation with resting-state functional MRI connectivity of the cerebellum involved in the attention/central executive and subcortical-cerebellum networks, which is consistent with the neural pathophysiology for ADHD. In rare variant analyses, we discovered that ADHD cases carried an elevated load of rare damaging variants in *TEP1*, *MTMR10*, *DBH*, *TBCC* and *ANO1*. These genes are annotated to be part of functional clusters relevant to the nervous system and immune response. ADHD genetic risk was associated with immune response/signalling, demonstrated in both common (pathway enrichment analysis) and rare (the gene-set-based burden test) variant analyses as well. *POC1B*, a gene previously found in a genome-wide significant locus of ADHD in the European population, was replicated in the current study. In addition, an accumulation of ADHD common-variant risks found in European Ancestry samples was found to be significantly associated with ADHD in the current study, potentially implicating genetic factors with a trans-ancestral effect in ADHD. These findings re-validate the abnormal development of the neural system, especially the cerebellum, in ADHD and extend the existing neuro-dysfunction hypothesis to a multi-system perspective.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1406 Coronary heart disease and type 2 diabetes metabolomic signatures in a middle eastern cohort.

Authors:

M. Elshrif[1], K. Isufaj[1], A. El-Menyar[2,3], K. Kunji[1], E. Ullah[4], R. Elsousy[5], H. R. Mokhtar[6], E. Ahmad[6], M. Al-Nesf[7], A. Beotra[6], M. Al-Maadheed[6], V. Mohamed-Ali[6], M. Saad[1], J. Al Suwaidi[5]; [1]Qatar Computing Res. Inst. (QCRI) - HBKU, Doha, Qatar, [2]Clinical Res., Trauma & Vascular Surgery, Hamad Med. Corp., Doha, Qatar, [3]Dept. of Clinical Med., Weill Cornell Med. Coll., Doha, Qatar, [4]Qatar Computing Res. Inst. (QCRI), HBKU, Doha, Qatar, [5]Dept. of Cardiology, Heart Hosp., Hamad Med. Corp., Doha, Qatar, [6]Anti-doping Lab Qatar, Doha, Qatar, [7]Dept. of Internal Med., Allergy and Immunology, Hamad Gen. Hosp., Hamad Med. Corp., Doha, Qatar

Background: The growing field of metabolomics has opened new venues for identifying biomarkers of type 2 diabetes (T2D) and predicting its consequences such as coronary heart disease (CHD). Middle Eastern populations, despite their large size, are underrepresented in omics research.**Methods:** In this study, we used a total of 641 metabolites from a large cohort of 3,679 Qatari adults from the Qatar Biobank (QBB; 272 T2D and 2438 non-T2D individuals) and Qatar Cardiovascular Biorepository (QCBio; all CHD patients; 488 T2D and 481 non-T2D individuals). Univariate analysis was performed to identify metabolites associated with T2D, in the absence or presence of CHD. Multivariate analysis, machine learning (ML) models, and metabolite risk scores were developed to assess the predictive power of the different combination of T2D and CHD.**Results:** Many metabolites were significantly associated with T2D in both QBB and QCBio cohorts. Among these, we observed 1,5-anhydroglucitol (1,5-AG) ($P = 1.33 \times 10^{-68}$ in QBB vs 9.82×10^{-33} in QCBio), glucose ($P = 7.14 \times 10^{-57}$ in QBB vs. 3.26×10^{-29} in QCBio), and mannose ($P = 2.61 \times 10^{-54}$ in QBB vs. 1.01×10^{-27} in QCBio). Other metabolites were significantly associated with T2D only in one cohort, e.g., glutamine ($P = 7.14 \times 10^{-57}$ and β = -4.87 in QBB vs. P= 0.26 and β = 0.3 in QCBio) and N6-carboxymethyllysine (P = 0.38 and β = 0.1 in QBB vs. P= 2.24×10^{-8} and β = 0.37 in QCBio). ML models performed well to predict all T2D and CHD combinations with accuracy reaching 80% for some groups. T2D was predicted with higher accuracy (>80% in both QBB and QCBio). The MRS developed in QCBio and tested in QBB while adjusting for HbA1C yielded an Odds Ratio of 21.18 for the top quintile vs. the remaining quintiles.**Conclusions:** Metabolomics profiling has the potential for early detection of metabolic alterations that precede clinical symptoms of T2D and CHD in the presence of T2D. This early detection potential allows for timely interventions and improved management strategies for both T2D and CHD.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1407 Correlation of Interleukin-6 gene (-174G/C) promoter polymorphism &It Human leukocyte antigen DQ genotypes with genetic susceptibility of Type 1 Diabetes Mellitus in Kuwaiti children

Authors:

M. Haider[1], M. Al-Rushood[2,3], H. Alsharhan[4,5], M. A. Rasoul[1], M. Al-Mahdi[6], H. Al-Kandari[7,5]; [1]Kuwait Univ., Coll. of Med., Pediatrics Dept., Safat, Kuwait, [2]Kuwait Univ. Coll. of Med., Pediatrics Dept., Safat, Kuwait, [3]Mubarak Al-Kabeer Hosp., Pediatrics Dept., Jabriya, Kuwait, [4]Kuwait Univ., Coll. of Med., Pediatrics Dept., Kuwait, Kuwait, [5]Farwania Hosp., Pediatrics Dept., Farwania, Kuwait, [6]Al-Adan Hosp., Pediatrics Dept., Al-Adan, Kuwait, [7]Dasman Diabetes Inst., Dept. of Population Hlth., Dasman, Kuwait

Type-1 diabetes mellitus (T1DM) is characterized by autoimmune destruction of the insulin-secreting beta cells by the lymphocytes and inflammatory cytokines. Cytokines play a crucial role in pathogenesis of autoimmune diseases possibly due to their effector and regulatory functions in immune and inflammatory responses. The interleukin-6 (*IL-6)* gene is located on chromosome 17 and carries a single nucleotide polymorphism (SNP, -174G/C, rs1800795) in its promoter region, which has been associated with transcription rates of IL-6 cytokine. We have determined the genotype frequency of *IL-6* gene (-174G/C) promoter polymorphism and human leukocyte antigen (HLA) DQ genotypes in Kuwaiti children with T1DM to investigate their role in genetic susceptibility. This study included 232 Kuwaiti children with T1DM and 200 controls. The criteria by International Society for Pediatric and Adolescent Diabetes (ISPAD) was used for diagnosis of T1DM. The control subjects were healthy Kuwaitis; none had close relative with T1DM and were evaluated by a specialist. The genotypes of *IL-6* (-174G/C) gene polymorphism, were identified by PCR-RFLP method. HLA-DQ alleles were determined by sequence-specific PCR method. The frequency of *IL-6* (-174G/C) gene polymorphism showed a positive correlation with T1DM in Kuwaiti patients in both the co-dominant and dominant models of genetic analysis (OR, 1.10 and 1.02 respectively). We have identified HLA DQ2/2, DQ2/8 and DQ8/8 as the high-risk genotypes for T1DM amongst Kuwaiti children. In this study, we also determined the co-inheritance of *IL-6* gene polymorphism in Kuwaiti T1DM children with different HLA-DQ genotypes. The proportion of Kuwaiti T1DM who carried at least one C-allele of the *IL-6* gene polymorphism was 92% in patients with HLA-DQ2/DQ2 genotype, 96% in those who had DQ2/8 genotype and 96% in patients with HLA-DQ8/DQ8 genotype. Our data highlights the role of C-allele of *IL-6* (-174G/C) gene polymorphism along with HLA-DQ2/2, DQ2/8 and HLA-DQ8/8 genotypes in determining the genetic susceptibility of T1DM in Kuwaiti children.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1408 Critical role of GABAergic neurons in driving CGG repeat toxicity associated with Fragile X premutation

Authors:

Y. Jin, Y. Cao, N. Rivera, W. Ma, J. Huang, Y. Li, Y. Kang, S. Yu, R. Li, R. Li, E. Allen, S. Qin, M. Epstein, P. Jin; Emory Univ., Atlanta, GA

Fragile X-associated tremor/ataxia syndrome (FXTAS) is a neurodegenerative disorder characterized by age-dependent symptoms. It is caused by a CGG repeat expansion (55-200) in the 5′ UTR of the fragile X messenger ribonucleoprotein 1 (FMR1) gene, associated with fragile X premutation carrier. The molecular mechanisms underlying FXTAS involve both CGG repeat RNA and a repeat-associated non-AUG (RAN) translation product called FMRpolyG peptide. To further understand the molecular pathogenesis of FXTAS, we performed a transcriptome-wide association study (TWAS) analysis using prediction models of the cortex and cerebellum. Through this analysis, we identified 45 genes significantly associated with the onset of FXTAS, which we further validated using a *Drosophila* model of FXTAS. Among these genes, ten were found to modulate the toxicity associated with expanded CGG repeats. Additionally, we performed single nucleus RNA-Seq on the prefrontal cortex of FXTAS and control mice, revealing cell type-specific gene expression signatures induced by CGG repeat expression. Notably, we observed pronounced gene expression changes in inhibitory neurons, suggesting their strong involvement in the overall transcriptional response. To determine the specific contributions of distinct neuronal cell types to FXTAS pathogenesis, we utilized conditional FXTAS mouse models and various Cre driver lines. Our findings indicated that the expression of CGG repeats in GABAergic neurons alone was sufficient to induce neuronal toxicity, whereas the expression in glutamatergic neurons had minimal impact. Furthermore, employing TRAP-seq, we identified several gene expression alterations specific to GABAergic neurons, some of which were found to be bound by hnRNPA2B1, an RNA-binding protein sequestered by CGG repeat RNA. Notably, PRKCG, a risk gene for SCA14, was among the identified genes and showed potential in modulating CGG repeat-associated toxicity. This suggests that hnRNP A2/B1-mediated regulation of PRKCG plays a significant role in FXTAS pathogenesis. Collectively, our integrated analyses highlight the critical role of GABAergic neurons in driving FXTAS pathogenesis and identify a potential therapeutic target for FXTAS.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1409 Cross-ancestry analyses identify new genetic loci associated with 25-hydroxyvitamin D.

Authors:

J. Revez[1], X. Wang[1], V. Hivert[1], S. Groot[1], Y. Wang[2], L. Yengo[1], J. J. McGrath[1], K. E. Kemper[1], P. M. Visscher[1], N. R. Wray[1]; [1]The Univ. of Queensland, Brisbane, Australia, [2]Massachusetts Gen. Hosp., Boston, MA

BACKGROUND: Vitamin D status - a complex trait influenced by environmental and genetic factors - is tightly associated with skin colour and/or genetic background. Yet very few studies have investigated the genetic underpinnings of vitamin D in non-European ancestry groups, and the ones that have, relied on small sample sizes, resulting in inconclusive results.
METHODS: We conduct genome-wide association analyses of 25 hydroxyvitamin D (25OHD) - the main circulating form of vitamin D - in 442,435 individuals from four genetically-inferred broad ancestry groups represented in the UK Biobank: European (N=421,867), South Asian (N=9,983), African (N=8,306) and East Asian (N=2,279). We test additive and dominance effects. Given that vitamin D metabolism occurs in the skin, we conduct GWAS of 25OHD stratified by skin colour. Lastly, we test the interaction between 25OHD-associated variants and skin colour.
RESULTS: We identify a new genetic determinant of 25OHD (rs146759773) in individuals of African ancestry, which was not detected in previous analysis of much larger European cohorts due to low minor allele frequency. We show genome-wide significant evidence of dominance effects in 25OHD that protect against vitamin D deficiency. In skin-colour-stratified analyses, we identify new genome-wide significant associations with 25OHD. Lastly, we identify two loci (rs10832254 and rs1352846) whose association with 25OHD differs in individuals of distinct complexions.
CONCLUSIONS: Collectively, our results provide new insights into the complex relationship between 25OHD and skin colour and highlight the importance of diversity in genomic studies. Despite the much larger rates of vitamin D deficiency that we and others report for ancestry groups with dark skin (e.g., South Asian), our study highlights the importance of considering ancestral background and/or skin colour when assessing the implications of low vitamin D.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1410 Cross-ancestry GWAS meta-analysis of keloids discovers novel susceptibility loci in diverse populations.

Authors:

C. Greene[1], G. Hampton[1], H. Seagle[1], Y. Luo[2], G. Jarvik[3], B. Namjou-Khales[4], A. Khan[5], T. Edwards[6], D. Velez Edwards[6], J. Hellwege[6,7]; [1]Vanderbilt Univ., Nashville, TN, [2]Northwestern Univ., Evanston, IL, [3]Univ. of Washington Med. Ctr., Seattle, WA, [4]Children's Hosp. Med. Ctr., Cincinnati, OH, [5]Columbia Univ., New York City, NY, [6]Vanderbilt Univ. Med. Ctr., Nashville, TN, [7]VA Tennessee Valley Hlth.care System, Nashville, TN

Keloids are benign fibroproliferative growths that form after injury to the skin and grow beyond the original wound boundaries. They can result in pain and disfigurement. Darker-skinned individuals are much more likely to develop keloids, particularly those of African and Asian descent. Prevalence estimates vary from less than 0.1% in European-ancestry populations to 16% in some African-ancestry populations. However, most genetic research has been limited to European and East Asian populations despite African-ancestry populations having a 20-fold increased risk for keloids. We have performed a large cross-ancestry genome-wide association study (GWAS) meta-analysis of keloids, incorporating data from 1,600,846 individuals (7,837 cases and 1,593,009 controls). These include 139,538 African-ancestry individuals (2,696 cases and 136,842 controls). We detected 142 novel loci in the cross-ancestry meta-analysis, which included seven replicated variants and 1,233 novel variants. The most significant result (P = 1.65 x 10-79) was at a variant (rs10863683) located downstream of LINC01705. LINC01705 was previously associated with keloids and implicated as a regulatory factor underlying tumorigenesis. Additionally, analysis of genetically-predicted gene expression with S-PrediXcan identified a significant association between decreased risk of keloids and increased expression of LINC01705 in fibroblasts (P = 7.62 x 10-21), which play an important role in wound healing. Other results include associations with NEDD4 and LSP1 in fibroblasts, as well as PHLDA3 in sun-exposed skin (P = 7.94 x 10-11, 5.96 x 10-08, and 3.81 x 10-14, respectively). The most significant result in the African-ancestry meta-analysis (P = 1.95 x 10-32) was an intergenic variant (rs34647667) in a conserved region downstream of ITGA11, though NEDD4 was the only gene in this region of chr15 to reach significance (P < 2.6 x 10-6) in the gene-based test. Previous research in our group identified an African ancestry-specific susceptibility locus at chr15q21-22, and our results support these prior findings. Keloid SNP-based heritability estimates using Linkage Disequilibrium Score Regression (LDSC) are 6%, 21%, and 34% for European, East Asian, and African ancestry, respectively. These results support a potential adaptive origin for keloid disparities in African-ancestry populations and significantly increase the yield of discoveries from keloid genetic association studies.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1411 Cross-trait genetic analysis highlights causal effects of chronic pain on older-adult-onset asthma with evidence for mediation by neuroimmune crosstalk in the airways.

Authors:

G. Koli Farhood, C. Chen, Y. Chen, N. Dimitrov, M. Parisien, L. Diatchenko, A. Grant; McGill Univ., Montreal, QC, Canada

Background: Chronic pain (persistence of pain experience $\geq$ 3 months) and asthma are heritable, with similar prevalence of approximately 10% worldwide. People with chronic pain usually suffer from other chronic comorbidities including autoimmune disorders and chronic pulmonary diseases, and recent genetic studies identified asthma as displaying the highest correlation with chronic pain among autoimmune disorders. While reciprocal interactions between the nervous and immune systems are considered to underlie the pathogenesis of both chronic pain and asthma, specific biological pathways have not been identified. We explored the extent to which shared genetics underlying chronic pain and asthma considering age-of-onset (AO) may lead to identification of common biological functions at separate life span stages. **Methods:** Genome-wide association (GWA) scans on chronic pain and AO-stratified asthma phenotypes (< 18 years (childhood), 18-40 years (younger-adult), and > 40 years (older-adult)) were conducted across 9.2 million variants on 457,461 European-descent subjects at baseline visit from the UK Biobank (UKB). We performed bidirectional Mendelian Randomization (MR) to assess whether effects of significant genetic variants in either chronic pain or AO-stratified asthma traits were mediated by a trait from the other set. We assessed the temporal order of cause-and-effect relationships between the two traits sets from baseline to subsequent visits of the UKB and the Canadian Longitudinal study of Aging (CLSA). We also constructed polygenic scores (PGS) in the UKB to predict risk in the CLSA across trait sets. We then implemented gene-based analysis on cross-trait GWA meta-analysis in conjunction with Gene Ontology, GTEx, and Gene Enrichment Profiler datasets. **Results:** We identified 1,416 causal variants with significant effects on older-adult asthma, which were mediated by chronic pain in MR analysis. Chronic pain to older-adult asthma causal directionality was further supported by the observation of increased risk of older-adult asthma at subsequent visits in both the UKB and the CLSA. The UKB PGS for chronic pain predicted increased risk for older-adult asthma in the CLSA. Gene-based analysis on the chronic pain and older-adult asthma meta-analysis highlighted differential nervous system expression, particularly in the cerebellum. Furthermore, we identified strong cell-specific heritability enrichment localizing to the central nervous system, lung, and immune system cells including B and T cells. **Conclusion:** Older adult-onset asthma is activated by long lasting neuroimmune cells communication in the airway, initiated by chronic pain.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1412 Cystic fibrosis risk variants confer a protective effect against inflammatory bowel disease in large-scale exome sequencing analysis

Authors:

M. Yu[1,2]; [1]Broad Inst. of MIT and Harvard, Cambridge, MA, [2]Analytic and Translational Genetics Unit, Massachusetts Gen. Hosp., Boston, MA

Genetic mutations that lead to the production of defective Cystic fibrosis transmembrane regulator (CFTR) protein are known to cause cystic fibrosis (CF), a lethal autosomal recessive Mendelian disorder. Existing studies have debated the roles of CFTR mutations in Inflammatory Bowel Disease (IBD), with inconclusive evidence supporting both risk and protective effects. Here, leveraging the largest IBD exome sequencing callset with 38744 cases and 69570 controls of European ancestries from the International Inflammatory Bowel Disease Genetics Consortium, we report the first compelling evidence establishing a protective role of CF-causing variants against IBD.

In our association analysis, we found delF508, a known CF-causing variant that accounts for 70% of all CFTR mutations observed in CF patients, protects against IBD with an effect size of -0.30 (SE=0.048, P=1.7E-10). Moreover, we observed a number of rare CF-causing variants in the CFTR2 database, e.g., G542X, G551D, and N1303K, which also conferred protective effects against IBD. To follow up on these observations, we conducted an unweighted gene burden test of CFTR restricted to all variants annotated as CF-causing in the CFTR2 database (excluding delF508). This variant set yields a composite allele frequency (CAF) of 1.2% and a notable protective effect of -0.23 (SE=0.047, P=1.1E-6). As a negative control, another burden test was performed on CFTR using all rare (MAF<0.001) nonsynonymous non-CF-causing variants (CAF=1.7%). This test showed no significant association between CFTR and IBD (P=0.42), demonstrating the importance of selecting relevant variants in burden tests. To evaluate variant selection using in silico evidence, we also performed burden tests using variants with high pathogenicity scores from CADD and REVEL, neither of which yielded comparable significance (P=3.7E-3 and 1.4E-3, respectively).

In this study, we established a protective role for CF-causing variants against IBD through large-scale exome sequencing analysis. We also demonstrated that existing in silico variant classifiers are ineffective at separating pathogenic and neutral variants in the case of CFTR, motivating the need for systemic functional studies to investigate the causal mechanisms for this association.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1413 Cytosine-to-Uracil RNA editing is upregulated by inflammatory stimulation of monocytes

Authors:

H. Seo[1,2,3,4,5], W. Cuddleston[1,2,3,4,5], M. Parks[1,2,3,4,5], E. Navarro[1,2,3,4,5], T. Raj[1,2,3,4,5], J. Humphrey[1,2,3,4,5]; [1]Dept. of Genetics and Genomic Sci., Icahn Sch. of Med. at Mount Sinai, New York, NY, [2]Nash Family Dept. of NeuroSci. & Friedman Brain Inst., Icahn Sch. of Med. at Mount Sinai, New York, NY, [3]Ronald M. Loeb Ctr. for Alzheimer's Disease, Icahn Sch. of Med. at Mount Sinai, New York, NY, [4]Icahn Inst. for Data Sci. and Genomic Technology, Icahn Sch. of Med. at Mount Sinai, New York, NY, [5]Estelle and Daniel Maggin Dept. of Neurology, Icahn Sch. of Med. at Mount Sinai, New York, NY

Genome-wide association studies of the neurodegenerative diseases Alzheimer's and Parkinson's have highlighted genes specific to myeloid cells as enriched in disease heritability. Myeloid cells are the innate immune system lineage, including monocytes and macrophages in the periphery and microglia in the central nervous system. By studying transcriptional regulation of myeloid cells we can uncover new insights into their role in neurodegeneration.

RNA editing is a post-transcriptional modification whereby specific enzymes alter the RNA base sequence. This can lead to various downstream consequences, including re-coding of the amino acid sequence or changes in RNA stability. The most well-known form of RNA editing in the brain is Adenosine to Inosine (A-I), facilitated by the ADAR enzyme family. In contrast, the conversion of Cytosine to Uracil (C-U) by the APOBEC enzyme family is largely understudied. C-U editing components are highly upregulated in microglia compared to the rest of the brain, suggesting a key role in myeloid cell function. Previous work has found that C-U editing sites are enriched in exonic re-coding events, and editing is a dynamically regulated process in response to inflammatory stimulation. However, the full extent of C-U editing sites throughout the myeloid transcriptome and the consequences for gene regulation are unknown.

As a proxy for microglia, we sorted monocytes from peripheral blood (n=52) and stimulated them with lipopolysaccharide (LPS), interferon beta (IFNb), or PBS (control) for 24 hours before performing RNA sequencing. We constructed a computational pipeline to measure A-I and C-U RNA editing and perform extensive quality control to remove false positive sites. We also estimated gene expression and isoform usage. We identified 6,278 editing sites in our monocyte samples, of which 1,077 sites were differentially edited in response to LPS and 3,795 in IFNb. Strikingly, 1,223 of the IFNb sites were C-U, which were enriched for re-coding and protein-truncating events within exons. By combining gene expression and splicing of differentially edited genes, we observed complex relationships between the three modalities and the two stimulation types, which also suggest potential regulatory relationships between RNA editing and gene expression levels of several APOBEC enzymes. Finally, we observed stimulation-specific editing of the Alzheimer's disease risk genes PILRA and PLCG2.

This work is the first step towards characterizing the C-U editing landscape in myeloid cells. We next plan to integrate common genetic variants into RNA editing analysis to interpret causality and genetic links to neurodegenerative disease.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1414 De novo mutations in bipolar disorder implicate genes involved in neurodevelopment and immunity.

Authors:

T. Shimada[1,2,3], S. Rachel[4,3], H. Sung[3], E. Besancon[3], F. Lopez[3], L. Kassem[3], N. Akula[3], F. J. McMahon[3]; [1]Dept. of Psychiatry and Behavioral Sci., Juntendo Univ. Graduate Sch. of Med., Tokyo, Japan, [2]Dept. of Psychiatry, Juntendo Univ. Urayasu Hosp., Chiba, Japan, [3]Section on the Genetic Basis of Mood and Anxiety Disorders Section, Human Genetics Branch, Natl. Inst. of Mental Hlth., NIH, Bethesda, MD, [4]Brain Mapping Unit, Dept. of Psychiatry, Univ. of Cambridge, Cambridge, United Kingdom

Background: Bipolar disorder (BD) is a debilitating disorder affecting ~2% of the world's population. Genome wide association studies have identified over 100 common risk loci, a few copy number variants, and one gene with increased burden of loss of function (lof) mutations, but there have been few studies of de novo variation in BD and no genes can yet be considered genome-wide significant de novo hits. For the first time, this study explores the contribution of de novo mutations (DNMs) to BD and related conditions in multiplex families ascertained from genetically isolated populations. **Methods:** Exomes of 199 complete trios ascertained as part of the Amish Mennonite Bipolar Genetics (AMBiGen) study were sequenced at Regeneron, quality controlled and analyzed using GATK. DNMs were called using HAIL, followed by stringent sample and variant filters and Sanger sequencing of selected putative DNMs. Genes with DNMs in cases were combined with genes hit by DNMs among 354 previously published BD trios for functional enrichment analysis using hypergeometric tests, topic modeling, and tests of cell-type enrichment. **Results:** A total of 42 lof or damaging missense DNMs in 42 genes, including *NRXN1*, *SHANK3*, and *SPECC1*, were detected among individuals with BD and related disorders. Three genes, *SUGP2*, *DICER1*, and *PLEC*, have been reported in previous studies. When combined with previously-published genes, the full set of 275 genes was significantly enriched for functions related to learning, post-synaptic organization, nervous system development, and calcium ion transport. These genes also significantly overlapped with brain co-expression modules associated with neurogenesis and immunity. These modules were significantly enriched in genes expressed in excitatory neurons, endothelial cells, and microglia. **Conclusion:** Although no single gene had a significant excess of DNMs, these findings suggest that DNMs in gene sets involved in neurodevelopment and immunity may contribute to BD. Future studies should focus on replicating these findings in additional samples. Genes with significant burdens of DNMs will be good candidates for functional genomic studies.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1415 Decoding disease and HLA-specific regulation of T cell receptor repertoire in Parkinson's disease

Authors:

T. Naito[1,2,3,4,5], R. Vialle[6], K. Ishigaki[7], Y. Okada[8,9,10], D. A. Knowles[5,11,12], T. Raj[1,2,3,4]; [1]Nash Family Dept. of NeuroSci. & Friedman Brain Inst., Icahn Sch. of Med. at Mount, New York, NY, [2]Dept. of Genetics and Genomic Sci., Icahn Sch. of Med. at Mount Sinai, New York, NY, [3]Ronald M. Loeb Ctr. for Alzheimer's Disease, Icahn Sch. of Med. at Mount Sinai, New York, NY, [4]Icahn Inst. for Data Sci. and Genomic Technology, Icahn Sch. of Med. at Mount Sinai, New York, NY, [5]New York Genome Ctr., New York, NY, [6]Rush Alzheimer's Disease Ctr., Rush Univ. Med. Ctr., Chicago, IL, [7]Lab. for Human Immunogenetics, Ctr. for Integrative Med. Sci., RIKEN, Yokohama, Japan, [8]Dept. of Statistical Genetics, Osaka Univ. Graduate Sch. of Med., Osaka, Japan, [9]Lab. for Systems Genetics, RIKEN Ctr. for Integrative Med. Sci., Kanagawa, Japan, [10]Dept. of Genome Informatics, Graduate Sch. of Med., the Univ. of Tokyo, Tokyo, Japan, [11]Dept. of Computer Sci., Columbia Univ., New York, NY, [12]Dept. of Systems Biology, Columbia Univ., New York, NY

Despite the accumulating evidence supporting the involvement of adaptive immune systems in the pathogenesis of Parkinson's disease (PD), the T cell receptor (TCR) repertoire associated with PD has not been well characterized. This is primarily due to the high diversity of the TCR repertoire resulting from human leukocyte antigen (HLA) types and the lack of sufficient resources informative about the TCR repertoire in PD patients that would allow us to characterize the feature stratified by HLA types. To address these issues, we investigated the TCR repertoire features associated with HLA types in PD by utilizing genomic and transcriptomic data from the Accelerating Medicines Partnership Parkinson's disease (AMP-PD) cohort. We targeted 1,127 PD cases and 700 controls from the AMP-PD dataset. We performed TCR repertoire reconstruction from whole blood RNA-seq and HLA imputation from genotype data. Through a comprehensive interaction analysis between HLA alleles of all classical HLA genes and TCR genes, we evaluated how specific HLA features influence the TCR repertoire specifically in PD cases. The strongest interaction was identified at HLA-DRB1*04 and TRAV13-2*01 among TRA genes ($P = 6.42 \times 10^{-7}$) and TRBV14*01 among TRBV genes ($P = 4.30 \times 10^{-6}$). HLA-DRB1*04 was previously shown to be the most risk-associated HLA allele with PD (Naito et al. 2021 and Yu et al. 2021). Notably, HLA-DRB1*04 exhibited the highest enrichment in interaction with various TCR genes compared to other HLA alleles. Furthermore, we conducted longitudinal TCR repertoire analysis and detected TCR features that significantly changed especially in PD cases. We are currently investigating longitudinal TCR regulation depending on specific HLA types. These findings present the evidence supporting the role of PD risk-associated HLA alleles in shaping TCR repertoire in a disease-specific manner. Additionally, specific TCR genes might serve as potential markers for the duration and progression of PD. We will validate our findings using curated samples of TCR sequencing and single-cell RNA-seq. Overall, these findings enhance our understanding of immune dysregulation in PD and may pave the way for future therapeutic interventions targeting the adaptive immune systems.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1416 † Decoding Genetic Modifiers of Disease Variants in Large-Scale Biobanks

Authors:

Z. Yu[1], A. Jowell[2], S. Koyama[1], M. Trinder[3], W. Hornsby[4], K. Cho[5], J. Gaziano[5], P. Wilson[6], Y. Sun[7], A. Bick[8], P. Natarajan[4]; [1]Broad Inst. of MIT and Harvard, Cambridge, MA, [2]Harvard Med. Sch., Boston, MA, [3]Univ. of British Columbia, Vancouver, BC, Canada, [4]Massachusetts Gen. Hosp., Boston, MA, [5]VA Boston Hlth.care System, Boston, MA, [6]Atlanta VA Med. Ctr., Decatur, GA, [7]Emory Univ., Johns Creek, GA, [8]Vanderbilt, Nashville, TN

Phenotypic expressions of DNA sequence variants are influenced by genetic, environmental, and stochastic factors. Unveiling these modifying factors can illuminate disease pathways and refine genetic counseling. However, the discovery of genetic modifiers in humans is limited. Here, we used large-scale genotyping and comprehensive phenotyping for a genome-wide modifier scan in the UK Biobank (UKB), with validation in the US VA Million Veteran Program (MVP). We selected disease mutations that are common, located in coding regions, and categorized as pathogenic, risk factors, or indicative of susceptibility to diseases in ClinVar. We then conducted phenome-wide association tests on these mutations across 1,601 PheCode disease phenotypes in UKB. We selected the five strongest associated disease outcomes for each mutation and conducted modifier scans for disease mutation-phenotype pairs whose disease phenotypes were prevalent and the associations were strong. Modifier searches included an interaction term between the disease mutation and potential modifier mutations in regression models. We modeled each disease mutation in two forms: as an individual SNP and as a binary variable of collapsed rare variants on the corresponding gene. From the analysis of 378,184 UKB EA participants, we pinpointed 16 disease mutations in 12 genes associated with 57 disease outcomes. Notably, we discovered 17,578 significant interactions (959 after pruning). We focused on interactions replicated in MVP EA (N~450K), particularly those where modifiers were coding variants, disease and modifier mutations were located in the cis-region, or disease mutation and its respective rare variant variable were both modified by the same modifier SNP. For instance, *ABCG8 p.Asp19His* allele carriers had a 1.9-fold increase in cholelithiasis risk. This risk was modified by rs6729865, an eQTL for *ABCG8* in liver tissue located in the close *ABCG5* and *ABCG8* 3' UTR ($P_{interaction}=4.5 \times 10^{-8}$ in UKB and 0.001 in MVP), which elevated the risk to 3.5 or 5.8-fold in participants carrying 1 or 2 A alleles at rs6729865, respectively. Also, we observed a significant modification of rs35445101 on the association between *LTA p.Thr60Asn* and celiac disease risk ($P_{interaction}=1.4 \times 10^{-14}$ in UKB and 1.1×10^{-31} in MVP). This missense mutation, located 100Kbp away on the *HLA* gene, posed a 1.5-fold or 2.4-fold higher risk for celiac disease, depending on whether the individual carried the G allele for rs35445101. Our findings underscore the potential of large-scale biobanks in identifying genetic modifiers, contributing to clinical diagnosis and prognosis, and paving the way for novel therapeutic interventions.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1417 Decoding pharmacogenomics of second-generation antipsychotic-Induced metabolic syndrome via biobanks and electronic health records.

Authors:

N. El Rouby[1], A. Owusu-obeng[2], M. Preuss[3], R. Nadukuru[4], S. Lee[2], S. L. Van driest[5], W-W. Hsu[1], G. Zhang[6], J. D. Mosley[5], M. DelBello[1]; [1]Univ. of Cincinnati, Cincinnati, OH, [2]Icahn Sch. of Med. at Mt Sinai, New York, NY, [3]Mount Sinai Sch. of Med., New York, NY, [4]Icahn Sch. Of Med. At Mount Sinai, New York, NY, [5]Vanderbilt Univ. Med. Ctr., Nashville, TN, [6]Cincinnati Children's Hosp., Cincinnati, OH

Second Generation Antipsychotics (SGA) are the mainstay treatment of bipolar and psychotic disorders; however, they are associated with metabolic adverse effects. We investigated the genetic determinants of SGA-induced metabolic syndrome (Mets) through a genome wide association study (GWAS) in the UK Biobank with validation in BioMe, Mt Sinai's electronic health record (EHR)-linked biobank data. We created an adult dataset (N=668 controls, 650 cases) for SGA-induced Mets using patient-reported medication use, lab values and ICD codes among European ancestry participants in the UK Biobank data. A case was a patient taking a prescription for an SGA and who met three out of five of the National Cholesterol Education Program (NCEP) criteria: 1) blood pressure ≥ 30/85 mmHg, antihypertensive use or hypertension diagnosis; 2) fasting serum glucose ≥100 mg/dL, antidiabetic medication use or type 2 diabetes diagnosis; 3) serum triglycerides ≥150 mg/dL; 4) HDL-cholesterol < 40 mg/dL in men, and < 50 mg/dL in women or antihyperlipidemic medication use; and 5) body mass index (BMI) ≥ 30 kg/m2. A control was an SGA user who met two or fewer criteria. We conducted a GWAS analysis of imputed genotypes using Scalable and Accurate Implementation of Generalized mixed model (SAIGE), and adjusting for age, sex, BMI, diabetes, and ten principal components of ancestry. We validated the top genomic locus using a meta-analyzed data of White, African American and Hispanic patients in BioMe (N=638 controls, 1825 cases). We identified an association at the *CHD2-RGMA* locus on chromosome 15 that met a suggestive level of significance ($p < 10^{-5}$). The lead SNP was rs12914956 (15:93504501G>A) with the A allele being associated with increased risk for SGA-Mets ($\beta=0.5$, $p= 3.6x10^{-7}$), with a frequency of 39%. According to GTEx, the rs12914956 SNP is an eQTL for *CHD2* in cultured fibroblasts. A different SNP (rs137889996) in the *CHD2-RGMA* region was nominally associated with SGA-induced at $p=6x10^{-5}$ in BioMe. The rs137889996 SNP is in low Linkage Disequilibrium with rs12914956 (R2= 0.0011; D'= 0.57) and was more common in African American with a frequency of 1%. The *CHD2-RGMA* locus has been associated with appendicular lean mass, BMI, insulin levels and high-density lipoprotein levels. In summary, we identified a novel association in *CHD2-RGMA* locus for SGA-induced Mets in the UK Biobank, with nominal association in BioMe. This region has documented association with multiple metabolic phenotypes. Replication and meta-analysis are underway in a third EHR-linked Biobank (BioVU).

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1418 † Decoding the genetic complexity of immunoglobulin heavy chain locus: Unveiling novel variants and structural complexity with implications for human health

Authors:

O. Rodriguez, K. Shields, C. Silver, S. Schultze, M. Smith, C. Watson; Univ. of Louisville Sch. of Med., Louisville, KY

Genetic variants within the immunoglobulin loci (IG) have been implicated in numerous human diseases and have a significant impact on the efficacy of broadly neutralizing antibodies. Our recent study demonstrated a direct influence of these genetic variants within the IG heavy chain locus (IGH) on the composition of the antibody repertoire. However, the repetitive nature of the IG loci, characterized by segmental duplications, simple repeats, and retrotransposon elements, has posed challenges in the sequencing and assembly of these loci across diverse populations in large cohorts. To address this complexity, we employed a combination of targeted long-read sequencing and a bioinformatics tool called IGenotyper, enabling us to comprehensively characterize single-nucleotide variants (SNVs), insertions and deletions (indels), and structural variants (SVs) in a haplotype-specific manner. In this study, we leveraged our approach to analyze a cohort of 860 healthy individuals representing 24 distinct populations. Through our analysis, we discovered several previously unknown large SVs that have a significant impact on the number of genes within the IGH locus. For example, we identified a 284 Kb deletion deleting 16 out of 49 (33%) IGH variable genes. Additionally, leveraging haplotype-resolved assemblies, we identified a substantial number of novel SNVs within segmental duplications and accurately determined the allele frequency of known SNVs with missing or skewed allele frequencies. These findings contribute to a deeper understanding of the genetic architecture of the IGH locus and provide insights into the potential implications for human health and disease. Our results underscore the importance of considering the genome complexity of the IG loci when investigating their association with diseases and antibody mediated phenotypes. The combination of long-read sequencing and IGenotyper provides a powerful toolset for dissecting the intricate variations within this region, thereby enabling more precise genotype-phenotype understanding and offering insights into disease susceptibility and therapeutic interventions. In conclusion, our study presents a comprehensive characterization of genetic variants within the IGH locus. We anticipate that these findings will enhance our understanding of the IG loci's role in human diseases and antibody function, ultimately contributing to personalized medicine approaches.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1419 Deconvoluting immunological, genetic, and clinical heterogeneity across autoimmune rheumatic diseases by cohort-wide immuno-phenotyping

Authors:

Y. Okada[1], H. Tanaka[2], S. Nakayamada[2], Y. Miyazaki[2], K. Sonehara[3], S. Namba[4], S. Honda[5], Y. Shirai[4], K. Yamamoto[4], K. Ikari[5], M. Harigai[5], K. Sonomoto[2], Y. Tanaka[2]; [1]The Univ. Tokyo / Osaka Univ. / RIKEN, Tokyo, Japan, [2]Univ. of Occupational and Environmental Hlth., Kitakyushu, Japan, [3]the Univ. of Tokyo, Tokyo, Japan, [4]Osaka Univ. Graduate Sch. of Med., Osaka, Japan, [5]Tokyo Women's Med. Univ., Tokyo, Japan

Deconvolution of immunological and clinical heterogeneity across autoimmune rheumatic diseases (AIRDs) is essential towards personalized medicine. Here, we conducted large-scale and cohort-wide immuno-phenotyping of 46 peripheral immune cells using the Human Immunology Protocol of comprehensive 8-color flow cytometric analysis. The dataset consisted of >1,000 Japanese patients of 11 AIRDs with deep clinical information registered at the FLOW study, including rheumatoid arthritis (RA), systemic lupus erythematosus (SLE), systemic sclerosis (SSc), ANCA related vasculitis (AAV), idiopathic inflammatory myopathy (IIM), psoriasis, IgG4 related disease (IgG4RD), mixed connective tissue disease (MCTD), ankylosing spondylitis (AS), Sjogren's syndrome (SjS), and giant cell aortitis (GCA). Multimodal clustering of immuno-phenotypes deciphered underlying disease-cell type network, providing novel immune cell type specificity shared or distinct across AIRDs (e.g., close immunological network between MCTD and SLE. Individual patient-level clustering deconvoluted the AIRD patients into several clusters with different immunological features. Of these, RA- or SLE-like clusters were exclusively dominant, showing immunological polarization between RA and SLE across AIRDs. By adopting RA as a flagship disease, in depth clinical analysis revealed that such patient clusters differentially defined clinical heterogeneity in disease activity and treatment responses, such as decreased regulatory T cells and treatment resistance in the RA patients with SLE-like immuno-phenotypes. Inborn human genetics represented by polygenic risk score (PRS) showed associations with dynamics in clinical manifestations and immuno-phenotypes. The PRS estimated based on the RA case-control genome-wide association study (GWAS) and within-case stratified GWAS were associated with patient characteristics, disease activity, and immune-phenotypes such as dendritic cells for RA-interstitial lung disease. Our study demonstrated a value of cohort-wide and cross-disease immuno-phenotyping to elucidate clinically heterogenous patient subtypes existing within the single disease in an immune cell type-specific manner.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1420 *Decrypting the Colon*: Using single colon crypts to study regional somatic mosaicism

Authors:

L. Hiatt, J. Kunisaki, S. Lulla, M. E. Goldberg, A. Hoffman, A. Uchida, X. Nie, Y. Guo, K. Aston, J. Hotaling, A. Quinlan; Univ. of Utah, Salt Lake City, UT

Somatic mosaicism arises from mutations accumulated in cells over an organism's lifespan and can play a pivotal role in disease pathogenesis, from developmental syndromes to cancer. Somatic mosaicism influences fitness and disease predisposition in healthy tissue, and recent experimental advances enable the exploration of mosaicism in healthy organs and throughout disease progression.

Given its uniquely monoclonal epithelial structure, the colon is an exemplary organ for studying somatic mosaicism. Somatic mutations are known to drive colorectal cancer and are hypothesized to contribute to or correlate with inflammatory bowel disease. These diseases impose significant morbidity and mortality, with colorectal cancer ranking as the second leading cause of cancer-related deaths in the United States and inflammatory bowel disease affecting one in 200 individuals of European ancestry.

Interestingly, distinct regional pathologies exist along the colon's length, such as "right" and "left" colorectal cancer and unique subtypes of inflammatory bowel disease. Yet, the genetic origins of these region-specific pathologies remain elusive due to previous experimental limitations. Assessing somatic mosaicism throughout the colon is essential to explore whether local mutagenic events, such as developmental processes or microbiome exposures, drive regional pathogenesis. A preliminary study by Lee-Six et al. reported variation in mutational signature proportions across three colon regions, as determined by biopsy metadata. However, low sequencing coverage (~15x), a lack of germline filtering, and the examination of a small number of non-standardized colonic subregions in a small cohort limited this analysis.

To systematically investigate the mutational landscape of distinct colorectal regions, we are studying cadaveric tissue and colonoscopy samples from the University of Utah Gastrointestinal Biobank taken from ten standardized regions along the colon. Utilizing laser capture microdissection, we isolate single colon crypts, which are monoclonal structures whose ~2000 cells originate from a common ancestral stem cell. By detecting mutations in different crypts with whole-genome sequencing, we can test the hypothesis that distinct mutagenic processes influence regional pathogenesis.

We will present our ongoing efforts to characterize regional somatic mosaicism in the colon, as this knowledge holds promise for innovative strategies in predicting disease risk, optimizing disease management, and advancing personalized therapeutic approaches.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1421 Deep learning based semantic segmentation of adipose tissue histology samples.

Authors:

E. Joersboe[1], J. Honecker[2], P. kubitz[3], M. Claussnitzer[4], H. Hauner[5], M. Blüher[6], P. D. Charles[1], C. Lindgren[7], C. Nellaker[8]; [1]Univ. of Oxford, Oxford, United Kingdom, [2]Technical Univ. of Munich, Munich, Germany, [3]Technical Univ. of Munich, Freising, Germany, [4]Massachusetts Inst Technology, Cambridge, MA, [5]Technical Univ. of Munich, Freising-Weihenstephan, Germany, [6]Univ. of Leipzig Med. Ctr., Leipzig, Germany, [7]Big Data Inst., Oxford Univ., Oxford, United Kingdom, [8]Univ Oxford, Oxford, United Kingdom

Fat distribution and macro structure of adipose tissue are important factors in predicting obesity-associated diseases. To investigate the relationship between mean adipocyte area and obesity-related traits, and identify genetic factors associated with adipocyte cell size, we performed the largest study of automatic adipocyte phenotyping using both histological and genetic data. Utilising visual deep learning based tools for doing semantic segmentation of subcutaneous and visceral adipose tissue histology samples (N=2,667) in 5 independent cohorts. We developed a pipeline using U-net for segmentation of adipocytes, and filtered on size (> 200 µm2) and roundness, using the Polsby-Popper test, of the segmented adipocyte. The pipeline was able to do rapid inference across 9,000 whole slide images, and segment over 27 million adipocytes. Mean adipocyte area in both depots is positively correlated with body mass index (BMI) (PSubcutaneous = 4.56·10-54, βSubcutaneous = 0.0020, R2Subcutaneous = 0.55, PVisceral = 5.90·10-31, βVisceral = 0.0019, R2Visceral = 0.47) using a linear model, adjusting for cohort. Additionally, the mean adipocyte area in subcutaneous depots is larger than their visceral counterparts using a simple linear model (P = 4.79·10-132 , β = 0.46, R2 = 0.27). Estimates of mean size of adipocytes were validated against Glastonbury et al. 2020 showing a high level of concordance (R2Visceral = 0.95 , R2Subcutaneous = 0.87). As expected from our novel phenotyping approach, where we segment all adipocytes in the whole slide image, and our new filtering approach for segmented adipocytes, we find a larger estimated mean size of adipocytes in the current study than Glastonbury et al. 2020 (βVisceral = 0.60, βSubcutaneous = 0.56). We performed the largest GWAS (NSubcutaneous = 2066, NVisceral = 1878) and subsequent meta-analysis of mean adipocyte area, however no genome-wide significant (P < 5·10-8) associations were found with mean subcutaneous or visceral adipocyte area, using imputed genetic data (almost 9 million markers with R2/imputation info > 0.3).

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1422 Defining cognitive phenotypes in Peurto Rican individuals for genetic studies using culturally sensitive norms

Authors:

A. Zaman[1], K. McInerney[1], P. Mena[1], S. Tejada[1], K. Celis[1], L. Adams[1], K. Hamilton-Nelson[1], C. Silva[1], V. Glenies[1], J. Sanchez[1], J. Vance[1], B. Feliciano-Astacio[2], M. Pericak-Vance[1], M. Cuccaro[1]; [1]Univ. of Miami, Miami, FL, [2]Univ. Celtral del Caribe, Bayamon, PR

Background: Cognitive testing is used to define phenotypes in genetic studies of Alzheimer's disease (AD). To evaluate cognitive test performance, normative (norm) data are used to adjust for demographics. For example, the National Alzheimer's Coordinating Center Uniform Data Set (NACC) is a nationally representative US dataset (83% White, 14% African American, 3% Other) with norms used to adjust and standardize scores. While these adjustments increase the sensitivity to detect cognitive impairment, they need to be adapted for diverse populations. This pilot study examines norm differences between individuals from the Puerto Rico Alzheimer Disease Initiative (PRADI) and NACC.
Methods: We restricted our PRADI dataset (N=1280) to 141 adults (11%) who were cognitively normal (CN), had a global CDR score = 0, and had no medical conditions of concern (76% Female, 13.6 +3.6 years of education, 70.5 + 10.0 years old). The NACC sample is mostly White and of European ancestry; the PRADI sample is admixed with European (71%), African (18%), and Amerindian (11%) contributions. For all cognitive tests, norms were estimated (population mean, sex, age, and years of education). We then compared the PRADI norms with NACC norms.
Results: The population mean adjustments (PMA) were higher in NACC compared to PRADI, reflecting lower test scores in the PRADI sample. The difference in PMA was large for the Craft Story Immediate (10.2) and Delayed (10.9) tasks and small for Category Fluency [Animal Naming (AN) 0.3; Vegetable Naming (VN) 2.8]. Age adjustment differences were small (<0.1) on all tests. Education adjustments were notably larger in the NACC than PRADI for AN (0.57 vs 0.34) and VN (0.31 vs 0.12) indicating that each educational year was associated with a higher predicted score in the NACC vs. PRADI. Notable sex adjustments were found for AN and VN, where adjustment for AN was large in PRADI where females scored lower than males (-1.3), while the adjustment was smaller and in the opposite direction in NACC (0.3). The sex adjustments on VN were smaller in PRADI where females scored slightly higher than males (0.3), while NACC females had much higher scores than males (2.5).
Discussion: PMAs were greater in the NACC compared to PRADI, and the overall result of using NACC norm adjustments in the PRADI sample could result in presumptively CN individuals showing more impairments on multiple cognitive tests. This suggests that the local PRADI norms may more accurately estimate cognitive performance than the NACC norms in the PRADI group. Culturally sensitive norms should be developed for cognitive tests to measure and define cognitive phenotypes for genetic studies of AD in diverse populations.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1423 Delineation of functionally important protein domains of CHD8 for assessment of variants identified in individuals with autism spectrum disorders

Authors:

E. Larsen[1], S. Spring-Pearson[1], C. Printzis[2], P. Gerges[1], S. Basu[1]; [1]MindSpec, Inc., McLean, VA, [2]Northwestern Univ., Niles, IL

CHD8, a member of the chromodomain-helicase-DNA binding protein family, is considered a high confidence gene for autism spectrum disorder (ASD) based on the identification de novo damaging variants in individuals with a primary diagnosis of ASD across studies and cohorts. This gene has also been shown to contribute to a neurodevelopmental disorder (NDD) presenting with intellectual disability, overgrowth, sleep disturbances, gastrointestinal problems, and dysmorphic features with or without ASD. However, it remains unclear if any correlation exists between the type and location of CHD8 variants and any specific pattern of phenotypes arising from these mutations. In order to delineate functionally-relevant protein domains, we assessed the genetic landscape of CHD8-associated syndrome using the richly annotated data curated in AutDB, our online genetic database for ASD (http://autism.mindspec.org/autdb/Welcome.do). Rare genetic variations in CHD8 identified in individuals diagnosed with ASD or NDD were extracted from 45 peer-reviewed scientific articles and annotated with multiple attributes for ongoing curation in AutDB. We assembled the largest CHD8 dataset to date comprising a total of 204 individuals carrying rare variants in CHD8, including 144 of whom were reported to have a diagnosis of ASD. Randomization across protein domains was performed on the standardized CHD8 rare variant dataset, as well as on a control dataset of CHD8 variants from gnomAD, to identify protein domains statistically enriched for variations linked to ASD and/or NDD. While disease-associated variants were distributed across the entire length of the CHD8 protein, randomization analysis demonstrated that three identified protein domains (the helicase ATP-binding, the helicase C-terminal, and the chromodomain 2 domains) showed statistically significant enrichment for ASD-associated variants. While no significant findings were observed for NDD-associated CHD8 variants in any identified protein domain, we observed that both the helicase ATP-binding and helicase C-terminal domains had significantly fewer than expected variants in the control cohort. Moreover, CHD8 missense variants in both the ASD and NDD cohorts displayed significantly higher CADD scores than missense variants from the control cohort. Our approach of protein domain-based enrichment of variants improves the understanding of functional consequences of CHD8 variations associated with ASD. Importantly, protein-domain mapping of variants will likely facilitate the interpretation of new CHD8 variants reported in large-scale genetic studies for diagnostic purposes.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1424 Determinants of clonal hematopoiesis expansion rate in *UK Biobank* and*All of Us* Cohorts

Authors:

T. Mack[1], C. Vlasschaert[2], B. Sharber[1], A. Bick[1]; [1]Vanderbilt Univ., Nashville, TN, [2]Queens Univ., Kingston, ON, Canada

Clonal Hematopoiesis of Indeterminate Potential (CHIP) is characterized by the acquisition of a somatic mutation in a hematopoietic stem cell that leads to the formation of a sub-clonal blood cell population. CHIP increases risk of blood cancer, cardiovascular disease, and overall mortality. Larger CHIP clones are associated with worse outcomes. However, CHIP clonal expansion rate is not well characterized because few cohorts have multi-timepoint DNA samples. To date, the largest of such analyses include a few hundred participants. To overcome this limitation, we recently developed a method to quantify clonal expansion rate from a single timepoint whole genome sequenced blood sample termed Passenger-Approximated Clonal Expansion Rate (PACER; Weinstock, Nature 2023). Here we identified 10,727 individuals with CHIP in 298,477 individuals with available whole genome sequence data in All of Us (CHIP N=4,446) and UK Biobank (CHIP N=6,281). We applied PACER to these two datasets to estimate the clonal expansion rate of each CHIP clone. We first identified the contribution of somatic driver mutation to clonal expansion. Consistent with prior observation from smaller multi-timepoint datasets, CHIP mutations in *JAK2* and splicing factors (*U2AF1, SRSF2, & SF3B1*) showed the fastest rates of clonal expansion. *DNMT3A* CHIP mutations were amongst the slowest expansion rates. Next, we identified the contribution of germline genetic variation to clonal expansion. We performed a single variant genome-wide association study (GWAS) of PACER in individuals with CHIP clonal expansion rate in each cohort and performed a fixed effect meta-analysis of the results. Covariates included age, sex, and the first ten genetic ancestry principal components. We identified 9 genome-wide significant loci. We robustly replicated TCL1A (rs2887399, $p= 3.5 \times 10^{-9}$), the only variant previously identified in our NHLBI TOPMed cohort PACER GWAS. Notable novel variants associated with PACER at genome-wide significance include intronic variants in CEP135, a centrosomal scaffold protein (rs73155551, $p=5.3 \times 10^{-9}$); MAU2(rs117352725, $p=3.5 \times 10^{-8}$) involved in cohesin loading and maintenance of mitotic sister chromatids; and DLEU7 (Deleted in Lymphocytic Leukemia 7, rs752066930, $p=1.2 \times 10^{-8}$). In summary, here we perform the largest analysis of CHIP clonal expansion to date, twice the size of our prior effort in TOPMed and >100 times larger than any analysis of multi-timepoint samples. Our findings highlight the power of PACER to estimate CHIP clonal expansion rate from a single timepoint sample and identify a previously unknown link between cell cycle progression and clonal expansion.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1425 Determining high-risk status using multiple types of genetic risk: Preliminary results from the eMERGE study.

Authors:

J. Linder[1], N. Lennon[2], W. K. Chung[3], C. Liu[3], C. Weng[3], C. A. Prows[4], J. Connolly[5], H. Hakonarson[5], M. Harr[5], L. Orlando[6], N. S. Abul-Husn[7], E. E. Kenny[8], S. Aguilar[9], M. Lebo[10], E. Perez[11], R. Freimuth[12], T. A. Manolio[13], R. Rowley[14], A. Gordon[15], E. M. McNally[16], M. J. Roy-Puckelwartz[15], D. R. Crosslin[17], J. Cortopassi[18], N. Limdi[18], G. P. Jarvik[19], J. Morse[1], D. M. Roden[1], D. Velez Edwards[1], eMERGE Network, J. F. Peterson[1]; [1]Vanderbilt Univ. Med. Ctr., Nashville, TN, [2]Broad Inst. of MIT and Harvard, Cambridge, MA, [3]Columbia Univ., New York, NY, [4]Children s Hosp Med Ctr, Cincinnati, OH, [5]Children's Hosp. of Philadelphia, Philadelphia, PA, [6]Duke Univ., Durham, NC, [7]Icahn Sch. of Med. at Mt. Sinai, New York, NY, [8]Icahn Sch. of Med. at Mount Sinai, New York, NY, [9]Invitae, San Francisco, CA, [10]Mass Gen. Brigham Lab for Molecular Med., Cambridge, MA, [11]Mass Gen. Brigham, Boston, MA, [12]Mayo Clinic, Rochester, MN, [13]Natl. Human Genome Res. Inst., Bethesda, MD, [14]NIH, Bethesda, MD, [15]Northwestern Univ., Chicago, IL, [16]Northwestern Univ, Chicago, IL, [17]Tulane Univ. Sch. of Med., Seattle, LA, [18]Univ. of Alabama at Birmingham, Birmingham, AL, [19]Univ Washington Med Ctr., Seattle, WA

Screening populations for genetic risks includes obtaining family health history (FHH) and testing for monogenic and polygenic diseases. As genomic screening becomes more routine, clinicians will need guidance on how these risks interact, and how to integrate them with clinical risk factors to establish risk-appropriate preventive care practices. The eMERGE network is enrolling a diverse 25,000-person cohort, ascertaining FHH, clinical sequencing for a limited number of monogenic risks (adults only), and genotyping for polygenic risk scores (PRS). We developed a REDCap based platform (R4: Recruitment, Retention, and Risk Reduction) to collect clinical-grade data from multiple sources, apply logical rules to trigger ‘high risk’ status, and generate a comprehensive risk report for clinicians, the electronic health record (EHR), and individuals. These data were obtained directly from individuals (survey), EHRs, clinical laboratories conducting genotyping and sequencing, and a family history data collection tool, MeTree. For each condition, a specific set of rules was developed based on clinical guidelines or expert opinion to trigger ‘high risk’ status. Three conditions (atrial fibrillation, coronary heart disease, prostate cancer) use PRS, limited monogenic, and FHH risk to independently trigger high risk status; chronic kidney disease uses PRS and FHH; hypercholesterolemia uses monogenic and PRS; and four conditions (asthma, obesity, type 1 diabetes, and type 2 diabetes) rely on the PRS score alone. Breast cancer uses a published integrated risk score, BOADICEA, which calculates lifetime risk based on FHH, clinical, and genomic data; high risk status is defined as ≥ 25% lifetime risk. Of the 11,897 individuals enrolled as of June 1st, 2023, 1009 risk reports have been generated, with 285 individuals classified as ‘high risk’; 193 with PRS above the study threshold, 18 with monogenic risk, and 139 with high risk FHH. Eleven individuals were assigned high breast cancer risk using BOADICEA. Individuals may have multiple risk categories and conditions. During R4 report generation testing, challenges were encountered specifically around missingness in survey and FHH data. Protocols to review, update, and regenerate reports and associated recommendations were developed. As the field of population genomic screening evolves, it will be critical to establish reliable and validated methods for capturing, processing, and returning genetic risks that are integrated and delivered in the context of an individual’s health while resilient to incomplete data about individuals’ personal and family history.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1426 Differential gene expression analysis reveals genes underlying the transition to T2D in Hispanic/Latino individuals.

Authors:

E. Frankel[1], W. Zhu[1], H-H. Chen[2], P. Sharma[3], L. E. Petty[1], R. Roshani[1], H. Highland[4], E. Gamazon[5], J. McCormick[6], S. Fisher-Hoch[7], K. North[8], J. Below[9]; [1]Vanderbilt Univ., Nashville, TN, [2]Academia Sinica, Taipei, Taiwan, [3]Univ. of North Carolina, Chapel Hill, NC, [4]Univ North Carolina at Chapel Hill, Chapel Hill, NC, [5]VUMC Clare Hall, Univ. of Cambridge, Nashville, TN, [6]UTH SPH Brownsville, Brownsville, TX, [7]Univ. of Texas Sch. of Publ. Hlth., Brownsville, TX, [8]Univ North Carolina, Wake Forest, NC, [9]Vanderbilt Univ Med Ctr., Nashville, TN

Background: Type 2 diabetes (T2D) is a multifactorial metabolic disorder that results from the interactions between genetic, environmental, and multiomic factors. Hispanic/Latino populations are often underrepresented in research studies despite having an increased prevalence of T2D compared to non-Hispanic whites. In this study, we sought to understand the changes in gene expression that occur during the transition to T2D and in quantitative measures such as longitudinal fasting blood glucose (FBG) and glycated hemoglobin (HbA1c) levels. **Methods:** We utilized 1150 longitudinal whole blood specimens from 575 participants (142 incident cases) in the Cameron County Hispanic Cohort. RNA-sequencing data from two time points per individual were used to profile the transition to incident T2D. Incident T2D was characterized using the ADA 2010 criteria, and we excluded individuals taking diabetes medication at either timepoint. We utilized joint mixed-effect regression models of longitudinal measures, accounting for sex, age, smoking status, body mass index (BMI), and probabilistic estimation of expression residual factors to compare changes in gene expression between individuals with incident T2D and controls, as well as in the quantitative traits FBG and HbA1c. We performed an additional analysis removing BMI as a covariate for incident T2D to capture genes highly associated with BMI that may contribute to the development of T2D. **Results:** Our preliminary analysis of incident T2D from a subset of samples revealed three genes independent of BMI that had a significant change in expression across T2D onset after false discovery rate (FDR) correction: *KLF10*, *BPI*, *OR2T2*. Excluding BMI as a covariate revealed six additional significant genes: *TP53INP2*, *PRICKLE2*, *CAMKK2*, *HP*, *MMP8*, and *CEACAM8*. Furthermore, preliminary analyses of longitudinal FBG and HbA1c levels identified 50 and 398 significant genes, respectively, after FDR correction. We further explored our significant genes in proteomic data from 259 CCHC individuals and found MMP8 protein levels are also associated with incident T2D (P=0.0018). **Conclusion:** Our analyses characterized transcriptomic changes related to incident T2D and longitudinal FBG and HbA1c levels in an understudied population with a high disease burden. Numerous genes discovered in these analyses have previously been identified in genome-wide association studies of T2D and other metabolic health and dysfunction measures. Our results provide functional evidence from a key timepoint in disease development that implicates a functional gene at these loci and a mechanism of effect.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1427 Differential gene expression analysis using RNA-Seq sheds light on the molecular etiology of chikungunya's chronic manifestations

Authors:

J. Noirel[1], R. Medina-Santos[1], S. Le Clerc[1], T. Labib[1], M. Rahmouni[1], M. J. Mathew[2], H. Hocini[3], L. Bruneau[4], L. Bruneau[4], A. Maillot[4], C. Fontaine[4], C. Fontaine[4], N. Ah-You[4], M. Marimoutou[4], P. Gérardin[4], J-F. Zagury[1]; [1]CNAM, Paris, France, [2]EFREI, Paris, France, [3]Hôpital Henri Mondor, INSERM, Créteil, France, [4]CHU Réunion, Saint-Pierre, France

Background: Chikungunya virus is responsible for a complex disease including fully asymptomatic, acute/post-acute self-limiting, and chronic incapacitating presentations. In the long run arthralgia/arthritis, musculoskeletal pain, fatigue and mental illness are observed in up to 52 % of the infected patients. In 2005-2006, a huge chikungunya outbreak hit La Reunion Island, which left 300,000 people infected and countless with chronic discomfort.
Methods: In the aim to better understand the molecular pathogenesis of the disease and the genetic and immunologic drivers of its chronicity, we conducted a genome-wide scan of a well-phenotyped cohort of 609 individuals from La Reunion Island, within which 244 adult patients were primarily classified as “asymptomatic” (none of the abovementioned manifestations), “nonchronic” (symptoms ≤ 3 months), or “chronic” (symptoms more than 3 months), and sampled for peripheral blood mononuclear cells (PMBCs) transcriptome analysis. We next identified differentially expressed genes (DEGs) and interactions associated with symptomatic status and with chronic symptoms through pathway and gene ontology (GO) analyses in comparing phenotypes (chronic vs nonchronic, asymptomatic vs nonchronic).
Results: In total, patients with chronic symptoms represented 55% of the subsample (136/244). Women were over-represented in this group (109 for 27 men), and to a lesser extent among patients with nonchronic symptoms (58 for 30 men) and asymptomatic individuals (17 for 3 men). DEGs included TMTC1, GSTM1, FRG1CP, CXCL16, RPL10P9, RPL10P6, MEGF6, IRGM and NBPF26, some of these being involved in viral infection susceptibility and/or severity, immune response and muscular diseases. The top gene pathways enriched according to GO deciphered mainly immune system processes, such as “complement activation, classical pathway” and “protein activation cascade”, while the KEGG pathways identified “Extracellular-Matrix receptor interaction” and “bone remodeling” as prominent mechanisms, which was consistent with shorter term RNA-Seq analyses of chikungunya virus infection.
Conclusion: This first RNA-Seq study of post-chikungunya chronic manifestations, performed on a large sample size, provides valuable insights into the molecular mechanisms underlying the complex chronic manifestations of chikungunya.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1428 Discovering genes associated with Alzheimer's Disease via multi-tissue and cell type Transcriptome-Wide Association Study.

Authors:

C. Liu[1,2], H. Sun[3], The FunGen-xQTL Consortium, G. Wang[3], F. Morgante[1,4]; [1]Ctr. for Human Genetics, Clemson Univ., Greenwood, SC, [2]Sch. of Computing, Clemson Univ., Clemson, SC, [3]Dept. of Neurology, Columbia Univ., New York, NY, [4]Dept. of Genetics and Biochemistry, Clemson Univ., Clemson, SC

Late-onset Alzheimer’s disease (AD) is a complex neurodegenerative disorder with a substantial genetic component unveiled through Genome-Wide Association Study (GWAS). Identifying genetic risk factors and understanding the mechanisms of their role in complex diseases remains challenging, particularly for non-coding variants which, on average, consist of >90% GWAS signals. Transcriptome-Wide Association Study (TWAS) has been developed to leverage expression Quantitative Trait Loci (eQTL) for gene mapping in GWAS, by imputing gene expression in GWAS samples using pre-trained models from eQTL studies and testing for their disease associations. Such imputed expression data aggregates functional evidence of non-coding variants, thus improving both the statistical power and biological interpretability of GWAS.

Here, we develop a new TWAS resource by training tissue and cell-specific models from multi-brain region RNAseq harmonized across several AD cohorts by the FunGen-xQTL Consortium. We apply a new Bayesian method, mr.mash, to exploit shared genetic effects across multiple tissues and cell types. Using a data-adaptive approach, mr.mash can improve the accuracy of expression imputation in conditions with inadequate power for TWAS --- such as rare cell types identified through single-nuclei RNAseq or brain regions with small sample sizes --- for genes with shared effects across multiple conditions. It will, nonetheless, retain tissue and cell-specific effects for genes uniquely regulated under specific conditions.

We apply this new TWAS resource to a recently published AD GWAS meta-analysis. To avoid false positives due to horizontal pleiotropy, we adopted the causal TWAS (cTWAS) framework to simultaneously identify putative causal genes and variants. Through cohort-specific analyses and cross-cohort meta-analysis, we report a list of genes with strong TWAS signals, including established AD genes such as APOE and PICALM, as well as novel genes that either exhibit consistent effect across conditions, or are associated with AD only in tissue and cell-specific fashion. Our results provide a powerful TWAS analysis framework and pre-trained models in human brains that can be applied to any other relevant complex traits of interest to improve our understanding of complex trait genetics.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1429 Discovery of microglia specific molecular drivers and targeted therapeutics for Alzheimer's disease using network-based deep learning integration of human brain single-cell RNA seq data and real-world patient data validation

Authors:

J. Xu[1], M. Danziger[2], Z. Xu[3,4], C. Zang[3,4], Y. Zhou[1], Y. Hou[1], E. Karavani[2], Y. Shimoni[2], F. Wang[3,4], M. Rosen-Zvi[2], F. Cheng[1,5,6]; [1]Cleveland Clinic, Lerner Res. Inst., Cleveland, OH, [2]AI for Hlth.care, IBM Res., Haifa, Israel, [3]Dept. of Population Hlth.Sci., Weill Cornell Med., New York, NY, [4]Inst. of Artificial Intelligence for Digital Hlth., Weill Cornell Med., New York, NY, [5]Dept. of Molecular Med., Cleveland Clinic Lerner Coll. of Med., Cleveland, OH, [6]Case Comprehensive Cancer Ctr., Case Western Reserve Univ. Sch. of Med., Cleveland, OH

Background: Microglia have been implicated in Alzheimer's disease (AD) for over 100 years. However, high microglial heterogeneities, including disease-associated microglia (DAM), tau microglia, and inflammatory microglia (neuro-toxic), largely limit the development of microglia-targeted treatment for AD and other neurodegenerative diseases.**Methods:** In this study, we conducted integrative analyses of ~0.7 million single-cell/nuclei RNA-seq transcriptomes derived from AD patient frozen brain samples across different brain regions at The Alzheimer's Cell Atlas (https://taca.lerner.ccf.org) using a variational autoencoder-based approach scVI. We used the trajectory analysis to identify microglial subtypes across AD progression, including DAM, tau and inflammatory microglia. We conducted the transition network analysis to identify potential molecular drivers of different microglial subtypes across varying degrees of AD and disease progression under the human protein-protein interactome network. We prioritized candidate drugs by specifically targeting molecular derivers using drug-gene signature enrichment analysis from the Connectivity Map (CMap) database and population-based drug-disease outcome analysis using with two independent real-word patient data (MarketScan insurance claims [172 million insured individuals] and INSIGHT Clinical Research Network (15 million patients).**Results:** We identified that tau microglia were significantly associated with synaptic processes. Compared to DAM, upregulated genes of inflammatory microglia were more significantly enriched with key immune pathways (e.g., toll-like receptor and th17 cell differentiation). We found that transition gene networks of inflammatory microglia and DAM contain potential AD pathobiology regulators (e.g., *SYK*, *CTSB*, and *PRKCA*) and genetic risk genes (including *INPP5D*, *BLNK*, and *ADAM10*). We further conducted network-based drug repurposing prediction by specifically targeting inflammatory microglial subpopulations. Consistent within real-world patient data observation in two independent clinical databases (i.e., MarketScan and INSIGHT CRN), we identified that usage of ketorolac could potentially reduce the risk of AD incidence in both MarketScan (hazard ratio [HR] = 0.81, 95% confidence interval [CI]: 0.69-0.91 after adjusting > 400 covariates); and INSIGHT (HR = 0.52, 95% CI 0.28-0.95, p-value = 0.03, after adjusting 267 covariates) databases. **Conclusion:** This study identifies potential underlying mechanisms of microglial subtypes involved in human AD brains and identified ketorolac as a potential microglia-targeted medicine for AD.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1430 Discovery of structural variants in the CARTaGENE cohort using whole-genome DNA sequencing

Authors:

R. Laflamme[1,2], M. Sayahian Dehkordi[3], D. Taliun[3], K. Lo[1], S. Gravel[3], S. Gagliano Taliun[1,2], G. Lettre[1,2]; [1]Montreal Heart Inst., Montreal, QC, Canada, [2]Université de Montréal, Montreal, QC, Canada, [3]McGill Univ., Montreal, QC, Canada

Structural variants (SVs), such as large deletions and duplications, are genetic variants that encompass 50 DNA base pairs or more. Although they are considered the largest source of genetic diversity, SVs have been less studied compared to smaller variants owing to the difficulty to discover and genotype them. As a result, whereas small SNVs and indels have been associated with a large number of human traits and diseases, associations between SVs and human phenotypes are rare. Our aim is to build the first catalog of SVs in the Quebec population and characterize their role in chronic diseases.The CARTaGENE cohort presents a unique opportunity to assess associations between SVs and human phenotypes. CARTaGENE is a population-based cohort of 43,000 individuals recruited from different metropolitan regions of Quebec (Canada) and aged between 40 and 69 years at recruitment. The aims of our project are to generate a map of SVs from high-coverage short-read whole-genome sequence (WGS) data of 2,173 participants of the CARTaGENE cohort, and to test their associations with clinically-relevant phenotypes implicated in chronic diseases. For this experiment, we sequenced the genome of 1,756, 131 and 163 individuals of French Canadian, Moroccan and Haitian ancestry, respectively. We used existing bioinformatics tools and built a pipeline for calling and genotyping SVs from WGS data. SVs were called in each participant with two algorithms (Manta and DELLY), and the intersection between the two callers were kept. Called SVs from all participants were then merged and genotyped using muCNV, a multi-sample SVs genotyper. Genotyped SVs were then used for association testing on 57 quantitative phenotypes with PLINK.We genotyped 25,835 SVs, including 24,034 deletions and 1,801 duplications. 74% of these SVs were present in the gnomAD-SV database, and their frequencies were concordant (Pearson's r=0.81 and r=0.58 for deletions and duplications, respectively). We identified rare SVs associated with known pathologies, including six carriers of the 15+kb deletion encompassing the promoter and first exon of the *LDLR* gene, which causes familial hypercholesterolemia.Finally, we are developing a haplotype reference panel to empower the imputation of these SVs in CARTaGENE and other cohorts to enable more powerful genetic association studies.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1431 Discovery of SVA-derived transcripts uncovers potential mechanism driving neurodegeneration in ventral forebrain organoid models of X-linked dystonia-parkinsonism.

Authors:

B. Sheehan[1], L. D'Ignazio[1,2], T. Evans[1,2], A. Lorenzetti[1], B. Qamar[1], A. Feltrin[1], A. McCord[1], A. Barbosa[1,2], E. Penney[3,4], B. Araujo[1], H. Giles[5,1], A. Guerrero Zuniga[2,1,6], Y. Wang[1], T. Sawada[1], M. Zabolocki[7], W. Hendriks[8,4], A. Gershman[9], C. Bardy[7], W. Timp[9,10], C. Bragg[3,4], A. Paquola[1,2], J. Erwin[1,11,2]; [1]Lieber Inst. for Brain Dev., Baltimore, MD, [2]Dept. of Neurology, Johns Hopkins Univ. Sch. of Med., Baltimore, MD, [3]Dept. of Neurology, Massachusetts Gen. Hosp. and Harvard Med. Sch., Boston, MA, [4]The Collaborative Ctr. for X-linked Dystonia-Parkinsonism, Massachusetts Gen. Hosp., Charlestown, MA, [5]McKusick-Nathans Inst., Dept. of Genetic Med., Johns Hopkins Univ. Sch. of Med., Baltimore, MD, [6]Biochemistry, Cellular and Molecular Biology (BCMB) Graduate Program, Johns Hopkins Univ. Sch. of Med., Baltimore, MD, [7]South Australian Hlth.and Med. Res. Inst., Adelaide, Australia, [8]Dept. of Neurology, Massachusetts Gen. Hosp. and Harvard Med. Sch., Boston, MA, United States, Boston, MA, [9]Dept. of Molecular Biology and Genetics, Johns Hopkins Univ., Baltimore, MD, [10]Dept. of BioMed. Engineering, Johns Hopkins Univ., Baltimore, MD, [11]Dept. of NeuroSci., Johns Hopkins Univ. Sch. of Med., Baltimore, MD

X-linked Dystonia-Parkinsonism (XDP) is an adult-onset fatal neurodegenerative disorder of the caudate nucleus predominantly affecting male descendants from the Panay Islands in the Philippines. It is characterized by progressive degeneration of the adult neostriatum. At molecular level, a disease-specific, hexamer-containing SINE-VNTR-Alu (SVA) retrotransposon insertion occurs within intron 32 of *TAF1* on the X chromosome, which alters splicing and decreases *TAF1* expression. The length of hexamer repeats is inversely correlated with age of onset, as XDP is also a repeat expansion disorder. Efforts have focused on rescuing altered *TAF1* splicing, however, modeling this hexamer expansion to better study the mechanism by which the SVA contributes to XDP pathogenesis remains unexplored.

Here, to understand the mechanism of degeneration driven by SVA insertion, we differentiated induced pluripotent stem cells (iPSCs) from male control and XDP individuals, and isogenic SVA-deleted lines from the same XDP individuals into ventral forebrain organoids. We cryosectioned and immunostained organoids for striatal markers to ensure our model recapitulated human striatal development. We also monitored electrophysiological activity from day 90 to day 180 on multi-electrode arrays. We found that striatal organoids from individuals with XDP displayed increased somatic instability of the hexamer DNA repeat throughout organoid maturation. The XDP organoids showed decreased electrical activity from day 120 to day 180 and increased apoptosis and DNA damage at day 120 compared to SVA-deleted organoids. We recapitulated neurodegeneration in our organoid model and uncovered potential molecular mechanisms by which the SVA contributes to XDP pathogenesis.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1432 Disentangling the effects of fat mass and fat-free mass on cardiometabolic risk by a genetic approach.

Authors:

X. Jiang[1], M. Christiansen[1], T. Kilpeläinen[2]; [1]Novo Nordisk Fndn. Ctr. for Basic Metabolic Res., Univ. of Copenhagen, Copenhagen, Denmark, [2]Univ. of Copenhagen, Copenhagen, Denmark

Background: The amounts of fat mass (FM) and fat-free mass (FFM) are linked to the risk of cardiometabolic diseases. However, disentangling the independent effects of FM and FFM on cardiometabolic risk is challenging because of the strong correlation between the traits (r=0.6-0.8) and shared heritability. We aimed to disentangle the effects of FM and FFM on cardiometabolic risk by identifying genetic variants specific to FM and FFM and studying their associations with cardiometabolic disease risk. **Method:** We constructed specific polygenic risk scores (PRS) for FM and FFM from variants that were associated with one trait (P<5E-08) but not the other (P>0.05 and P>0.01, respectively). The associations between the PRS and outcomes were estimated in up to 414,377UK Biobank participants whose body composition was measured by bioimpedance. The specificity of the PRS for FM and FFM were independently validated in 35,010 additional UK Biobank participants whose body composition was measured by dual-energy X-ray absorptiometry. We tested associations between the PRS and cardiometabolic traits and diseases in up to 414,377 UK Biobank participants, as well as by using summary statics from published GWAS for coronary heart disease (CHD), type 2 diabetes (T2D), blood glucose, glycated hemoglobin (HbA1c), low-density lipoprotein cholesterol (LDL-C), high-density lipoprotein cholesterol (HDL-C), triglycerides, and systolic and diastolic blood pressure (SBP and DBP). **Results:** We identified 109 specific SNPs for FM and 73 for FFM and constructed the respective PRS. The FM-PRS was strongly associated with FM (?:7.16, P<0.001) and FFM-PRS with FFM (?:-4.84, P<0.001) without showing association with the counterpart (P=0.08, P=0.38). One Kg higher FM was associated with OR 1.25-fold (95%CI 1.10-1.42) higher risk of T2D. Higher FM was also associated with higher HbA1c (P<0.001), SBP (P=0.002), DBP (P<0.001), and lower HDL-C (P<0.001). One Kg lower FFM was not associated with the risk of CHD or T2D (P>0.05). Lower FFM was associated with higher blood glucose (P<0.001), LDL-C (P<0.001), triglycerides (P<0.001), and SBP (P<0.001). **Conclusion:** Through a novel genetic approach to assess the independent effects of FM and FFM on cardiometabolic risk, we find evidence that higher FM increases HbA1C, SBP, DBP and the risk of T2D, and lowers HDL-C, while lower FFM increases blood glucose, LDL-C, triglycerides, and SBP. [2,300 characters, excluding spaces, for the body of the abstract.]

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1433 † Disentangling the link between maternal influences on birth weight and disease risk in 36,211 genotyped mother-child pairs

Authors:

J. Leinonen[1], M. Pirinen[2], FinnGen, T. Tukiainen[3]; [1]FIMM, Univ. of Helsinki, Helsinki, Finland, [2]Univ Helsinki, Helsinki, Finland, [3]Inst. for Molecular Med. Finland (FIMM), Helsinki, Finland

Epidemiological studies have robustly linked lower birth weight to later-life disease risks, including cardiovascular diseases and type 2 diabetes. These observations have been hypothesized to reflect the adverse impact of intrauterine growth restriction on a child's health. However, causal evidence from human data supporting such a mechanism is largely lacking. We assessed the relationship between intrauterine growth restriction and five disease outcomes using 36,211 genotyped mother-child pairs (child mean age = 45.0 years) from the FinnGen study. Using a Mendelian Randomization (MR) framework, we tested how maternal genetic scores for child's birth weight, proxying intrauterine growth restriction, associate with disease in the children, accounting for the direct transmission of alleles between the mother and child. Despite the robust epidemiological associations, we find no significant association between the maternal birth weight genetic scores and offspring disease after adjustment for child's own genetic score of the same SNPs. Rather, in particular for coronary heart disease (CHD) and hypertension, the child's own genetic score associated with disease risk independently of the maternal genetic score (logOdds -0.135, SE 0.039, p = 0.000470 for CHD, and -0.042, SE 0.020, p= 0.0322 for hypertension). These results generally agree with and extend the previous suggestions that genetic pleiotropy in the child's genome may largely account for the known epidemiological associations between birth weight and disease risk in adulthood. In summary, our work provides further support for the concept that modest intrauterine growth restriction is likely not a major risk factor for later-life disease. Based on a large sample of genotyped parent-child pairs identified from a population-based biobank, our study illustrates the opportunities that such data provides for addressing causal questions related to maternal influences.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1434 Dissecting Genetic Associations of Hypercholesterolemia in the Personalized Environmental Genes Study

Authors:

J. House, E. Brown, A. Burkholder, Z. Xu, D. Lloyd, F. S. Akhtari, C. S. Schmitt, D. C. Fargo, J. E. Hall, A. A. Motsinger-Reif; Natl. Inst. of Environmental Hlth.Sci., RTP, NC

Hypercholesterolemia (HC) results from a combination of genetic and lifestyle factors. HC is a prevalent comorbidity for multiple medical conditions including cardiovascular disease, diabetes, hypertension, non-alcoholic fatty liver disease, and metabolic syndrome, although causality is unclear.
The North Carolina-based Personalized Environment and Genes Study (PEGS) cohort consists of more than 9,000 individuals with diverse race, education, and socioeconomic status characteristics for which extensive health and exposure data have been collected, with whole genome sequencing (WGS) data on ~4,700 participants. Individuals who answered yes to, “Have you ever been diagnosed with high cholesterol?” were defined as a case (n = 1,616) and controls numbered 3,084. We conducted a Genome Wide Association Study (GWAS) on common variants (MAF > 1%) using logistic regression adjusting for age, age-squared, sex, and the first 10 principal components. We conducted a structural variation analysis using the same covariates. Lastly, we conducted rare variant analysis (RVA) on variants with MAF < 0.01 using kernal-set based approaches (SKAT).
In GWA analysis, we found variants in *APOE* associated with HC status (lead variant *hg38_19:44909976_TGGT*, OR (95%CI) = 0.51(0.43-0.62), p = 1.07e-12, CAF = 0.09). Genetic variation in other previously identified genes with HC were suggestive genome wide significant (*PCSK9* : p=6.41e-8, *LDLR* : p=2.80e-6). Genome-wide RVA revealed rare variants in and near the coding region of *DBX2* associated with HC (SKAT.p = 1.07e-6). In addition, a targeted RVA in the lead GWAS finding (*APOE*) revealed associations with HC (p=0.004). In an analysis of structural variation, we found a deletion in *DNAH6* (*hg38_2:84691636-84991687*, CAF= 0.08, p = 3.4e-7) significantly associated with HC.
APOE has previously been associated with multiple lipid traits and rare variants in *APOE* have been associated with biomarkers of Alzheimer’s Disease. *DNAH6*, a microtubule-associated motor protein, has been implicated in cognitive traits, as well as late-onset Alzheimer’s disease. Additional work is ongoing to leverage the extensive exposure data on these individuals to identify gene-by-environment interactions that contribute to understanding of HC.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1435 Distilling causality between physical (in)activity and obesity

Authors:

Z. Wang[1], G. D. Smith[2,3], R. Loos[1,4], M. den Hoed[5]; [1]The Charles Bronfman Inst. for Personalized Med., Icahn Sch. of Med. at Mount Sinai, New York, NY, [2]MRC Integrative Epidemiology Unit, Univ. of Bristol and NIHR Bristol BioMed. Res. Ctr., Bristol, United Kingdom, [3]Population Hlth.Sci., Bristol Med. Sch., Univ. of Bristol and NIHR Bristol BioMed. Res. Ctr., Bristol, United Kingdom, [4]Novo Nordisk Fndn. Ctr. for Basic Metabolic Res., Faculty of Hlth.and Med. Sci., Univ. of Copenhagen, Copenhagen, Denmark, [5]The Beijer Lab. and Dept. of Immunology, Genetics and Pathology, Uppsala Univ. and SciLifeLab, Uppsala, Sweden

Many studies have suggested that lower physical activity and more sedentary time are associated with higher odds of obesity, while obesity itself may contribute to more sedentary behavior and less physical activity. However, observational studies can suffer from confounding and reverse causation. Better instrumental variables and a more thorough consideration of potential confounding variables that may influence the causal inference between physical activity and obesity are needed.Leveraging results from our recent genome-wide association study that identified 11 loci for leisure time moderate-to-vigorous intensity (MV) physical activity and 88 loci for leisure screen time, we now have stronger instruments for causal inference. Given the strong genetic correlations of education with physical activity and obesity, we aim to examine the causal effects between physical (in)activity traits, education (defined by years of schooling), and body mass index (BMI). We applied various advanced univariable Mendelian Randomization (MR) methods - including the Causal Analysis Using Summary Effect estimates and Latent Heritable Confounder MR - as well as multivariable MR methods to disentangle these causal relationships. Univariable MR analyses suggest bidirectional causal effects between years of schooling and physical (in)activity traits. However, multivariable MR analyses show that the estimated causal effects of physical (in)activity traits on years of schooling are abolished when BMI is considered, such that more years of schooling is causally related to more MV physical activity and less screen time, but not vice versa (e.g., screen time on schooling: $\beta_{total} = -0.49$, $P = 7.8\times10^{-24}$; $\beta_{direct} = -0.04$, $P = 0.28$; schooling on screen time: $\beta_{total} = -0.33$, $P = 8.5\times10^{-76}$; $\beta_{direct} = -0.40$, $P = 8.8\times10^{-55}$). In addition, we find that years of schooling has a causal effect on BMI, but BMI does not affect years of schooling. With that, years of schooling may confound the causal relationship between physical (in)activity traits and BMI. Indeed, while univariable MR analyses suggest bidirectional causal effects, multivariable MR analyses that take years of schooling into account reveal that more MV physical activity is causally associated with a lower BMI **(**$\beta_{total} = -0.25$, $P = 2.0\times10^{-3}$; $\beta_{direct} = -0.20$, $P = 1.8\times10^{-6}$**)**, and a higher BMI with more screen time **(**$\beta_{total} = 0.16$, $P = 1.4\times10^{-74}$; $\beta_{direct} = 0.16$, $P = 1.0\times10^{-34}$), but not vice versa. In conclusion, our results highlight the beneficial effect of education on improved health and suggest that a more physically active lifestyle leads to lower BMI, whereas sedentary behavior is a consequence of higher BMI.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1436 Do iron homeostasis biomarkers mediate the associations of liability to type 2 diabetes and glycemic traits in non-alcoholic fatty liver diseases: A two-step Mendelian Randomization study

Authors:

Y. Liang[1], Y. Zhou[2], X. Bai[3], S. Au Yeung[1]; [1]The Univ. of Hong Kong, Hong Kong, China, [2]Natl. Clinical Res. Ctr. for Dermatologic and Immunologic Diseases (NCRC-DID), Beijing, China, [3]Peking Union Med. Coll. and Chinese Academy of Med. Sci., Beijing, China

Background. Previous Mendelian randomization (MR) studies showed liability to type 2 diabetes (T2D) and glycemic traits likely impacted iron homeostasis biomarkers, as well as increased risk of nonalcoholic fatty liver (NAFL) disease. Whether iron homeostasis markers mediate these associations have not been explored using MR design.

Methods. Using a two-step MR design, we assessed the mediating role of iron homeostasis biomarkers in the associations of liability to T2D, glycemic traits in liver diseases using summary statistics of genome-wide association studies (GWAS) of T2D (DIAMANTE n=933,970), glycemic traits (n≤209,605), iron homeostasis biomarkers(n≤246,139), and liver diseases (NAFL and liver cirrhosis (n≤972,707)) and its biomarkers (alanine aminotransferase (ALT) and Magnetic Resonance Imaging (MRI)-derived proton density fat fraction (PDFF) (n≤344,136)) with exclusion of potential pleiotropic genetic variants in *ABO* and *HFE* (rs1800562 and rs1799945). Inverse-variance weighted (IVW) was the main analysis, followed by sensitivity analyses (e.g., MR-Egger, weighted median, Robust Adjusted Profile Score, and MR-PRESSO) and mediation analysis using product of coefficients methods.

Results. In the main analysis, liability to T2D and fasting insulin (FI) increased NAFL ($OR_{\text{liability to T2D}}$ 1.20 per log odds, 95%CI 1.13 to 1.27; OR_{FI} 3.45 per log pmol/l, 95%CI 1.95 to 6.08), as well as ALT and PDFF. Liability to T2D also increased liver cirrhosis (OR 1.11, 95% CI 1.05 to 1.18). Amongst all iron homeostasis biomarkers considered, higher ferritin and serum iron were associated with increased risk of NAFL and liver cirrhosis respectively. In the mediation analyses, we only found ferritin partially mediated the associations of FI in NAFL (mediation effect: 7%, 95%CI 2% to 13%).

Conclusion. Ferritin likely mediated the associations of FI in NAFL. Targeting ferritin may reduce risk of NAFL in people with elevated insulin.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1437 Drastic weight loss alters expression of genes in immune pathways and induces changes in adipose cell-type composition dependent on sex and type 2 diabetes status

Authors:

X. Guo[1], S. Lee[1], D. Kaminska[2,3], V. Männistö[4], U. Arasu[5], T. ÖRd[5], M. Kaikkonen[5], J. Pihlajamäki[2,6], P. Pajukanta[1,7]; [1]Dept. of Human Genetics, David Geffen Sch. of Med. at UCLA, Los Angeles, CA, [2]Inst. of Publ. Hlth.and Clinical Nutrition, Univ. of Eastern Finland, Kuopio, Finland, [3]Dept. of Med., Div. of Cardiology, UCLA, Los Angeles, CA, [4]Inst. of Clinical Med., Univ. of Eastern Finland and Kuopio Univ. Hosp., Kuopio, Finland, [5]A. I. Virtanen Inst. for Molecular Sci., Univ. of Eastern Finland, Kuopio, Finland, [6]Dept. of Med., Endocrinology and Clinical Nutrition, Kuopio Univ. Hosp., Kuopio, Finland, [7]Inst. for Precision Hlth., David Geffen Sch. of Med. at UCLA, Los Angeles, CA

Obesity induces low-grade inflammation in adipose tissue, hampering key adipose tissue functions, i.e. lipogenesis and lipolysis, and ultimately drives obesity comorbidities, such as type 2 diabetes (T2D). We hypothesized that when individuals with obesity undergo bariatric surgery, their immune cell composition in adipose tissue will change in response to the substantial weight loss. Here we leverage longitudinal bulk RNA-sequencing (RNA-seq) data of subcutaneous adipose biopsies from 168 individuals collected at the time of bariatric surgery and 12-month follow-up to identify transcriptional changes during the drastic weight loss. We performed differential expression (DE) analysis between the baseline and follow-up stratified by sex using edgeR while adjusting for key technical factors. We performed adipose single nucleus RNA-seq (n=8) and used Bisque to estimate and adjust for cell-type proportions in the bulk RNA-seq data. In both sexes, we observed a high proportion of T-cell marker genes (58.6% in males and 57.4% in females) among the DE genes up-regulated in the follow-up. We tested for pathway enrichment of these DE genes and observed 40 T-cell-related pathways. Among the top 10% most shared genes (n=20) across the pathways (FDR<5%) for both sexes, there are two T-cell marker genes, *PRKCQ* and *RASGRP1*, associated with Type 1 diabetes and T2D in genome-wide association studies, respectively. To examine whether the enrichment in immune pathways and T-cell marker genes can be explained by changes in adipose immune cell-type composition during the weight loss, we compared cell-type proportions estimated in the adipose bulk RNA-seq data between the baseline and follow-up. We observed a decrease in macrophage proportions in both males (p.adj=1.0x10-5) and females (p.adj<1.8x10-15) and a decrease in stromal proportions in females (p.adj=3.0x10-6) at the follow-up using the Wilcoxon rank-sum test. To investigate the effect of T2D, we further stratified the data by T2D in each sex. We observed a significant decrease in macrophage (p.adj<3.5x10-15) and stromal (p.adj=7.2x10-5) proportions from baseline to follow-up for non-diabetic females, and a decrease in macrophage proportions in diabetic females (p.adj=1.1x10-5), whereas there was a decrease in the macrophage proportions only in the non-diabetic males (p.adj=1.1x10-3). These longitudinal human transcriptomics findings suggest that after weight loss, males and females with obesity decrease their macrophage proportions, and when stratifying by T2D, obese males with T2D do not change their macrophage proportions, while the obese non-diabetic males and females with and without T2D do.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1438 Drug repurposing for Alzheimer's disease: targeting the genetics-enriched, neuropathology-associated SPI1 regulon in microglia

Authors:

Y. Tan[1,2], R. Chen[1,2], Q. Wang[1,2], A. Tiwari[1,2], Y. Yan[3], B. Li[1,2], X. Zhong[1,4]; [1]Vanderbilt Genetics Inst., Nashville, TN, [2]Molecular Physiology and Biophysics, Vanderbilt Univ., Nashville, TN, [3]Dept. of Biostatistics, Vanderbilt Univ., Nashville, TN, [4]Dept. of Med., Div. of Genetic Med., Vanderbilt Univ. Med. Ctr., Nashville, TN

Alzheimer's disease (AD) is characterized by progressive decline in memory and cognition and involve multiple brain cell types. Among them, microglia play a crucial role and are the major cell type implicated in AD genetics. Evidence accumulates suggesting that drugs with support from genetics are more likely to succeed to the market. Here, we aim to identify genetics-enriched, AD-relevant regulatory programs in microglia to guide the search of approved drugs for repurposing toward AD. To do this, we first applied SCENIC to single-cell RNAseq data from the prefrontal cortex of 48 individuals with varying degrees of AD pathology to infer regulons (a regulon is module of co-expressed genes composed of a transcription factor (TF) and genes the TF regulates). We detected 241 regulons, with 42 of them being significantly enriched in AD heritability based on LDSC ($p<0.05$) on the latest AD GWAS summary statistics. Most of these genetics-enriched regulons are active in microglia, as quantified by the regulon activity score based on gene expressions in the regulons. The SPI1 regulon, named by the master regulatory TF SPI1 (PU.1), ranked with the highest AD heritability. The SPI1 regulon activity score in microglia increased as the AD pathological burden increased. Specifically, increased SPI1 regulon activity is associated with an increase in neurotic plaque burden ($p=0.03$), neurofibrillary tangle burden ($p=0.02$), tangle density ($p=0.04$), global neocortical pathology ($p=0.02$), braaksc (a measure of neurofibrillary tangles, $p=0.03$) and with a worse cognitive status ($p=0.03$). The SPI1 regulon genes are enriched for GO terms including mononuclear cell differentiation, wound healing, endocytic vesicle and GTPase regulator activity etc. Several AD risk genes appear in the SPI1 regulon, including APOE, INPP5D and CD74 etc. Finally, we collected drug targets of all FDA-approved drugs (small molecules) and evaluated the proximity between the drug target(s) and the SPI1 regulon genes on a gene-gene interaction network. We identified several promising candidates including Salsalate and Baricitinib. We are currently working on validation of Salsalate in electronic medical records (EHR). In conclusion, we showed that combining single cell RNAseq, AD GWAS summary statistics and neuropathological traits of AD enabled the dissection of genetics-enriched, AD-relevant regulatory programs in human microglia, which have value in guiding the drug purposing for AD.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1439 Drug-target Mendelian randomization analysis supports lowering plasma ANGPTL4 and APOC3 but not ANGPTL3 levels as a strategy for reducing cardiovascular disease risk

Authors:

F. Landfors[1,2], S. K. Nilsson[2], S. Kersten[3,4]; [1]Dept. of Publ. Hlth.and Clinical Med., Section of Med., Umeå, Sweden, [2]Lipigon Pharmaceuticals AB, Umeå, Sweden, [3]Nutrition, Metabolism and Genomics group, Div. of Human Nutrition and Hlth., Wageningen Univ., Wageningen, Netherlands, [4]Div. of Nutritional Sci., Cornell Univ., Ithaca, NY

Aim: APOC3, ANGPTL3, and ANGPTL4 inactivation are actively pursued pharmacological strategies for the treatment of dyslipidemia and the prevention of atherosclerotic cardiovascular disease. Their key mechanism of action is inhibiting the lipases that facilitate lipolysis and removal of lipoprotein from circulation. However, which target would yield the highest clinical benefit for cardiometabolic disease and have a favorable side-effect profile in humans is still unclear. One way to approach this problem is to use human genetics. Therefore, we used drug-target Mendelian randomization (MR) methods to compare ANGPTL4, ANGPTL3, and APOC3, to determine their relative suitability as drug targets for reducing disease risk. *Methods*: MR is a genetic epidemiology method that aims to estimate causal effects in humans while limiting bias from confounding and reverse causation. We obtained exposure and outcome data on common variants in the proximity of the target genes associated with protein levels from large-scale genome-wide association studies (N=30,565-1,296,908) and corrected for linkage disequilibrium-related correlation using generalized least squares. We evaluated six key cardiometabolic endpoints and examined potential adverse effects across 2202 health registry endpoints, 53 routine laboratory tests, and 11 MRI measurements of internal organs. *Results*: Genetically reduced circulating ANGPTL4 protein levels demonstrated a beneficial profile characterized by a lower risk for coronary artery disease (CAD) (odds ratio [OR], 0.57 per s.d. protein [95%CI, 0.47-0.69]) and type 2 diabetes (T2D) (OR, 0.74 per s.d. protein [95%CI, 0.58-0.93]). Genetic APOC3 lowering had a favorable effect on CAD (OR, 0.86 per s.d. protein [95%CI, 0.80-0.92]), while ANGPTL3 lowering did not positively affect the cardiometabolic endpoints. Genetic lowering of ANGPTL3, ANGPTL4, and APOC3 levels did not raise any safety concerns. *Conclusion*: Human genetic evidence suggests that therapies reducing the amount of ANGPTL4 and APOC3 would reduce the risk of CAD. In contrast, ANGPTL3-lowering was not observed to reduce the risk of CAD.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1440 *DUSP1* and *DUSP16* are novel IBD genes implicated in the protection of the intestinal epithelial barrier integrity.

Authors:

J. Ntunzwenimana[1,2], R. Huard[1,2], G. Boucher[2], A. Alikashani[2], C. Beauchamp[2], J. Paquette[2], P. Goyette[2], J. Rioux[3]; [1]Université de Montreal, Montreal, QC, Canada, [2]Montreal Heart Inst., Montreal, QC, Canada, [3]Montreal Heart Inst. | Res. Ctr., Montreal, QC, Canada

Background: Crohn's disease and ulcerative colitis are two chronic inflammatory bowel diseases (IBD). Genome-wide association studies have identified and validated over 200 genomic regions associated with IBD. To pinpoint the causal gene within these regions, we screened 145 genes from validated IBD loci using an intestinal epithelial cell model (HT-29). The screen revealed that *DUSP1* and *DUSP16*, two phosphatases targeting the phosphorylation of MAP kinases (MAPKs), modulated the expression of 452 and 457 genes respectively, with 40.9% of targets in common, including genes crucial for the protective function of intestinal epithelial cells. To date, no coding variants in these two genes have been associated with IBD, whereas the index SNPs in these loci (or their proxies) co-localize with different regulatory elements. Thus, our hypothesis is that a decrease in *DUSP1* or *DUSP16* expression negatively impacts their protective functions in intestinal epithelial cells. **Methodology:** To understand the impact of *DUSP1* and *DUSP16* on epithelial functions, we generated stable knockdowns (KD) of these genes in Caco-2 cells; a model known for its ability to form a monolayer of polarized cells. Intestinal barrier structure in this model was evaluated via confocal immunofluorescence imaging (IF) of apical junction complexes in 2D monolayers and 3D spheroids. Intestinal barrier function was evaluated by measuring transepithelial electrical resistance (TEER) across monolayers. **Results:** In the 2D epithelial monolayers, the knockdown of *DUSP1* or *DUSP16* altered cell shape and size and increased permeability as measured by a decrease in TEER. In addition, lowering the expression of *DUSP1* or *DUSP16* led to a diminished capacity to form polarized structures in 3D cultures. Through Western Blot and IF analyses, we observed that *DUSP1* and *DUSP16* knockdown perturbed the phosphorylation patterns of MAPKs, leading to an upregulation of myosin light chain kinase (*MLCK*) expression and phosphorylation of myosin II regulatory light chain (MLC). Consequently, this affected the localization of tight junction proteins, particularly Occludin and Zo-1, which are partially controlled by MLC. Further confirmation through IF analysis revealed the deregulation of tight junction protein localization upon *DUSP1* or *DUSP16* knockdown. **Conclusion:** By identifying functional links between two genes located in independent genomic regions associated with IBD, namely *DUSP1* and *DUSP16*, our study highlights their significant roles in maintaining intestinal homeostasis and their contribution in the development of IBD.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1441 Effect of ABCA1-R219K polymorphism in serum lipid parameters in patients under statin therapy visiting tertiary cardiac center of Nepal.

Authors:

B. Yadav[1], S. Ghimare[2]; [1]Madhesh Inst. of Hlth.Sci., Janakpurdham, Nepal, [2]Shahid Dharma Bhakta Natl. Transplant Ctr., Bhaktpur, Nepal, Kathmandu, Nepal

Introduction:ATP-binding cassette transporter A1 (ABCA1) encoded by ABCA1 gene is one of the important protein involved in lipid metabolism. The effect of statin therapy on dyslipidemia varies among individuals and it may be due to different genetic polymorphism. The R219K polymorphism of ABCA1 gene is found to have a significant role in the response of statin. **Objective:**This study was designed to evaluate the effect of R219K polymorphism in lipid-lowering action of statin in patients with dyslipidemia.**Material and Methods**:This study was conducted in 88 patients. Blood samples were taken from patients before and at the end of 3months of statin use and were analyzed for lipid profile. Whole blood was analyzed for R219K Polymorphism using PCR-RFLP.**Results:**R219K polymorphism was associated with significant percentage reduction of serum TG/HDL ratio and TC/HDL ratio in atorvastatin users. However, there was no significant association of polymorphism with change in serum TC, HDL-C, LDL-C, TG and VLDL. Among KK genotype individuals, value of TG, VLDL, TG/HDL, and TC/HDL were significantly lower than in RR genotypes. Also TG/HDL and TC/HDL were significantly lower in RK genotype than in RR. Treatment of dyslipidemia with statin was found to be comparatively better in patients having the genotypes KK and RK.**Conclusion**:Our study demonstrated association of R219K polymorphism with the significant reduction of TG/HDL and TC/HDL and particularly the KK genotype was associated with significant improvement of lipid parameters following atorvastatin treatment. **Keywords:**ABCA1-R219K polymorphism, PCR-RFLP, Lipid Profile, Single Nucleotide Polymorphism, Dyslipidemia, Atorvastatin, statin therapy

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1443 Effects of somatic and germline variants on coronary heart disease in Chinese population

Authors:

K. Zhao[1], X. Shen[1], H. Liu[1], J. Li[1], S. Chen[1], F. Liu[1], K. Huang[1], J. Cao[1], D. Gu[1,2], X. Lu[1], Z. Lin[1]; [1]Fuwai Hosp., Natl. Ctr. for Cardiovascular Diseases, Chinese Academy of Med. Sci. and Peking Union Med. Coll., Beijing, China, [2]Sch. of Med., Southern Univ. of Sci. and Technology, Shenzhen, China

Background: The genetic basis of coronary heart disease (CHD) has expanded from the germline genome to somatic mutations in haematopoietic cells, termed clonal haematopoiesis of indeterminate potential (CHIP). However, little is known about CHIP and its relationship with CHD development in East Asians, especially those with different genetic backgrounds. **Objective**: Here, we aimed to characterize CHIP in the general Chinese population and investigate its effect on incident CHD risk considering germline predisposition. **Methods**: This analysis included 6,181 participants (mean age, 54.15 years) from three large prospective cohorts with a median follow-up of 12.17 years. Fix-effect meta-analysis was conducted to calculate pooled results of the three cohorts. Targeted deep sequencing for 90 leukemia-related genes was used to detect CHIP mutations. A predefined CHD polygenic risk score (PRS) comprising 531 germline variants was utilized to evaluate CHD germline predisposition. We then measured the concentrations of six inflammatory cytokines using an automated enzyme-linked immunosorbent assay. **Results**: A total of 1,153 (18.65%) individuals harbored 1,555 CHIP mutations at baseline, demonstrating an increased risk of incident CHD (HR = 1.40, 95% CI: 1.16-1.70, $P = 5.00 \times 10^{-4}$). A risk gradient of CHD was observed across variant allele fraction (VAF) categories (P for trend = 3.90×10^{-4}). We found CHIP carriers had higher CHD risk in the intermediate and high PRS groups, but not in the low PRS group. Specifically, individuals with both high PRS and CHIP demonstrated a 2.16-fold increase in risk (95% CI: 1.47-3.18, $P = 1.04 \times 10^{-4}$) compared to those with low PRS and without CHIP. The concentrations of IL-6, CRP, and SAA were significantly higher in CHIP carriers than those without CHIP, and a dose-response relationship was observed between VAF and the levels of IL-6 and SAA. **Conclusion**: Our study comprehensively characterized the profile of CHIP in the general Chinese population and showed the effect of acquired somatic mutations on CHD was potentially modified by germline predisposition, providing new evidence for preventive precision approaches in CHD occurrence.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1444 Elucidating sex-specific gene expression at the single cell level in visceral fat.

Authors:

K. Gelev[1], M. Alvarez[1], R. Mancina[2], F. Tavaglione[3], S. T. Lee[1], O. Jamialahmadi[2], A. Kar[1], Z. Wang[1,4], D. Kaminska[5], J. Pihlajamäki[6], S. Romeo[2,3], P. Pajukanta[1,4,7]; [1]Dept. of Human Genetics, David Geffen Sch. of Med. at UCLA, Los Angeles, CA, [2]Dept. of Molecular and Clinical Med., Inst. of Med., Univ. of Gothenburg, Gothenburg, Sweden, [3]Clinical Med. and Hepatology Unit, Fondazione Policlinico Univ.rio Campus Bio-Med., Rome, Italy, [4]Bioinformatics InterDept.al Program, UCLA, Los Angeles, CA, [5]Div. of Cardiology, Dept. of Med., UCLA, Los Angeles, CA, [6]Inst. of Publ. Hlth.and Clinical Nutrition, Univ. of Eastern Finland, Kuopio, Finland, [7]Inst. for Precision Hlth., David Geffen Sch. of Med. at UCLA, Los Angeles, CA

Abdominal obesity is a growing health problem with well-established sexual dimorphism, cardiometabolic comorbidities, and inflammatory changes in the visceral fat depot. Previous studies have shown that sex affects bulk tissue gene expression; however, it is not known whether sex affects cell-type and cell state level gene expression in human visceral fat. To address this knowledge gap, we investigated visceral fat biopsies from 11 Italians with obesity, 5 females (mean BMI 41.6; SD 2.88) and 6 males (mean BMI 43.4; SD 4.53), who underwent bariatric surgery. We performed single nucleus RNA sequencing in the visceral fat and conducted quality control to remove contaminated reads and doublets. Samples were integrated with Harmony, normalized and clustered using Seurat, and annotated with SingleR. We then performed cell-type level differential expression (DE) analysis that identified 5,166 DE genes between males and females in adipocytes, 2,567 in T cells, 2,566 in B cells, 1,840 in mesothelial cells, 1,474 in adipose stem progenitor cells (ASPCs), 755 in macrophages, 278 in natural killer cells, 180 in dendritic cells, 137 in lymphatic endothelial cells, 110 in endothelial cells, 41 in mast cells, and 20 in pericytes (padj<0.05), respectively. To find subcellular cell state differences by sex, we ran Milo and found differentially abundant cellular neighborhoods (DAs) (FDR<0.10) between males and females in ASPCs, T cells, macrophages, B cells, and adipocytes. Of these DAs, we identified 1,018 significant DE genes by sex in ASPCs, 682 in T cells, 171 in macrophages, 27 in B cells, and 10 in adipocytes (FDR<0.05). We also found GO enrichment in cardiovascular development for upregulated female DE genes in ASPC DAs (FDR<0.05). Using MAGENTA, regional variants around female upregulated macrophage DE genes were found to have significant (p=0.0075) enrichment in female-specific BMI GWAS results, suggesting that these DE genes genetically contribute to the cellular differences in obesity between sexes. To reveal transcription factors (TF) that regulate the sex DE genes, we ran HOMER which showed *de novo* motif enrichment in female downregulated T cell genes for the TF and nuclear receptor, Estrogen Related Receptor Alpha (*ESRRA*). This result suggests hormonal involvement drives some sex differences we observed in the visceral adipose tissue cell-types and subcellular cell states. Taken together, we performed cell-type-specific DE analysis between males and females in each visceral fat cell-type, followed by cell state level DA and DE by sex analyses. Our results show that obese visceral fat exhibits prominent sex-specific gene expression in multiple cell-types.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1445 Elucidating the genetic architecture of autism spectrum disorder compared to more broadly defined developmental disorder

Authors:

F. Satterstrom[1], J. Fu[2], Z. Zhang[3], D. Cutler[4], B. Devlin[5], K. Roeder[6], S. Sanders[7], P. Kruszka[3], J. Buxbaum[8], M. Talkowski[2], M. Daly[2], Autism Sequencing Consortium; [1]Broad Inst., Cambridge, MA, [2]Massachusetts Gen. Hosp., Boston, MA, [3]GeneDx, Gaithersburg, MD, [4]Emory Univ. Sch. of Med., Atlanta, GA, [5]Univ. of Pittsburgh, Pittsburgh, PA, [6]Carnegie Mellon Univ., Pittsburgh, PA, [7]Univ. of Oxford, Oxford, United Kingdom, [8]Icahn Sch. of Med. at Mount Sinai, New York, NY

The Autism Sequencing Consortium (ASC) recently published an analysis (Fu et al., 2022, *Nature Genetics* 54:1320) which drew from the exomes of over 20,000 individuals with autism spectrum disorder (ASD). Some samples were obtained through ASC collaborating clinicians, and others were from Simons Foundation efforts such as the Simons Simplex Collection (SSC) and SPARK. Parents were sequenced for about three-quarters of the individuals, allowing the identification of *de novo* variants. The analysis identified 72 genes associated with ASD at a false discovery rate <= 0.001. In addition, the autism data was combined with *de novo* variants from 31,058 developmental disorder trios (Kaplanis et al., 2020, *Nature* 586:757), and a mixture model was used to identify genes mutated more often in autism than in more broadly defined developmental disorder, and vice versa. The analysis identified 36 “autism-predominant” genes and 82 “DD-predominant genes.”

Currently, the ASC is working toward a data freeze for its next analysis. We have called *de novo* variants on an additional 1,100 individuals with ASD from ASC collaborators, 5,000+ individuals with ASD from SPARK, and 11,300 individuals with ASD from the genetic testing company GeneDx. (GeneDx contributed 18,783 trios to the Kaplanis et al. study, and in fact all 11,300 trios shared with us meet the inclusion criteria for that study, and 4,000 overlap.) Together, these new samples more than double the number of individuals with ASD for whom we can call *de novo* variants compared to the Fu et al. study, and they will lead to an expanded list of ASD-associated genes. In addition, they grant new power for investigating the genetic architecture of ASD compared to more general developmental disorder/delay.

Looking within the ASD sample sets ascertained by GeneDx, SPARK, and the ASC, we find that individuals with ASD+ID/DD have a similar rate of *de novo* synonymous variation compared to individuals with ASD only (~0.29 variants/person) but a nearly doubled rate of *de novo* protein-truncating variation in constrained genes (0.152 vs 0.089 variants/person in the three most-constrained LOEUF gene deciles), with both groups significantly enriched beyond control trios. This stratification allows the identification of genes mutated more often in one condition than the other; for example, across all sample sets, 15 of 20 protein-truncating *de novo* variants in *KDM5B* are found in individuals with ASD and no DD/ID. We are currently working to deepen this analysis and will show up-to-date results at the meeting.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1446 Elucidating the genetic architecture of type 2 diabetes in 125,000 admixed Mexican adults.

Authors:

E. Trichia[1,2], J. Berumen[3], D. Aguilar-Ramirez[1], M. Turner[1,2], F. Rivas[3], M. Hill[1,2], R. Collins[1], J. Emberson[1,2], J. Alegre-Díaz[3], P. Kuri-Morales[4,5], R. Tapia-Conyer[4], J. Torres[1,2]; [1]Clinical Trial Service Unit & Epidemiological Studies Unit, Nuffield Dept. of Population Hlth., Univ. of Oxford, Oxford, United Kingdom, [2]MRC Population Hlth.Res. Unit, Nuffield Dept. of Population Hlth., Univ. of Oxford, Oxford, United Kingdom, [3]Experimental Res. Unit from the Faculty of Med., Natl. Autonomous Univ. of Mexico, Mexico City, Mexico, [4]Faculty of Med., Natl. Autonomous Univ. of Mexico, Mexico City, Mexico, [5]Monterrey Inst. of Technology and Higher Ed., Nuevo Leon, Mexico

Type 2 diabetes (T2D) is a physiologically heterogeneous disease that is common in Mexico. In the Mexico City Prospective Study (MCPS) the excess mortality risk associated with previously-diagnosed diabetes accounted for one third of all deaths before age 75. However, Mexican participants remain underrepresented in genetic studies of T2D. To address this gap and characterise the genetic architecture of diabetes in this Mexican population with high relatedness and admixture, we conducted a genome-wide association study (GWAS) of previously-diagnosed type 2 diabetes among 125,042 adults (19,431 cases) using a generalised estimating equation based approach implemented in SUGEN to account for extended family networks. Regression models were adjusted for age, sex and admixture (first seven principal components). Conditional analyses yielded 86 independent variants, of which 21 were potentially novel: 6 were in known loci, all non-coding, and 15 were mapped in novel loci and included two missense variants (rs139343836 at the *RFX6* locus and rs35409673 at the *WDFY2* locus). The loci with the most conditionally independent signals were, in descending order, *KCNQ1* (5 signals), *CCND2* (3), *TCF7L2* (2), *CDKN2B* (2), *ANKRD55* (2), *RTL1* (2), *and MC4R* (2). Heritability was ascertained using a variety of approaches to account for relatedness, admixture and genotype imputation, and yielded estimates ranging from 22%-30%. Islet enhancer and beta-cell chromatin accessibility regions presented the highest enrichment of T2D-associated variants in an analysis of over 200 tissue and single-cell annotations. To better understand the potential mechanisms through which the 86 variants are linked to diabetes risk, their associations with metabolic traits such as glycemia, adiposity, and dyslipidemia from MCPS and from external GWAS consortia were used to perform physiological clustering of the variants. Known T2D variants were assigned to previously-specified clusters, while most of the potentially novel variants mapped within or close to clusters related to insulin secretion and/or insulin action. Our work highlights the need for genetic discovery efforts in understudied populations such as Mexico, a highly-admixed population, where diabetes is common and a major contributor to mortality.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1448 Enhancer marks and Epstein-Barr virus (EBV)-encoded transcription co-factors (TcFs) are closely related to the more dominant genetic mechanisms operating among the 330 now known risk loci for systemic lupus erythematosus (SLE) in two ancestries.

Authors:

V. Laurynenka[1,2], X. Chen[1], S. Parameswaran[1], L. Kottyan[1,3], M. Weirauch[1,3], I. Chepelev[2,4], K. Kaufman[1,2,3], J. Harley[2,4]; [1]Ctr. for Autoimmune Genomics and Etiology (CAGE), Cincinnati Children's Hosp. Med. Ctr., Cincinnati, OH, [2]Res. Service, US Dept. of VA Med. Ctr., Cincinnati, OH, [3]Dept. of Pediatrics, Univ. of Cincinnati Coll. of Med., Cincinnati, OH, [4]Cincinnati Ed. and Res. for Veterans Fndn. (CERVF), Cincinnati, OH

Background: The associations of EBV infection with SLE, of anti-EBNA1 molecular mimicry with SLE autoantigens, of EBNA2 TcF binding at SLE loci of EBV-infected B cells (EBBCs), and of B-cell tropic EBV mechanisms suggest that the interaction between host and EBV is responsible for SLE autoimmunity and pathogenesis. We have identified the dominant histone mark at SLE loci in East Asian (EAS) and European (EU) ancestries in EBBCs. Methods: We identified published SLE risk loci. We curated 18,075 ChIP-seq datasets assessing DNA binding by 1,339 transcription factors and co-factors (TFs) (9,862 datasets) and 81 histone marks (HMs) (8,213 datasets), evaluated in 3,531 different "cellular sources". We modified RELI to mark DNA binding at SLE loci and assess allelic distortion with MARIO (PMID: 2966164). Results: We found 330 SLE risk loci at $p<5x10^{-8}$, published prior to 2023. Of these, 255 are reported in EAS, 106 in EU, and 30 in other ancestries, with 40 loci confirmed in both EAS & EU. KEGG analysis of 330 loci, show associations with IFNγ, IL-12, IL-23, BCR, & EBV infection ($6.1<OR<16$, $10^{-30}<p_a<10^{-8}$) and others. 200 of 1,339 tested TFs from 1,871 of the 9,862 TF ChIP-seq datasets, are associated with SLE loci at $p_c<10^{-6}$ in both ancestries. The TF dataset association ranks in EAS & EU are highly correlated ($r=0.73$, $p<0.0001$). TF DNA binding at SLE loci in EBBCs is enriched relative to other cell types in both EAS & EU ($OR>19$, $p<10^{-300}$). Removing the SLE risk loci separately in EAS & EU that are bound by the EBV-encoded EBNA2, EBNA3C, and EBNALP TcFs reduces the significantly associated human TF datasets by >95%, which is also found with many human-encoded TFs. Of the individual HM datasets, H3K27ac is the most closely associated with both the EAS & EU SLE loci ($2.43<OR<2.6$, $10^{-56}<p_a<10^{-24}$), binding a majority (>66%) of EAS & EU SLE loci. The same pattern of significant association for H3K27ac as found in EBBCs, is found in other cell types, but the proportion of significant H3K27ac EBBC datasets is far higher than any other cell type ($466<OR<882$, $10^{-12}<p<10^{-10}$), including B cells that are not EBV-infected ($121<OR<208$, $10^{-7}<p<10^{-6}$). >60% of both EAS & EU SLE loci are bound by H3K27ac more frequently in EBBCs than in B cells that are not EBV-infected. H3K27ac binding is distorted at heterozygotes for 58 EAS loci and 44 EU loci (by MARIO with score>0.4). Conclusion: TF and H3K27ac binding to SLE risk loci is relatively cell specific and highly significant in EBBCs. These results are consistent with variants influencing histone modifications and EBV-encoded TcFs in EBBCs being important for many of the mechanisms that alter genetic risk in the pathogenesis of SLE.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1449 Enhancing Alzheimer's disease prediction and portability through integration of transcription wide association study and polygenic risk scores in diverse populations

Authors:

M. Shivakumar, S-H. JUNG, E. Suh, J. Bao, L. Shen, D. Kim; Univ. of Pennsylvania, Philadelphia, PA

The prevalence of Alzheimer's disease (AD) is considerably higher among individuals of African ancestry compared to those of European or Asian ancestry. Despite this discrepancy, the use of polygenic risk scores (PRS) for AD prediction remains primarily applicable to individuals of European descent. However, recent research has demonstrated that combining PRS with TWAS-based PRS can enhance its applicability to other ancestries. While numerous prior TWAS studies have identified significant genes associated with AD, specific TWAS PRS studies on AD prediction have been limited. Therefore, the objective of this study was to investigate whether integrating TWAS-derived PRS with PRS can a) enhance AD prediction in the European population and b) improve its portability to the African American and South Asian populations. By exploring this integration, we aim to address the gap in understanding the potential benefits of TWAS-derived PRS in AD prediction for different populations.
We generated PRS using The International Genomics of Alzheimer's Project (IGAP) summary statistics. We imputed transcript expressions in 13 brain tissues using PrediXcan in Alzheimer's Disease Sequencing Project (ADSP) and Alzheimer's Disease Neuroimaging Initiative (ADNI) whole genome sequencing data. Using ADSP European population as training data (N = 10,121), we performed association analysis with AD to obtain effect sizes for the imputed transcripts. Subsequently, we utilized these effect sizes to calculate the TWASPRS by summarizing across all expressions weighed by effect size in ADNI dataset (N = 1,040), ADSP African American (N = 4,754) and ADSP South Asian (N = 2,628) dataset. Finally, we assessed the predictive performance by measuring the area under the receiver operating characteristic curve (AUC) using PRS, TWASPRS and PRS+TWASPRS as predictor and using AD as outcome.
The AUC for the PRS alone in the ADNI dataset was 0.727, while the AUC for TWASPRS was 0.781 in ADNI. However, after integrating TWASPRS with PRS, the AUC improved to 0.815. In the African American dataset, the AUC for PRS alone was 0.647, and for TWASPRS, it was 0.627. After integrating TWASPRS with PRS, the AUC increased to 0.649. In South Asian dataset, the AUC for PRS alone was 0.721, and for TWASPRS, it was 0.725. After integrating TWASPRS with PRS, the AUC increased to 0.729.
The findings from this study shows that integrating TWASPRS with PRS significantly enhances our ability to accurately predict an individual's susceptibility to developing AD in European ancestry and slightly increases portability of PRS to other ancestries.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1450 Enrichment of *AKAP11* consequential variants in bipolar disorder patients: Whole genome sequencing analysis.

Authors:

S. Chadwick, S. P. Smieszek, R. Torres, C. Xiao, E. L. Czeisler, C. M. Polymeropoulos, G. Birznieks, M. H. Polymeropoulos; Vanda Pharmaceuticals Inc., Washington, DC

Bipolar disorder (BD) is a heritable neuropsychiatric disorder characterized by episodes of mania and depression. It has a prevalence between 1-2% of the population, often with onset in early adulthood. Numerous twin studies estimated broad heritability of BD around 67%. GWAS meta-analysis of common SNPs have identified roughly 70 independent loci that contribute to BD susceptibility. These implicated genes encode ion channels, neurotransmitter transporters, and synaptic and calcium signaling pathways. Latest meta-analyses led to delineation of enrichment of rare consequential variants in patients with BD. *AKAP11*, which encodes A-Kinase Anchoring Protein 11 (AKAP-11), emerged as the strongest candidate and a definitive risk gene. It is highly expressed in the brain and has been shown to interact with GSK3B, the hypothesized target of lithium therapy.

In this study we investigated the frequency and type of *AKAP11* variants in a large cohort of whole genome sequencing samples obtained from patients who participated in a multicenter, randomized, double-blind, placebo-controlled, parallel-group study of iloperidone in patients with acute manic or mixed episodes associated with bipolar I disorder. We specifically tested the hypothesis of enrichment of rare MAF<5, pLOFs and missense variants in *AKAP11* in the BD cases (40 variants were tested). 456 samples were analyzed, and genetic analyses were adjusted for principal components, sex and age.

We report a significant enrichment of *AKAP11* variants. Specifically, we detect alleles: 133/456 in cases versus internal ancestry matched controls: 654/2845. This is a significant effect (p-value=0.0042), OR 1.4 (CI 1.1-1.7). Altogether we report 105 carriers of at least one *AKAP11* variant. The carriers were 60% males, mean age of 42 years old, and of multiple ancestries. On average, the age of onset was 24 years in carriers, 26 years in non-carriers, and 20 years in carriers of multiple variants within *AKAP11*.

This analysis demonstrates the role for rare coding variation as a significant risk factor in BD etiology. Importantly, *AKAP11* is known to confer risk for schizophrenia. *AKAP11* codes for AKAP-11, one of a family of scaffolding proteins that bind to the regulatory subunit of the protein kinase A. AKAP-11 binds to GSK3B and mediates PKA-dependent inhibition of GSK3B. PKA inhibits the activity of GSK3B bound to AKAP-11 more strongly than GSK3B in general, and modifications to AKAP-11 have the potential to affect downstream pathways. BD is a significant source of disease burden worldwide. Further understanding of genetic underpinnings will lead to better understanding of the etiology and optimal treatments.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1451 Enrichment of melanopsin genetic variants in Delayed Sleep-Wake Phase Disorder (DSWPD) patients: A whole genome sequencing analysis.

Authors:

J. Brzezynski, S. Smieszek, C. Johnson, A. Kaden, A. Goldberg, C. Miller, C. Polymeropoulos, G. Birznieks, M. Polymeropoulos; Vanda Pharmaceuticals, Washington, DC

Melanopsin (*OPN4*) is a blue light-sensitive opsin-type G-protein coupled receptor. It is highly expressed in photosensitive retinal ganglion cells which mediate responses to light, including regulation of sleep, circadian photoentrainment, and pupillary light response. Mutations in *OPN4* were shown to affect responses to light, ultimately affecting the regulation of circadian rhythm and sleep regulation. Previously, we described a male carrier of an *OPN4* missense variant diagnosed with Delayed Sleep-Wake Phase Disorder (DSWPD), with a consistent recurrent pattern of delayed sleep onset. The rs143641898 [NM_033282.4:c.502C>T p.(Arg168Cys)] variant in the *OPN4* gene was shown in a functional study to render the *OPN4* protein non-functional.

We have conducted a rigorous observational research study in suspected DSWPD patients with the aim of detecting consequential variants that may be associated with the delayed sleep phenotype. Altogether, 117 samples were collected from DSWPD patients and compared to 315 healthy controls.

We report an enrichment of *OPN4* rare coding variants (MAF<1) in DSWPD samples compared to healthy controls. This significant enrichment likely implies that other rare predicted loss-of-function variants can similarly contribute to the delayed sleep phenotype. A significant association with rs1079610, a common variant in *OPN4* with delayed bedtime in DSWPD patients, was found. This variant has been previously reported in association with altered pupillary responses. This effect is not seen in in a large set of healthy controls without DSWPD diagnoses. The discussed rare *OPN4* variants likely increase the risk of DSWPD via its direct effect on pathophysiology along the melanopsin axis.

This study offers useful insights for the differential diagnosis and ultimately treatment of DSWPD in patients that carry pathogenic variants in the *OPN4* gene.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1452 Epilepsy gene variants associated with seizure presentation in brain arteriovenous malformations.

Authors:

H. Kim[1], S. Chowdhury[2], L. Huang[1], J. Nelson[3], C. E. McCulloch[3], C. Fox[3], A. Numis[3], S. M. Weinsheimer[3]; [1]UCSF Ctr. for Cerebrovascular Res., San Francisco, CA, [2]Univ. of Toronto, Toronto, ON, Canada, [3]Univ. of California San Francisco, San Francisco, CA

Background: Brain arteriovenous malformations (BAVM) are a rare but important cause of hemorrhagic stroke in younger adults and children. The second most common presentation is seizures, which occur in 20-45% of patients and significantly impacts quality of life. While male sex and younger age have been associated with seizure onset, it is unknown whether epilepsy-related genes contribute to risk of seizure in BAVM. We hypothesized that common genetic variants in known epilepsy or epilepsy-related genes are associated with risk of seizure at initial presentation in BAVM patients. **Methods:** A total of 338 Caucasian BAVM cases were previously genotyped on the Affymetrix Genome-Wide Human SNP Array 6.0. Quality control assessment removed SNPs and samples with low genotyping call rates (<95%) and SNPs out of Hardy-Weinberg equilibrium (P<0.00001). We tested 28,312 single nucleotide polymorphisms (SNPs) mapping within 5kb of 935 genes identified as epilepsy-associated genes or genes related to epilepsy with minor allele frequency ≥1%. Using multivariable logistic regression, we tested whether the number of minor alleles was associated with seizure at presentation, adjusting for age at enrollment, sex, and the first 3 principal components to adjust for population stratification. We report odds ratio (OR) and 95% confidence interval (CI), and significance was based on Bonferroni adjustment for multiple comparisons (P=0.05/28,312 SNPs =1.8E-06). SNPs with P≤0.001 were considered *nominally* associated with seizure risk. **Results:** There were 90 BAVM patients with seizures at presentation and 248 without seizures (mean age 38.7 +/- 17.6 yrs). A total of 18 SNPs mapping to 11 epilepsy genes were nominally associated with seizure presentation in BAVM patients, although none were statistically significant after correction for multiple testing. Five of the 18 SNPs mapped to introns of the gene, teneurin transmembrane protein 2 (*TENM2)*, including rs13176788 which was the most significantly associated SNP for seizure risk (OR=2.2; CI: 1.5-3.2; P=6.3E-05). Additional SNPs associated with seizure mapped to introns of the following genes: *PRICKLE2, DLG2, SOX5, GRIN2B* and *RBFOX1*; upstream of *NALCN, SCN1A, SLC25A12* and *HERC1*; and downstream of *RBPJ*. **Conclusion:** SNPs mapping within or nearby 11 epilepsy genes including *TENM2* were associated with seizure risk at initial presentation in sporadic BAVM patients, suggesting shared genetic susceptibility. Further work is needed to confirm these findings in a larger, more diverse cohort. The identification of BAVM patients carrying epilepsy gene variants may warrant greater attention for epilepsy management.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1453 Establishing UK Biobank data augmented by electronic Medical Records for genetic association analyses on kidney function

Authors:

M. Gorski[1], T. W. Winkler[1], K. J. Stark[1], S. Wiegrebe[1,2], I. M. Heid[1]; [1]Dept. of Genetic Epidemiology, Univ. Med. Ctr., Regensburg, Germany, [2]2Statistical Consulting Unit StaBLab, Dept. of Statistics, LMU Munich, Geschwister-Scholl-Platz 1, Munich, Germany

Impaired kidney function is a substantial risk for cardiovascular disease and early mortality. Kidney function is typically assessed as estimated glomerular filtration rate (eGFR) based on serum creatinine (Scr). Hundreds of loci were identified for eGFR, including data from UK Biobank based on centralized Scr measurements (enzymatic) from baseline blood drawn in study centers (SC). Emerging electronic Medical Record (eMR) data in UK Biobank present a promising resource to augment the data on eGFR longitudinally. However, the potential of eMR-based eGFR in UK Biobank for genetic association analyses is yet unclear.

We curated a UK Biobank data set on eMR-based Scr (various labs, Jaffé or enzymatic) using a published algorithm (Denaxas et al., JAMIA Open 2020). We computed eGFR (CKD-Epi 2021) and described the resulting longitudinal information. We compared eMR-and SC-based eGFR phenotypically (selecting the eMR-Scr closest in time to SC assessment per person). We also compared genetic association effect sizes of the 634 independent variants known for eGFR (Stanzick et al., Nat Comm, 2021) on each of the two phenotypes (eMR-based, SC-based; linear regression adjusted for age, sex, 20 PC).

We obtained 1,775,433 eMR-based Scr measurements for 170,772 participants (92% European ancestry) with a median number of measurements per person of 6.0 (min=1, max=289, IQR=3.0-12.0).

When selecting the eMR-eGFR closest in time to SC assessment (n=138,362 European ancestry, unrelated), median time difference between assessments was 0.86 years (IQR= 0.27-2.63; min=0.01; max=20.46). Values of eGFR were moderately correlated and on average lower when using eMR compared to SC (Spearman r=0.67; mean eGFR =84.09 vs. 93.91 mL/min/1.73m², p<0.0001). When restricting to 55,013 persons where eMR-eGFR assessment was from the same calendar year as the SC assessment, the correlation was r=0.71.

When analyzing the 634 variants (n=138,362), we identified 206 Bonferroni-corrected significant variants for both SC- and eMR- eGFR (p<0.05/634=7.89x10-5; 296 for SC, 237 for eMR). Genetic effect estimates were highly correlated (r=0.83; similar for the n=55,013, r=0.81); standard errors were perfectly corrected and by factor 1.18 larger when using eMR- eGFR as outcome versus SC-based.

We found that the 634 variants explained a comparable proportion of the eGFR variance based on eMR- versus SC-assessment (7.0% vs. 7.4%).

Overall, we established eMR-eGFR in UK Biobank for genetic association analyses. The larger standard errors of genetic effects for eMR-eGFR might be explained by larger measurement error in eMR-Scr compared to the centralized, standardized SC-Scr measurements.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1454 † Estimating common and rare variant convergence of complex traits and diseases across diverse biological pathways

Authors:

D. Weiner, L. O'Connor; Broad Inst., Cambridge, MA

A major open question in human genetics is whether common and rare variants have mechanistically convergent effects on complex traits and common diseases. This question has critical importance for understanding disease biology and therapeutic development (Claussnitzer et al. 2020 Nature). Recently, systematic comparison of common and rare variant genetic architecture has become feasible due to biobank scale exome-sequencing resources and new analytical approaches (Karczewski et al. 2022 Cell Genomics, Weiner et al. 2022 AJHG, Weiner and Nadig et al. 2023 Nature). Recent work demonstrated convergence at the level of disease-relevant tissues and cell types (Weiner and Nadig et al. 2023 Nature). However, it remains unclear whether common and rare variants are convergent at the level of biological pathways.

To address this question, we applied recently developed tools for estimating gene set enrichments to large-scale common and rare variant association data from the UK Biobank (median GWAS/exome sample size: 382,765 individuals; rare variant analysis restricted to loss-of-function variants AF < 1e-5) and to diverse pathway annotations (n = 1,742 annotated biological pathways with at least 50 genes; Kim et al. 2019 AJHG). In a pilot analysis of 200 randomly selected pathways and 24 traits, we observed correlated common and rare enrichments (Spearman rho = 0.12 (P < 1e-10) for 4,632 trait-pathway pairs), indicating broad convergence. We observed that the relative degree of common vs. rare variant enrichment at least partially reflected inclusion in the pathway of an exome-wide significant gene. For example, filtering to lipid traits we observed rare variant enrichments 2-3x larger than common variant counterparts across a wide range of enrichments (49 < pathway size < 178), which reflected inclusion in these pathways of APOB, a gene previously shown to explain a large fraction of rare variant heritability for lipid traits. Unexpectedly, we observed many trait-pathway pairs with common-variant-specific enrichments (n = 63), such as a common variant enrichment of 46x (SE = 18x) of rheumatoid arthritis heritability in a pathway of naïve CD8+ T cells, compared to a rare variant enrichment of 0x (SE = 6x). In summary, our results are the first to systematically characterize the relationship between common and rare variant enrichments across a range of disease-relevant biological pathways for complex traits and common diseases.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1455 Evaluating a coronary artery disease polygenic risk score screening strategy in a diverse patient population at UCLA

Authors:

K. Boulier, Y. Ding, V. Venkateswaran, B. Pasaniuc; UCLA, Los Angeles, CA

Introduction: Given the existence of preventative therapies for coronary artery disease (CAD), there have been many efforts to predict and mitigate CAD risk including, more recently, incorporating polygenic risk scores (PGS). Few studies examine the clinical impact of PGS screening strategies or quantify their performance in clinical terms across a diverse patient population. Here we estimate the number needed to screen (NNS) for a PGS screening strategy for CAD based using a large and diverse electronic health record (EHR) linked biobank at UCLA Health. **Methods:** We leveraged the longitudinal EHR data to evaluate the PGS screening strategy in a synthetic manner. We assumed PGS screening to have been performed on 6/1/2013 for patients established prior to 1/1/2013 or 6 months into follow time if established after 2013. We then evaluate CAD outcomes for different strata to evaluate clinical utility (NNS). To evaluate our strategy realistically, we focused only on lower risk patients with no evidence of cardiovascular disease; patients with preexisting cardiovascular disease or strong statin indications (LDL > 190, diabetes or already on statins) are excluded. Patients with PGS scores in the top 5% of their genetic ancestry groupings (European, East Asian, African American, Hispanic) are assumed to have been started on statins for the median 6.9 years of follow up. CAD diagnosis during follow up time was determined by ICD codes. NNS to prevent one CV event is calculated from cohort size and events assuming a statin number needed to treat of 50. **Results:** Of 33,789 adults, 17,074 were low risk and eligible for PGS screening (35% male with a mean age of 44). NNS to detect one CAD diagnosis was 351, 359, 814, 159 in the European, Hispanic, East Asian and African ancestry groups and NNS to prevent one cardiovascular event was 17,550, 17,950, 40,700, 7,950 respectively. **Conclusion:** Although PGS accuracy is lower in African ancestry individuals as compared to Europeans, the clinical utility of PGS-based screening is highest in African Americans. This is due to the increased baseline rates of disease in the African American genetic ancestry grouping. Our results clearly demonstrate that PGS accuracy is distinct from clinical utility of PGS-based strategies. Evaluation of screening strategies rather than only accuracy will be required before PGS screening implementation.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1456 Evaluating integration of genetic risk into clinical risk prediction models of cardiovascular disease using polygenic risk score decomposition

Authors:

N. Larson, P. A. Decker, J. Killian, A. Batzler, S. Manemann, S. Moon, S. Bielinski; Mayo Clinic, Rochester, MN

Modern atherosclerotic cardiovascular disease (ASCVD) risk prediction models, such as the Pooled Cohort Equations (PCE), are widely used for clinical risk stratification. These models include multiple heritable risk factor phenotypes, including total and HDL cholesterol, systolic blood pressure, hypertension, diabetes, and smoking status. With the growing availability of trained PRS models across diverse phenotypes, there is increased interest in integrating PRSs into existing clinical risk prediction models. However, PRS models are typically developed from GWAS without adjustment for clinical risk factors, and thus likely include many genetic effects that are mediated through these factors via vertical pleiotropy. Thus, adding a previously trained ASCVD PRS to an existing risk prediction model may attenuate the potential benefit of integrating genetic risk information into such models, and previous studies have demonstrated only modest increases in discrimination with the addition of an ASCVD PRS. One strategy for addressing this source of redundancy is aggregating PRS model scores for disease risk itself as well as corresponding heritable clinical risk factors. For example, by linear regression we can decompose the original PRS into a residual polygenic component and complementary risk-factor-associated component. Herein, we explored this approach using data on 16,832 participants of European ancestry from the Mayo Clinic Biobank (MCB) free of ASCVD at enrollment and with available genome-wide genotyping data generated in collaboration with Regeneron Genetics Center. Representative PRS models for ASCVD risk as well as the six heritable clinical factors included in the PCE listed above were downloaded from the PGS Catalog and applied to all eligible MCB participants. Among these individuals, ASCVD risk factor PRSs explained ~13.8% of the variability in the ASCVD PRS. At a median 10.9 years of follow-up, we observed a total of 1,325 (7.9%) incident ASCVD events (i.e., myocardial infarction, CABG, PCI, stroke, CVD death). The Harrell's concordance index for the ASCVD PRS was 0.592 (95% CI = [0.577,0.608]). Cox proportional hazards regression for time-to-ASCVD was applied to fit three separate models, defined as M1: PCE risk factors alone; M2 = M1 plus the ASCVD PRS; and M3 = M2 plus the residualized ASVCD PRS. We observed statistically significant results for both the ASCVD PRS in M2 ($P<0.0001$) and the residualized PRS after adjusting for the full ASCVD PRS in M3 ($P = 0.025$). We anticipate more sophisticated methods can address heritable mediators of genetic risk and fully leverage the growing available information on polygenic risk across the phenome.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1457 Evaluating somatic mosaicism in post-mortem Parkinson's disease brains.

Authors:

Y. Heng[1], E. Chew[1], M. Lian[1], M. Tandiono[1], R. Reynolds[2], J. Foo[1]; [1]Nanyang Technological Univ. Singapore, Singapore, Singapore, [2]Imperial Coll. London, London, United Kingdom

Parkinson's disease (PD) is a common neurodegenerative movement disorder resulting from dopaminergic neuronal loss in the substantia nigra. Previous genetic studies have focused on identifying germline mutations which are present in all cells of an affected individual. However, these mutations only account for <10% of PD cases and explain a small percentage of PD risk. Beyond such heritable genetic variations, somatic mutations which are present in only a subset of cells, are gradually shown to be implicated in PD. Somatic mutations in the brain may perturb essential genes and pathways, thereby contributing to individual variation in the severity of PD. Characterizing them at high sensitivity in post-mortem brain tissues would enable us to gain biological insights into the genetic mechanisms driving PD pathogenesis.

We performed high coverage whole-exome and mitochondrial DNA sequencing on 116 well-characterized post-mortem human brain tissues, retrieved from 73 PD patients and 43 age-matched controls, to uncover genes harbouring recurrent somatic mutations that were enriched in PD patients compared to controls. The mean sequencing coverage was 176× and an average of 95.8% of bases were covered by 50 or more reads in individual samples. Using the MosaicForecast bioinformatics pipeline which was designed to identify somatic mutations from NGS data in the absence of paired samples, this study identified 22 genes that were recurrently mutated in >1 PD brain samples from between 2 to 7 PD patients. Interestingly, some of these genes were implicated in alpha-synuclein aggregation (i.e., *G3BP1* and *CAPN15*), one of the widely recognized hallmarks of PD, as well as in neuronal and axonal pathways (i.e., *DOCK7*, *DVL1* and *NAV1*). However, none of the genes identified were significantly enriched in patients compared to pathologically normal brains ($p > 0.05$) and none are known PD genes with germline variants previously identified by GWAS or family studies. Similarly, even though the control region of mitochondrial DNA was found to be enriched for genetic variants in PD patients compared to controls, none of the mutations identified in this region - which includes known mitochondrial transcription and replication regulatory sites - passed Bonferroni correction for multiple testing.

The high inter-individual genetic heterogeneity of PD noted in this present research as well as in past studies highlights the need to expand the study to include as many samples as available. This will enable characterization of the full spectrum of somatic variants and a more comprehensive assessment of their distribution in diseased vs pathologically normal, age-matched brains.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1458 Evaluating the cardiovascular impact of genetically proxied PCSK9 and HMGCR inhibition in East Asian and European populations: a drug-target Mendelian randomization study.

Authors:

L. Park[1], D. B. Rosoff[1,2], A. S. Bell[1], L. A. Mavromatis[1], A. M. Hamandi[1], J. Jung[1], J. Wagner[1], P. Pacher[1], D. Ray[2], G. Davey Smith[3], F. Lohoff[1]; [1]NIH/NIAAA, Bethesda, MD, [2]Univ. of Oxford, Oxford, United Kingdom, [3]Univ. of Bristol, Bristol, United Kingdom

Background: Lipid-lowering pharmacotherapies involving proprotein convertase subtilisin/kexin 9 (PCSK9) inhibition and/or 3-hydroxy-3-methylglutaryl coenzyme A reductase (HMGCR) inhibition are effective in reducing cardiovascular risk; however, it remains unknown whether there are differences in long-term efficacy between cohorts of East Asian (EAS) and European (EUR) ancestries. **Methods:** Using single-nucleotide polymorphisms (SNPs) in the *PCSK9* and *HMGCR* loci from EAS and EUR genome-wide association studies (GWAS) summary-level data of low-density lipoprotein cholesterol (LDL-C), we performed drug-target Mendelian randomization (MR) analysis evaluating their cardiovascular impact in these populations and assessed potential differences of the drug target-cardiometabolic relationships, including across ancestries and drug classes. **Results:** PCSK9 and HMGCR proxied LDL-C lowering (per standard deviation decrease in circulating LDL-C) were associated with reduced coronary artery disease (CAD) risk in both EAS and EUR [PCSK9 EAS: odds ratio (OR)=0.524, [95% CI, 0.430, 0.638], P-value=1.1610^{-10}; PCSK9 EUR: OR=0.503, [95% CI, 0.413, 0.613], P-value=$8,8810^{-10}$; HMGCR EAS: OR=0.681, [95% CI, 0.557, 0.833], P-value=2.0010^{-4}; HMGCR EUR: OR=0.681, [95% CI, 0.529, 0.876], P-value=0.003]. Analysis of additional cardiovascular diseases showed genetically proxied PCSK9 inhibition was associated with reduced arrhythmia, atrial fibrillation, and congestive heart failure in EUR but not EAS while peripheral artery disease was reduced in both EUR and EAS. Interestingly, genetically proxied HMGCR inhibition had no effect on any of the additional outcomes in either population. MR estimates were generally consistent across complementary MR methods and tests for heterogeneity and pleiotropy were null, strengthening causal inference. **Conclusions:** We demonstrate similar efficacy of genetically proxied PCSK9 and HMGCR inhibition for CAD risk reduction among EAS and EUR cohorts. PCSK9 inhibition had additional beneficial effects on various cardiovascular outcomes, highlighting potential additional therapeutic value. Together, these findings underscore the utility of genetics-based drug-target analysis to compare efficacy, which might help our understanding of the impact of long-term lipid lowering therapies for cardiovascular disease in diverse populations.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1459 Evaluating the effects of sex chromosome dosage on ASD risk and cognitive performance

Authors:

A. Berry[1], B. Finucane[2], S. Myers[3], M. Oetjens[3]; [1]Geisinger, Danville, PA, [2]Geisinger Hlth.System, Lewisburg, PA, [3]Geisinger, Lewisburg, PA

Autism spectrum disorder (ASD) is 3.8 times as prevalent in boys as in girls, however the biological mechanism for this sex bias is unknown. To explain the large ASD sex difference, theories such as a protective effect of two X chromosomes in females compared to the single X found in males have been proposed. However, genetic studies of ASD have largely overlooked the Y chromosome. Sex chromosome aneuploidies (SCAs) are common human genetic conditions characterized by an atypical number of X or Y chromosomes and are therefore ideal candidates for the study of X and Y chromosome dosage effects on ASD. Here, we use several population-based cohorts to assess how changes in sex chromosome dosage affect ASD risk. We assembled an ASD case-control cohort by combining individuals with an ASD diagnosis from the Simons Foundation SPARK study with controls from Geisinger's MyCode, NIH's All of Us, and the UK Biobank. Sex chromosome complements were identified using genotype array data. Logistic regression was used to determine whether ASD risk differed between each SCA and 46,XX. 46,XY and 47,XXY had ~4-fold (46,XY OR [95% CI]: 3.6 [3.5-3.7]; 47,XXY: 4.7 [3.4-6.7]) increased ASD risk compared to 46,XX. 47,XYY and 45,X had the highest ASD risk relative to 46,XX (47,XYY: 9.3 [6.6-13.1]; 45,X: 11.1 [5.9-21.0]), and were also at increased ASD risk relative to 46,XY (47,XYY: 2.6 [1.9-3.7]; 45,X: 3.1 [1.6-5.9]). Next, we examined the effect of SCAs on cognitive performance measured by a fluid intelligence test taken by UK Biobank participants. Within the UK Biobank cohort, we found that only individuals with an additional sex chromosome had significantly reduced cognitive performance (47,XXX SD [95% CI]: -0.7 [-1.0- -0.5]; 47,XXY: -0.9 [-1.1- -0.7]; 47,XYY: -0.7 [-0.9- -0.5]) relative to 46,XX and 46,XY. Comparisons between SCA groups revealed that 45,X had a significantly increased risk of ASD relative to 47,XXX (OR [95% CI] 5.4 [2.1-14.0]) and yet had significantly better cognitive performance compared to 47,XXX (SD [95% CI] 0.7 [0.3-1.1]). All adjusted $P<0.05$. In summary, our results suggest that the addition of a Y chromosome, but not an X chromosome, increases ASD risk relative to a 46,XX or 46,XY background, while the addition of either an X or Y chromosome reduces cognitive performance. Despite our observation that 45,X conferred the highest ASD risk among females, we did not observe an association between 45,X and cognitive performance typical of an ASD-risk variant of large effect. The study of SCAs may provide an untapped avenue to investigate the effects of X and Y-related gene dosage on mental health conditions that differentially affect males and females in the general population.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1460 Evaluating the interaction of polygenic risk and child sexual abuse on longitudinal trajectories of posttraumatic stress disorder symptoms

Authors:

K. Passero[1], K. O'Donnell[2], C. Selin[1], J. G. Noll[1], M. A. Hall[1]; [1]The Pennsylvania State Univ., State College, PA, [2]Yale Sch. of Med., New Haven, CT

Posttraumatic stress disorder (PTSD) has a modest estimated heritability and a suspected polygenic architecture, yet is predicated on the occurrence of a traumatic event. Although most individuals will experience a potentially traumatic event, developing PTSD is relatively rare in the general population albeit elevated in high-risk groups, such as children exposed to adverse childhood events. In adolescent victims of sexual assault, the rate of PTSD is estimated above 30%. Furthermore, child sexual abuse (CSA) is hypothesized to exacerbate risk for negative outcomes when in conjunction with other risk factors. We tested whether polygenic risk of PTSD and CSA interacted to affect the developmental trajectory of PTSD in a cohort of female adolescents. Participants were stratified into African or European ancestry groups, and ancestry-specific polygenic risk scores (PRSs) of PTSD were computed. Their PTSD symptoms over three timepoints were discretized into four trajectory classes - chronic, recovering, emerging, or resilient - using growth mixture modeling. Using multinomial logistic regression, we compared how CSA, PRS, and their interaction affected trajectory classification within each ancestral group. We computed gene-environment interactions on both the additive and multiplicative scales. Within the African ancestry participants, only CSA significantly contributed to a greater odds of a chronic trajectory, even after controlling for potentially confounding psychosocial risk factors, including alternative potentially traumatic events. Among the European ancestry participants, both PRS and the PRS×CSA interaction significantly differentiated the chronic and recovering trajectories. The interaction suggested that having higher PRS alongside CSA elevated the risk of a chronic trajectory versus a recovering trajectory. Together, these results suggest that background genetic risk may play a variable role in the development of PTSD, with a greater negative effect displayed in the presence of a salient traumatic event such as child sexual abuse.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1461 Evaluation of the association between suicidality and polygenic risk scores for anorexia nervosa, suicide attempts, and suicidal ideation.

Authors:

R. Zhang[1,2], J. Termorshuizen[2], A. Birgegård[2], L. M. Thornton[1], P. Sullivan[1], C. M. Bulik[1], A. Abbaspour[2]; [1]UNC Chapel Hill, Chapel Hill, NC, [2]Karolinska Inst.t, Stockholm, Sweden

Suicidality, including suicidal thoughts and attempts, is elevated among individuals with anorexia nervosa (AN), as evidenced by an increased risk of suicide attempts and suicide being one of the leading causes of death in this population. Twin studies suggest that shared genetic factors may underlie the co-occurrence of AN and suicide. To further explore the genetic factors that contribute to suicidality in AN, we computed three polygenic risk scores (PRS): for AN, suicide thoughts, and suicide attempts, and evaluated their associations with both ICD-based and self-reported suicidal behaviors in individuals with AN. Using data from the Swedish site of the Anorexia Nervosa Genetics Initiative (ANGI-SE) study comprising 3,189 AN cases born after 1977 with linkage to the Swedish National Registers, we examined the association of AN PRS, suicide thought PRS, and suicide attempt PRS on self-harm, suicide thoughts/plans, suicide attempts, and suicide death.

Among ANGI-SE cases, higher suicide attempt PRS was significantly associated with higher risk of ICD-based suicide attempt measured by odds ratio (OR) with 95% confidence interval (CI) (OR [95%CI]: 1.30 [1.14, 1.49]) and higher risk of self-reported self-harm (1.22 [1.08,1.38]). Higher suicide attempt PRS was also significantly associated with more frequent self-reported suicide thoughts (1.18 [1.02,1.37]), self-reported self-harm (1.19 [1.06,1.34]), and self-reported suicide attempts (1.34 [1.09,1.64]). Suicide thought PRS was significantly, albeit less strongly, associated with higher risk of ICD-based suicide attempts (1.14 [1.02,1.27]) and self-reported self-harm, both as presence/absence (1.13 [1.02,1.25]), and frequency (1.12 [1.02,1.24]). Although a moderate correlation existed between AN PRS and suicide attempts PRS (r_g=0.26), we did not observe any association between AN PRS and the suicidality measures.

These results suggest individuals with AN who have a genetic predisposition to suicide attempts are more likely to experience a range of suicide-related thoughts and behaviors. Furthermore, genetic risk factors from common variants for AN itself (i.e., AN PRS) may not directly contribute to the high risk of suicidality in individuals with AN. Risk factors including other genetic factors (e.g., rare variants) as well as environmental factors contributing to suicide risk in individuals with AN merit further investigation.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1462 Evidence for reduced somatic T-cell receptor sequence diversity profiles in Alzheimer disease among Midwestern Amish

Authors:

L. Cruz[1,2], S. Liu[1,2], K. Miskimen[1,2,3], Y. E. Song[1,2], R. A. Laux[1,2], P. Ogrocki[4], A. J. Lerner[4], A. Lynn[1], S. L. Fuzzell[1], S. D. Hochstetler[1], I. Konidari[5,6], J. L. Mccauley[5,6], W. K. Scott[5,7], M. Pericak-Vance[5,6], J. L. Haines[1,2,3], D. C. Crawford[1,2,3]; [1]Dept. of Population and Quantitative Hlth.Sci., Case Western Reserve Univ., Cleveland, OH, [2]Cleveland Inst. for Computational Biology, Case Western Reserve Univ., Cleveland, OH, [3]Dept. of Genetics and Genome Sci., Case Western Reserve Univ., Cleveland, OH, [4]Univ. Hosp., Cleveland Med. Ctr., Cleveland, OH, [5]John P. Hussman Inst. for Human Genomics, Univ. of Miami Miller Sch. of Med., Miami, FL, [6]Dr. John T. Macdonald Fndn. of Human Genetics, Univ. of Miami Miller Sch. of Med., Miami, FL, [7]Dr. John T. Macdonald Fndn. of Human Genetics, Univ. of Miami Miller Sch. of Med., Miami, FL, Cleveland, OH

Late-onset Alzheimer disease (LOAD), a progressive neurodegenerative disease characterized by brain-related changes such as amyloid-β plaque deposition, is the most common form of dementia among adults ≥65 years of age. Known risk factors include female sex, increased age, and LOAD-associated genetic risk alleles, most notably *APOE* e4. Recent lines of evidence implicate the adaptive immune system as an important component of amyloid-β plaque clearance and suggest in limited sample sizes there are LOAD-associated immune profiles detectable in blood. Here, we characterized somatic T-cell receptor (TCR) sequence diversity profiles from DNA extracted from blood (99%) and saliva (1%) in clinically diagnosed cognitively impaired (CI; n=38), cognitively borderline (CB; n=17), and cognitively unimpaired (CU; n=32) Amish participants residing in Ohio. Participants were predominantly female (63.8%), and CI participants were on average older at last clinical exam than CU participants (85.40 vs 78.70 years; t-test, $p<0.0001$). Sequencing was performed using Adaptive Biotechnologies immunoSEQ targeting the beta chain, resulting in >160,000 productive templates and >100,000 productive rearrangements. TCR beta chain sequences were characterized using Simpson's productive clonality, which ranges from 0 to 1 representing diverse and completely even sequences (e.g., 0) to monoclonal or single dominant clone (e.g., 1). We also examined the frequency of recurrent sequences for evidence of recent clonal expansion. As hypothesized, in this relatively homogeneous population, TCR sequence diversity was lowest among CI (0.123) and highest among CU (0.083), a difference (t-test, $p=0.038$) that was no longer significant after adjusting for age ($p=0.156$). TCR diversity among CB participants was intermediate (0.1043). Overall, CI participants had more dominant clones (defined here as productive rearrangements greater ≥10% in frequency) compared with CU participants (35% versus 22%, respectively). While this pattern is suggestive of increased clonal expansion among CI compared with CU participants, the difference was not significant ($p=0.255$). Relatively few clonotypes (exact nucleotide sequences) were shared across participants; of those few shared include the Epstein Barr virus associated clonotype (CASSLGQAYEQYF), with at least one count detected in 45% and 44% of the CI and CU participants, respectively. Collectively, these data suggest that reduced TCR diversity is associated with CI, but further study in larger and additional study populations is needed to establish the association and its relationship to the development of LOAD independent of age.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1463 Evidence that GWAS-derived genetic risk scores are not associated with vitiligo in African Americans, Hispanics, or Asians drawn from a clinical population

Authors:

D. Drouet, S. Liu, D. C. Crawford; Case Western Reserve Univ., Cleveland, OH

Vitiligo is an autoimmune dermatologic condition characterized destruction of melanocytes subsequently resulting in the depigmentation of the affected skin areas. The causes of vitiligo, affecting approximately 1.9 to 2.8 million adults residing in the United States, are unknown. Risk factors include a diagnosis of autoimmune diseases such as Addison's disease, pernicious anemia, psoriasis, rheumatoid arthritis, systemic lupus, erythematosus, thyroid disease, and type 1 diabetes and well as having a family history of vitiligo. Recent genome-wide association studies (GWAS) have identified >70 common single nucleotide variants (SNVs) associated with vitiligo in mostly European-descent populations. No GWAS has been published for African Americans (AA) or Hispanics (H), and it is unclear if vitiligo GWAS-identified SNVs or risk scores derived from them are also associated with vitiligo in diverse populations. In the present study, we accessed the Epidemiologic Architecture for Genes Linked to Environment (EAGLE) BioVU, a resource of de-identified electronic health records (EHRs) linked to Illumina Metabochip genotypes assayed on ~15,000 DNA samples extracted from clinical blood draws at a major medical center in Davidson County, Tennessee. Using International Classification of Diseases (ICD)-9 codes 709.01, 374.53, 624.8, we identified 118, 9, and 6 cases of vitiligo for (AA), Asian (A), (H) patients, respectively. Controls or non-cases (11,072 AA, 1,011 A, and 1,474 H) were defined as absence of case-defining ICD-9 codes. We queried the NHGRI-EBI GWAS Catalog for genome-wide association SNVs (at $p<5x10^{-8}$), risk alleles, and risk allele effect. Of the 79 identified SNVs, 10 were assayed directly on the Metabochip, and after tests of Hardy Weinberg Equilibrium and additional quality control, the final risk score included 7, 6, and 7 SNVs for AA, A, and H, respectively. We calculated both unweighted and weighted risk scores for each patient and tested each for an association with vitiligo using logistic regression. Mean weighted vitiligo risk score was 2.22 ± 0.45 (AA), 1.64 ± 0.44 (A), and 2.32 ± 0.52 (H) in controls and 2.15 ± 0.49 (AA), 1.55±0.21 (A), and 2.27 ± 0.62 (H) in cases. Despite previous reports of strong genetic associations for vitiligo in European-descent populations, weighted risk scores adjusted by sex risk scores not associated with vitiligo (p-values>0.05) in any of the diverse patient groups tested here (ORs = 0.70; 95% CI 0.46-1.06 (AA), 0.61; 95% CI 0.12-2.78 (A), and 0.83; 95% CI 0.18- 3.93(H)). The lack of association underscores the need for both additional resources and discovery efforts for vitiligo in diverse populations.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1464 Evolutionary meta-analysis of exomes identifies new Type 2 Diabetes risk genes and mechanistic pathways

Authors:

K. Wilhelm, J. K. Asmussen, K. Lee, M. Samieinasab, O. Lichtarge; Baylor Coll. of Med., Houston, TX

Type 2 Diabetes (T2D) is a leading cause of morbidity and mortality. It is characterized by impaired insulin sensitivity and production, leading to hyperglycemia and progressive metabolic, vascular, renal, and neurological complications. While environmental factors, exercise, and body mass index (BMI) are important, genetics is a major risk factor in T2D. Genome-wide association studies (GWAS) have suggested over 5,000 risk loci, yet missing heritability remains. To fill the missing heritability gap, we analyzed whole exome sequencing data from 3,273 T2D cases and 3,273 age-, sex-, and BMI-matched healthy controls from the UK Biobank. We annotated missense variants by their inferred functional impact calculated from phylogenetic distances and amino acid substitution odds. Next, genes were tested for differences in the functional burden of mutations between cases and controls. For this, we used three complementary machine learning algorithms. Finally, we combined the results by meta-analysis into a high confidence set of 178 genes hypothesized to be associated with T2D risk. These 178 genes overlapped with known risk genes (n = 6/72, p = 4.5x10^{-5}), co-localized with T2D GWAS loci (p = 0.02), and interacted with known risk genes in the STRING protein-protein interaction network (z = 5.4). In the same network, these genes were also mutually linked, grouping into clusters that discovered T2D-related pathways, such as “Glucose metabolism” (FDR = 0.003) and “Regulation of insulin secretion” (FDR = 0.009). Interestingly, we also recovered potentially novel pathways, such as “Negative regulation of stem cell proliferation” (FDR = 0.008). Additionally, when we added the 72 known risk genes into the network, we observed that many of the novel candidate genes interacted with them to augment functional clusters with annotations such as “Insulin receptor signaling pathway” (FDR = 1x10^{-10}), “Glycolysis” (FDR = 2.2x10^{-8}) and “Negative regulation of fat cell differentiation” (FDR = 0.004). Some of these novel genes also carried mutations associated (FDR < 0.1) with T2D cases, including *TNS1* and *PTCH1*. In contrast, we found other genes associated with healthy controls that may be protective, such as *PLSCR1* and *DLG1*. Finally, the 178 candidates predicted early-onset T2D in a holdout cohort (n = 988, AUROC = 0.58), on par with other genetics-only polygenic risk scores and suggesting that our results may be complementary to prior models to uncover the full genetic architecture of T2D. In summary, this study combined machine learning with evolutionary analysis of coding variants in a relatively small cohort to identify new genes contributing to risk in T2D.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1465 Examination of the genetic architecture of nephrotic syndrome in large-scale biobanks replicates known and identifies 20 novel loci.

Authors:

H. M. Seagle[1], J. H. Breeyear[1,2], E. S. Torstenson[3], Y. Zhang[4], G. P. Jarvik[5], O. Dikiltas[6], M. T. M. Lee[7], I. J. Kullo[6], R. Rowley[8], K. Kiryluk[9], T. L. Edwards[1,2,10], K. Susztak[11], J. N. Hellwege[1,2,12]; [1]Vanderbilt Univ., Nashville, TN, [2]VA Tennessee Valley Hlth.care System (626), Nashville, TN, [3]Vanderbilt Univ. Med. Ctr., Nashville, TN, [4]Geisinger Hlth.Systems, Danville, PA, [5]Univ. of Washington Med. Ctr., Seattle, WA, [6]Mayo Clinic, Rochester, MN, [7]Galatea Bio, Inc., Hialeah, FL, [8]NIH, Bethesda, MD, [9]Columbia Univ., New York, NY, [10]Vanderbilt Univ. Med. Ctr., Div. of Epidemiology, Nashville, TN, [11]Univ. of Pennsylvania and Children's Hosp. of Philadelphia Kidney Innovation Ctr., Philadelphia, PA, [12]Vanderbilt Univ. Med. Ctr., Div. of Genetic Med., Nashville, TN

Nephrotic syndrome (NS) is a collection of disorders characterized by excess protein in the urine (>3g/day). Nephrotic syndrome has heterogeneous histopathological manifestations, including minimal change disease, focal segmental glomerulosclerosis and membranous nephropathy. Our goal was to identify loci associated with adult NS across European (EUR), African (AFR), and East and Central Asian (ASN) ancestry groups. We integrated genotype data from Vanderbilt University's BioVU and the Electronic Medical Records and Genetics (eMERGE) network with summary statistics from FinnGen, Million Veteran Program (MVP), Biobank Japan (BBJ), and UK Biobank (UKB), for a cross-ancestry total of 5,214 cases (3,060|790|1,364; EUR|AFR |ASN and 1,601,060 controls (1,279,630|135,664|185,766; EUR|AFR|ASN). Cases were identified with diagnostic codes (FinnGen) or phecodes (BioVU, eMERGE, MVP, BBJ, UKB) for NS, while controls were without renal disease diagnosis codes. Associations with NS were modeled as a function of additive genotype, sex, and the top 10 principal components, followed by inverse-variance weighted meta-analysis within and across ancestral groups. We identified 31, 15, 43, and 13 significant (r^2<0.1, >1Mb, $p<5x10^{-8}$) loci in the multi-ancestry, EUR, AFR, and ASN analyses, respectively. The strongest association in the multi-ancestry analysis was rs1265889 (class II MHC region, $p=1.4x10^{-20}$, OR=1.50 (1.43-1.57)). We significantly ($p<5x10^{-8}$) replicated loci associated with pediatric NS (*ZNF345*) and membranous nephropathy (*PLA2R1*). In the AFR analysis, we identified 16 novel non-HLA loci, as well as replicated the *APOL1* G1 variants ($p=4.5x10^{-15}$, OR=2.56). We evaluated associations between NS and transcriptome wide predicted gene expression using GTEx v8 tissue models. We identified 28 unique genes from 167 significant ($p<1.55x10^{-6}$) gene-tissue pairs in the multi-ancestry, 33 unique genes at 337 gene-tissue pairs in EUR, and one gene out of two gene-tissue pairs in AFR. We interrogated significant SNPs in these genes with the Human Kidney eQTL Atlas (susztaklab.org) and observed 1,080 kidney microstructure and bulk tissue eQTLs (FDR<0.05) in *AGPAT1, C4A, HLA-DQA1, HLA-DQB1, HLA-DQB2, NOTCH4,* and *RNF5* genes. This study provides insight into adult NS through identification of novel genetic loci and translation to predicted renal gene expression in the largest multi-ancestry meta-analyses to date.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1466 Examining Dry Eye Disease Phenotype in UK Biobank

Authors:

B. Riley-Gillis[1], M. Liu[1], J. Lee[1], A. Acevedo-Canabal[2], K. Held[2], A. Abbasi[1], N. Smaoui[1]; [1]AbbVie, North Chicago, IL, [2]AbbVie, Irvine, CA

Dry eye disease (DED) is a common, and often under-diagnosed disease, that affects five to fifty percent of the population. DED is a heterogenous disease with symptoms such as blurred vision and irritation, and treatments vary from over-the-counter artificial tears to surgical closure of the tear ducts. Symptoms of DED can be driven by multiple factors, including environmental (e.g., arid climate), behavioral (e.g., prolonged use of computers) or medication use. Twin studies have estimated the heritability of DED from 50%-80% depending on the phenotype, which implies a genetic component of the disease.

The mechanism of how these factors, individually or together, can lead to dry eye symptoms has not been well understood. Furthermore, precisely identifying DED as a phenotype is a challenge as the different causes and symptoms are often classified under one umbrella medical code.

In the current study, we utilized the UK Biobank to investigate how genetic factors may play a role in DED. First, we defined dry eye disease using a combination of medical records and prescription information to precisely define DED cases with established dry eye symptoms. Next, we limited cases to individuals of European ancestry to avoid any population confounders. In total, we identified a case cohort of 2,327 and a control cohort of 137,672 who do not have medical records or prescriptions related to dry eye. A case-control genome-wide association study (GWAS) was performed but did not identify any genome-wide significant loci. Stratification of the cohort by sex or known dry eye disease factors also did not find any genome-wide significant loci.

In conclusion, we characterized a dry eye disease cohort within a large biobank using medical records and prescription information, however GWAS analysis did not identify any genetic factors associated with dry eye disease. The heterogeneity of DED is a key challenge that may be addressed by precise phenotyping (e.g., Schirmer values) to more accurately classify DED patients for future genetic analyses.

Disclosures:

ML, JL, AAC, KH, NS, BRG are employees of AbbVie. The design, study conduct, and financial support for this research were provided by AbbVie. AA reports contract work with AbbVie at the time of study. AbbVie participated in the interpretation of data, review and approval of the publication.

This research was carried out using the UK Biobank resource under application number 26041.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1467 † Examining the link between 179 lipid species and 8 diseases

Authors:

L. Ottensmann[1], R. Tabassum[1], S. E. Ruotsalainen[1], M. J. Gerl[2], C. Klose[2], D. L. McCartney[3], E. Widen[1], FinnGen, K. Simons[2], S. Ripatti[1,4,5], V. Vitart[6], C. Hayward[6], M. Pirinen[1,4,7]; [1]Inst. for Molecular Med. Finland (FIMM), University of Helsinki, Finland, [2]Lipotype GmbH, Dresden, Germany, [3]Ctr. for Genomic and Experimental Med., Inst. of Genetics and Cancer, Univ. of Edinburgh, Western Gen. Hosp., Edinburgh, United Kingdom, [4]Dept. of Publ. Hlth., Clinicum, Faculty of Med., Univ. of Helsinki, Helsinki, Finland, [5]Broad Inst. of the Massachusetts Inst. of Technology and Harvard Univ., Cambridge, MA, [6]Med. Res. Council Human Genetics Unit, Inst. of Genetics and Cancer, Univ. of Edinburgh, Western Gen. Hosp., Edinburgh, United Kingdom, [7]Dept. of Mathematics and Statistics, Univ. of Helsinki, Helsinki, Finland

Background: Human plasma lipidome captures risk for cardiovascular diseases and beyond. Genome-wide association studies (GWAS) of individual lipid species have identified many lipid-associated loci, however the influence of associated genetic variants on disease risk remains poorly understood.
Methods: We derived polygenic risk scores (PRS) for 131 lipid species measured in two independent Finnish cohorts, GeneRISK and FinRISK. PRS were calculated using GWAS summary statistics from the GeneRISK cohort for 7,174 individuals with PRS-CS and PLINKv1.9 and validated using the FinRISK cohort (N=1,032). Regional heritability for 179 lipid species from the GeneRISK cohort was estimated with FINEMAP. For lead variants, association analyses of 8 disease endpoints were performed in 3 independent cohorts- FinnGen (N=377,277), UK Biobank White British subset (N=408,959) and Generation Scotland (N=20,032).
Results: PRS explained >4% of variance for 25 lipid species. For each of these lipid species, we identified the locus, which when removed from the PRS calculation resulted in a drop of variance explained > 2%. There were six loci (*AL161670.1*, *CERS4*, *FADS2*, *LINC01722*, *LIPC*, *ZPR1*) for which variance explained was < 2.5% after removal of the corresponding loci. These six loci and five other loci (*ABCG8*, *APOE*, *GLTPD2*, *NTAN1*, *TMC4*) reached regional heritability values > 2%. The GWAS lead variants and top variants of credible sets identified by fine-mapping of these 11 high impact loci revealed their associations with cholelithiasis and alzheimer's disease in all 3 cohorts, ischaemic heart disease, statin medication, type 2 diabetes and pure hypercholesterolemia in 2 cohorts, and age-related macular degeneration in 1 cohort (P < 6.25e-9). Non-alcoholic fatty liver disease was nominally associated in 2 cohorts (P < 6.25e-3). We observed several distinct signals in some loci. For instance, one of five independent *LIPC* signals (missense variant rs113298164) had larger effect sizes for both age-related macular degeneration and phosphatidylethanolamines compared to another *LIPC* signal (intronic variant rs1800588).
Conclusion: PRS explained over 4% of variance for 25 lipid species. We identified 11 loci with regional heritability estimates over 2%. Lipid species associated variants within these loci revealed significant associations with 8 disease endpoints. In some loci, the effect sizes on lipids and diseases varied between independent signals. These results might help in determining the most suitable genetic instruments required to tease out causal relationships between lipid species and disease endpoints.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1468 Examining the relative predictive accuracy of polygenic scores across ancestral populations

Authors:

R. Alemu[1,2], P. Turley[3], A. Okbay[4], D. Benjamin[5,6]; [1]Univ. of California Los Angeles, Los Angeles, CA, [2]Broad Inst. of MIT and Harvard, Cambridge, MA, [3]Univ. of Southern California, Los Angeles, CA, [4]Vrije Univ.it Amsterdam, Amsterdam, Netherlands, [5]Univ. of California Los Angeles, Santa Monica, CA, [6]NBER, Boston, MA

Genome-wide association studies (GWASs) are conducted to identify replicable genomic risk loci associated with diseases and other traits. Polygenic Scores (PRS) are genetic predictors of phenotypes computed using GWAS-generated summary statistics for the corresponding phenotype. The recent increase in the sample size of GWASs has helped improve the predictive performance of PGS. However, this improvement is unbalanced across ancestral groups due to the over-representation of European ancestry individuals in most GWASs (Martin et al., 2019). The predictive accuracy of PGS falls by an average of 37, 50, and 78 percent in individuals of South-Asian, East-Asian, and African ancestries, respectively (Martin et al., 2019). Prior studies have primarily focused on a few biological phenotypes, such as anthropometrics, blood pressure and hemoglobin levels. This paper examines the loss of predictive accuracy in behavioral traits such as educational attainment, substance use, and neuropsychiatric conditions. For such traits, we may expect the relative predictive accuracy to be lower, given the complex environmental pathways through which these PGS operate. Furthermore, we estimate what share of the PGS accuracy loss is explained by differences in minor allele frequencies (MAF), linkage disequilibrium, and heritability, closely following the decomposition approach recently developed by Wang et al. (2020). The SNP weights were generated from the largest training set containing European ancestry individuals excluding a validation set, while the PGS were computed using SBayesR, a Bayesian multiple regression method. Our preliminary results show PGS relative accuracy for educational attainment (EA) in African ancestry individuals in the HRS and Add-health cohorts are 11 and 15 percent, respectively. Moreover, we find that the share of predictive accuracy shrinkage attributable to minor allele frequency (MAF) and linkage disequilibrium (LD) is 65 percent, implying that the remaining loss is likely due to environmental factors or an imperfect cross-population correlation of causal SNPs. Quantifying the relative contribution of these factors may have behavioral, clinical, and policy implications.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1469 Expanded multi-ancestry genome-wide association meta-analysis of IBD identifies 85 new loci and implicates potential therapeutic drug targets

Authors:

L. Fachal, on behalf to the International IBD Genetics Consortium; Wellcome Sanger Inst., Hinxton, United Kingdom

Genome-wide association studies (GWASs) have identified 320 loci associated with inflammatory bowel disease (IBD) susceptibility and its two main subtypes, Crohn's disease (CD) and Ulcerative Colitis (UC). About 15% of these loci include genes that lie in pathways targeted by drugs currently approved to treat IBD. Identifying new loci can reveal novel therapeutic targets for IBD.

We carried out a multi-ancestry GWAS meta-analysis of 67 cohorts, from European (EUR), East Asian (EAS) and South Asian (SAS) populations. The final dataset included 105,116 IBD patients (49,625 with CD; 47,378 with UC) and more than 1.2 million population controls.

We identified 194 novel (P<5x10-8) disease associated signals, with no evidence of effect heterogeneity across populations. Out of these, 85 are located >1Mb from any previously reported loci, thus representing a more conservative estimation of new loci implicating new causal genes. To prioritise effector genes and generate therapeutic hypotheses we colocalised IBD loci with bulk and single cell QTL data from relevant tissues, cell types and environmental conditions. 154 IBD loci showed strong evidence of colocalisation (PP>0.8), including two independent new loci associated with UC and IBD that colocalised with *SIK2* and *VSIR*, respectively.

SIK2 encodes a protein kinase that helps maintain a pro-inflammatory status by inhibiting IL10 production in macrophages. The UC protective allele (OR=0.94;P=3.1E-9) is associated with decreased SIK2 expression in macrophages, monocytes, neutrophils and T-cells, which suggests that molecules inhibiting SIK2 could have a therapeutic effect on UC.

VISTA, the protein encoded by *VSIR*, is an immune checkpoint receptor. Knockout mice show increased expression of IL23 and develop chronic inflammation in multiple tissues. The IBD protective allele (OR=0.95;P=7.5E-9) is associated with increased *VSIR* expression in colon and oesophagus, indicating that compounds with an agonist effect on this protein might have therapeutic value.

For both genes, we have added genetic support for the mechanism of action of compounds (SIK2 inhibitors; VISTA agonists) in ongoing pre-clinical and clinical trials for immune-mediated traits, including IBD.

In summary, we performed the largest multi-ancestry GWAS meta-analysis of IBD to date, identifying 85 new IBD loci and providing further insights into the biology of the disease. Integrating molecular data from multiple cell types and conditions not only expands our ability to determine candidate causal genes, but it also provides additional evidence of the utility of genetics as an instrument to guide the development of new therapeutic drugs.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1470 Exploiting pleiotropy of corneal resistance factor genetic associations.

Authors:

X. Jiang[1,2], T. Boutin[1], V. Vitart[1]; [1]MRC Human Genetics Unit, Inst. of Genetics and Cancer, Univ. of Edinburgh, Edinburgh, UK, Edinburgh, United Kingdom, [2]Ctr. for Genetics and Molecular Med., Inst. of Genetics and Cancer, Univ. of Edinburgh, Edinburgh, UK, Edinburgh, United Kingdom

Background: Genome-wide association studies (GWAS) have identified over 100 genomic loci contributing to inter-individual variation in corneal resistance factor (CRF). Those have proved particularly useful to flag susceptibility loci for keratoconus, a postnatal ocular disease characterized by progressive thinning of the central cornea. Collagen and extracellular matrix (ECM) regulation pathways were prominently implicated by geneset enrichment analyses, based on closest genes to lead variants. Better understanding of the mechanisms underlying CRF associations promises to shed light on keratoconus poorly understood pathogenesis but also plausibly on ECM regulation in diverse tissues of the human body.
Methods: In absence of GWAS of gene expression in the cornea, we performed colocalization analysis between CRF fine-mapped GWAS signals and those affecting expression (eQTL) or splicing (sQTL) of genes in 49 diverse, non-corneal, tissues from the GTEx consortium. We assume that a subset of CRF causal variants affecting gene regulation in a GTEx tissue would also be acting in a corneal cell type in the same molecular way. To support this hypothesis, genes linked to regulatory loci by colocalization were evaluated in adult corneal single cell RNA Atlas. Conditions other than those affecting the cornea that the prioritized CRF variants might be associated with will be highlighted using phenome wide association studies (PheWAS) and multitrait colocalization harnessed to help refine causal variants and target genes.
Results: Twenty-six percent of 181 CRF independent causal signals colocalize with e/s QTLs from GTEx. This subset appears enriched in keratoconus susceptibility loci, and depleted in loci close to genes implicated in Mendelian connective tissue conditions. The shared causal signals with GTEx e/s QTLs provide many new insights into mediation of CRF genetic effects, including modulation of splicing events. All the nominated causal protein-coding genes were expressed in the adult cornea, with significant enrichment found in the stromal cells. Extra and intracellular known roles of several implicated genes products in providing tensile strength, mechano-sensing and signaling make the corresponding genes and regulatory variants prime CRF candidates to be validated and their roles and effects understood within specific contexts.
Conclusion: Although limited, the set of CRF and keratoconus associated causal variant and gene candidates exposed here presents the advantages of being relevant to more than one tissue and likely involved in maintenance of adult cornea rather than (or as well as) developmental processes.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1471 Exploring causal mechanisms of sex dimorphism at lipid associated loci.

Authors:

D. Zanetti[1], C. Carland[2], F. Busonero[1], A. Maschio[1], F. Crobu[1], J. Petrie[3], L. Lind[4], T. Assimes[5,6], S. Sanna[1,7]; [1]Inst. of Genetic and BioMed. Res. (IRGB), Natl. Res. Council (CNR), Cagliari, Italy, [2]Univ. of Pennsylvania, Philadelphia, PA, [3]Univ. of Glasgow, Glasgow, United Kingdom, [4]Uppsala Univ., Uppsala, Uppsala, Sweden, [5]Stanford Univ. Sch. of Med., Palo Alto, CA, [6]VA Palo Alto Hlth.Care System, Palo Alto, CA, [7]Univ. of Groningen, Groningen, Netherlands

Background: Large scale genetic studies on lipid traits have highlighted several loci with evident sex-heterogenous effects for which the underlying causal genes remain unknown. Intriguingly, a protein quantitative trait loci (pQTLs) mapping study recently identified 118 pQTLs associated with 48 circulating proteins showing significantly heterogeneous effects between sexes. Out of these 48 proteins, 15 are mapping loci previously associated with lipid traits. We explored the roles of such proteins in explaining the observed sex dimorphism in lipids traits. **Methods:** We searched for shared mechanisms and causal links between lipid-associated loci and proteins located at these loci using colocalization and two-sample Mendelian Randomization (MR) methods. We used sex-specific summary statistics of high-density lipoprotein cholesterol (HDL-C), low-density lipoprotein cholesterol (LDL-C) and triglycerides (TG) from the Global Lipids Genetics Consortium (DOI: 10.1186/s13059-022-02837-1), and sex-specific pQTL summary statistics for 15 proteins measured using the Olink's Metabolism panel in a total of 10,885 women and 12,112 men from 12 cohorts from the SCALLOP consortium (DOI: 10.21203/rs.3.rs-2621829/v1). For MR, we used the two-sample MR package to select for each protein independently ($r2<0.001$) associated variants at $p<5x10^{-8}$, either in cis- or in trans-, and performed MR using Inverse Variance Weighted, Weighted Median and MR-Egger tests. **Results:** The colocalization analyses supported the existence of a potential shared causal variant for the proteins: COMT and LILRA5 with LDL-C; MEP1B with TG; and SEMA3F with HDL-C - all only in males. Using a Bonferroni-adjusted threshold of $p<10^{-4}$, the MR analyses in females supported evidence of a significant causal association with TG for a total of 2 proteins, TYRO3 and ANGPT2, and with HDL-C for MEP1B. In males, a significant causal association was found for the protein COMT with HDL-C, thus corroborating findings from colocalization analyses. In contrast, protein NECTIN2 showed causal links with all lipid traits in both sexes, with similar causal effect. The leave-one-out and sensitivity analyses removing outliers due to horizontal pleiotropy (MR-PRESSO) provided consistent results. **Conclusions:** Our study identified possible sex-specific causal roles of several proteins located at loci associated with lipids traits. This approach evidences the need for future genetic studies to perform sex-stratified QTL mapping and integrative analysis as a route to understanding sex dimorphism in lipid metabolism.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1472 Exploring cytokine pathways as potential determinants of multimorbidity: a Mendelian randomisation study

Authors:

N. Hukerikar[1], F. Schmidt[2,3,4,5], C. Finan[2,3,4], F. Asselbergs[1,2,4,5], A. Hingorani[2,3], S. Chopade[2,3]; [1]Inst. of Hlth.Informatics, Faculty of Population Hlth., Univ. Coll. London, London, United Kingdom, [2]Inst. of Cardiovascular Sci., Faculty of Population Hlth., Univ. Coll. London, London, United Kingdom, [3]UCL BHF Res. Accelerator Ctr., London, United Kingdom, [4]Dept. of Cardiology, Div. Heart and Lungs, Univ. Med. Ctr. Utrecht, Utrecht Univ., Utrecht, Netherlands, [5]Dept. of Cardiology, Amsterdam Cardiovascular Sci., Amsterdam Univ. Med. Ctr., Univ. of Amsterdam, Amsterdam, Netherlands

Background: Multimorbidity, the co-occurrence of two or more distinct diseases, presents a challenge in research and healthcare. Cytokines and their regulators play an important role in acute and chronic inflammatory responses. Moreover, they often influence multiple disease-causing pathways, thereby having implications for multimorbid diseases. This study aimed to characterise the role of cytokines as determinants of multimorbidity, and identify druggable cytokines and pathways relevant for therapeutic interventions in multiple diseases.
Methods: Using data from five large genome-wide association studies on the plasma proteome (between 1328 and 35559 samples), *cis*-acting genetic variants for 139 cytokines and regulatory proteins were identified. Mendelian randomisation was employed to determine the association between these immune proteins and 64 diseases and biomarkers relevant to multimorbidity. A knowledge graph was constructed from the significant protein-trait associations, which was further annotated with information on druggability, licensed indications and trait-tissue associations. This entire network was queried to identify cytokine communities and prioritise a subset of pleiotropic cytokines and pathways, representing the subset of immune-proteins strongly involved with multimorbidity.
Results: Diseases frequently implicated for cytokine involvement include inflammatory bowel disease (affected by 23 proteins), lung cancer (21 proteins), type 2 diabetes (20 proteins), Crohn's disease (20 proteins) and Asthma (19 proteins). These proteins were also frequently associated with cardiometabolic risk factors such as cholesterol and blood pressure. Based on the graph communities and graph node importance, 24 pleiotropic cytokines, associated with multiple clinically relevant traits, were prioritised. This included 23 druggable proteins, such as C4B, CX3CL1 and IL1RN.
Conclusion: This study found strong genetic support for plasma levels of immune regulatory and effector proteins acting pleiotropically and partially determining the onset of multimorbidity. Druggable proteins informing *de novo* and ongoing drug development programs against sets of diseases were also identified.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1473 Exploring Molecular Interactions between SARS-CoV-2 Infection and Diabetes: Transcriptomic Analysis of PBMCs Reveals Converging Pathways

Authors:

K. Kumar; Columbia Univ., New York, NY

Patients with diabetes and COVID-19 are at a heightened risk of developing severe complications, such as acute respiratory distress syndrome, multi-organ failure, and death. This increased susceptibility may be attributed to the underlying inflammation associated with obesity, insulin resistance, and other comorbidities typically seen in patients. Patients with hyperglycemia exhibit elevated expression of angiotensin-converting enzyme 2 (ACE2), the cellular receptor for viral entry, which further facilitates viral infection. The pre-existing chronic inflammation, augmented inflammatory response, and increased viral load in diabetic patients contribute to a systemic immune response known as the "cytokine storm," which is strongly associated with the severity of COVID-19. Moreover, SARS-CoV-2 infection can also induce dysregulation of metabolic factors and trigger the onset of diabetes. However, the molecular mechanisms underlying the bidirectional relationship between SARS-CoV-2 infection and diabetes are not fully understood. In this study, comparative transcriptomic analysis of peripheral blood mononuclear cell (PBMC) samples using publicly available scRNA-seq and bulk RNA-seq datasets from COVID-19 patients and type 1 and type 2 diabetes patients are used to elucidate the common molecular pathways that are involved in these conditions. An ensemble learning-based regression model informed by cell state, differential expression, predicted miRNA interactions, age, and sex distinguished significant features in altered metabolic signaling and immune profiles across measures of SARS-COV-2 infection severity. Further, functional assessment with gene ontology (GO) and pathway analyses revealed common mechanistic links between SARS-CoV-2 infected and type I and type II diabetes datasets paired with overlapping miRNA-mRNA-TFs regulatory networks through co-modulated genes involved in cytokine signaling. Such insights may contribute to the development of targeted therapeutic strategies and improved clinical management for patients with diabetes and COVID-19 at heightened risk of severe complications.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1474 Exploring SETD1A LoF models to elucidate genetic etiology in schizophrenia risk

Authors:

H. Giles[1,2], T. Sawada[2], A. S. Feltrin[2], Y. Wang[2], B. A. Torres[2], A. E. McCord[2], H. Shizhong[2,1], E. Radulescu[2], Q. Chen[2], B. Qamar[2], A. R. Barbosa[2], R. S. Jacomini[2], V. Dimitrova[3], R. Kaneva[3], V. Vladimirov[4], T. M. Hyde[2,5], D. Weinberger[2,5], A. C. Paquola[6,1], J. A. Erwin[2,1]; [1]Johns Hopkins Univ. Sch. of Med., Baltimore, MD, [2]Lieber Inst. for Brain Dev., Baltimore, MD, [3]Med. Univ. of Sofia, Sofia, Bulgaria, [4]Univ. of Arizona, Phoenix, AZ, [5]Johns Hopkins Sch. of Med., Baltimore, MD, [6]Lieber Inst. for Brain Dev. / Johns Hopkins Med., Baltimore, MD

Heterozygous loss-of-function (LoF) mutations in SETD1A are rare variants presenting significantly increased risk for schizophrenia (SCZ). However, the mechanism by which SETD1A LoF leads to SCZ-associated phenotypes, and the relationship between SETD1A haploinsufficiency and common variant-associated SCZ, are unknown. We investigated possible connections using isogenic human pluripotent stem cell models engineered with SETD1A LoF variants (such as the most common pathogenic mutation: c.4582-2delAG>-) as well as human prenatal brain samples. We found that SETD1A preferentially acts at promoter regions of SCZ GWAS risk loci, with enrichment for genes regulating histone modification, DNA repair, and synaptic signaling. We also found that genes with significantly decreased expression were enriched for DNA repair pathways, especially homology directed repair. Taken together, these findings suggest that increased genomic instability and reduced DNA repair may undergird a potential mechanism of SCZ genetic risk. SETD1A LoF pluripotent stem cell models provide an important avenue to further investigate these mechanisms and identify therapeutic targets.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1475 Exploring the common genetic factors modifying the germline de novo mutation rate

Authors:

O. Garcia Salinas[1], S. Hwang[2], Q. Huang[1], A. Scally[2], R. Rahbari[3], H. Martin[4]; [1]Wellcome Sanger Inst., Cambridge, United Kingdom, [2]Univ. of Cambridge, Cambridge, United Kingdom, [3]Wellcome Trust Sanger Inst., Cambridge, United Kingdom, [4]Wellcome Sanger Inst., Hinxton, United Kingdom

Somatic mutations accumulate across all cell types during the normal ageing process. However, germline cells accumulate these mutations at a rate at least 27 times lower than somatic cells (Moore et al., 2021). Approximately 80% of the variance in germline *de novo* mutation (DNM) rate is explained by parental age and technical covariates, while the rest of the variance remains unexplained (Kaplanis et al., 2022). Previous studies suggest that genetics could contribute to DNM variance among unrelated individuals, but these studies lacked sufficient statistical power to test this hypothesis effectively (Sassani et al., 2019).

In this study, we capitalised on one of the largest DNM catalogues, consisting of approximately 11,000 trios from the Genomics England initiative, to investigate the role of common variants in modifying the DNM rate. We carried out a genome-wide association study for DNM rate in mothers and fathers separately and combined. The SNP heritability, as estimated by LD score regression, was not significantly different from zero, and no genome-wide significant hits were obtained. However, recent work has shown that a polygenic score (PGS) for age of menopause (which is linked with DNA repair and depletion of the ovarian reserve) is significantly associated with decreased DNM rate in mothers (Stankovic et al., 2022), suggesting there is a polygenic contribution at least in females, but that the heritability may just be too low to detect at this sample size. We then tested for associations between DNM counts and PGSs for 82 complex traits, including cancers, as well as other non-cancer diseases, and non-disease continuous phenotypes.

We found that both parents PGSs for reproductive tract malignancies, such as ovarian and cervical cancers, were significantly associated with total DNM counts per trio (p &lt 0.06, FDR 10%). PGSs for age at smoking initiation, were also nominally significantly associated with total DNM counts (p &lt 0.06, FDR 10%), suggesting that earlier initiation of tobacco use may contribute to an increased DNM rate.

In summary, we present exploratory analyses investigating the role of common genetics in modifying the DNM rate, finding that the SNP heritability is not significantly different from zero. However, we found suggestive evidence that PGSs for certain cancers and for smoking initiation are associated with DNM rate, which we now seek to replicate in an independent sample.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1476 Exploring the genetic basis of Leukodystrophies in resource-poor settings through an in-house targeted panel approach

Authors:

Y. Sharma[1], A. G. Saini[1], P. Bhatia[1], S. Attri[2], P. Srivastava[1], N. Sankhyan[1], R. Suthar[1], V. Bhatia[1], S. Vyas[1]; [1]Postgraduate Inst. of Med. Ed. and Res., Chandigarh, India, [2]Postgraduate Inst. of Med. Ed. & Res., Chandigarh, India

Background: Leukodystrophies are a group of genetically determined neurodegenerative disorders that affect predominantly the white matter. We aimed to study the clinical, radiological and genetic features of childhood-onset Leukodystrophies by in-house genetic testing using a targeted genetic panel. **Methods**: All children with a clinico-radiological suspicion of leukodystrophy were evaluated by an in-house Ion Torrent next-generation sequencing technology. A custom gene panel was designed with the most common causes of leukodystrophies in the given setting. Genes included in the panel were *GFAP, ASPA, EIF2B1-B5, GALC, L2HGDH, DARS2, EARS2, MLC1, HEPACAM, ARSA, PSAP, SUMF1, PLP1, GJC2, RNASET2, ALDH3A2, POLR3A, POLR3B* and *ABCD1*. **Results**: Twenty two children with suspected leukodystrophies were tested by our in-house panel. Of these, 14 patients (63%) were genetically confirmed. The most common diagnoses were: X-ALD 42.85% (n=6), Vanishing-white-matter disease 21.42% (n=3), L-2-hydroxyglutaric aciduria 14.28% (n=2), Metachromatic leukodystrophy 7.14% (n=1), Alexander disease 7.14% (n=1) and POL-III related leukodystrophy 7.14% (n=1). The pathogenic variations were detected in *ABCD1*, *EIF2B5, GFAP, L2HGDH, ARSA* and *POLR3A* genes respectively. Majority of cases were males 71.42% (n=10/14). The mean age at presentation was 7.8 years (range 0.3-15 years). The common clinical features were developmental delay (100%), increased tone (60%), gait impairment (50%) and seizure (40%). Magnetic resonance imaging showed characteristic bilateral symmetrical white matter involvement in all children. **Conclusion:** The advent of next generation sequencing has helped in the early genetic confirmation and prenatal counselling for these disorders. Use of in-house targeted genetic panels help in significant cost reduction and feasibility for the patients in resource-poor settings.

Keywords: *Leukodystrophies, neurodegenerative disorders, genetic panels, pathogenic variations, next-generation sequencing*

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1477 Exploring the genetic landscape of obesity through variant-level, gene-level, sex-specific, and pathway enrichment analyses

Authors:

N. Baya[1,2,3], F. H. Lassen[1,2], B. Hill[2], S. S. Venkatesh[1,2], T. Ferreira[2], S. R. Myers[1,3], D. S. Palmer[2], C. M. Lindgren[1,2]; [1]Wellcome Ctr. for Human Genetics, Oxford, United Kingdom, [2]Big Data Inst., Univ. of Oxford, Oxford, United Kingdom, [3]Dept. of Statistics, Univ. of Oxford, Oxford, United Kingdom

Obesity, a pervasive health challenge, demands a comprehensive understanding of its genetic underpinnings. Leveraging UK Biobank's whole exome sequencing data we performed variant and gene burden analysis for three obesity-related traits (n>377,000): body mass index (BMI), waist-to-hip ratio adjusted for BMI (WHRadjBMI), and body fat percentage (BFP). Secondary analyses included six traits derived from DXA scans (android fat percentage, gynoid fat percentage, total body fat percentage, android-gynoid fat percentage ratio) and MRI scans (abdominal fat ratio, visceral adipose tissue volume). Our goal was to unravel obesity's causal mechanisms and provide novel insights into its genetic architecture.

In our variant-level analysis, we identified 643 non-synonymous genome-wide significant (GWS: $P<5x10^{-8}$) variants associated with BMI (177 variants), WHRadjBMI (270), BFP (193), gynoid fat percentage (2), and android-gynoid fat percentage ratio (1). Known obesity loci were replicated (e.g. *MC4R*, *PDE3B*, and *PLIN1*) and possible novel associations were discovered in *TRMU* (chr22:46335792:G:T, $P_{BFP}=4.9x10^{-9}$, effect allele frequency=0.89, predicted damaging missense) and *SLC17A3* (chr6:25862238:C:T, $P_{BMI}=2.1x10^{-8}$, $P_{WHRadjBMI}=2.4x10^{-8}$, $P_{BFP}=3.6x10^{-8}$, effect allele frequency=0.76). We performed a sex-differential analysis and identified 15 non-synonymous GWS variants for WHRadjBMI. All these variants lie in genes that were also significant in our sex-combined analysis, except for one variant (chr3:101347873:T:A, $P_{diff}=1.8x10^{-8}$, effect allele frequency=0.62, $beta_{female}=-0.021$, $beta_{male}=0.006$) in *SENP7*.

Next, we performed gene burden analysis and identified 20 significant genes (Cauchy combination test $P<2.5x10^{-6}$, Bonferroni correction) for BMI (6 genes), WHRadjBMI (13), and BFP (5). Four genes were significant in multiple phenotypes (*MC4R*, *PLIN4*, *HECTD4*, and *GPM6A*). To our knowledge, *GPM6A* is a novel association for BMI and BFP. The female- and male-only analyses identified 11 and 6 significant genes, respectively, including 3 genes significant exclusively in the male-only analysis (*GIGYF1*, *HERC6*, and *PLXNA2*).

Gene set enrichment analysis using curated gene sets highlighted pathways associated with WHRadjBMI including extracellular matrix organization ($P=3.1x10^{-7}$), adipogenesis ($P=1.0x10^{-6}$), and integrin signaling ($P=1.7x10^{-6}$). BFP enrichment analysis implicated cell cycle regulation, including the mitotic metaphase and anaphase pathway ($P=1.4x10^{-6}$).

Our study furthers the understanding of the genetic architecture of obesity and provides opportunities for interventions in combating this health burden.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1478 Exploring the genotypic and phenotypic significance of Polycystic Kidney Disease-2 (PKD2) variants in the UK Biobank using REVEAL

Authors:

Z. Pitluk[1], S. Sarangi[1], J. Zahiri[1], C. Guerriero[2], J. Brodsky[2]; [1]Paradigm4, Waltham, MA, [2]Univ. of Pittsburgh, Pittsburgh, PA

Polycystic Kidney Disease (PKD) is a progressive, irreversible, genetic disorder that is characterized by the development of multiple cysts within the kidney, leading to loss of function over time. There are two main types of PKD - autosomal dominant (ADPKD) and autosomal recessive (ARPKD). The former is one of the most common genetic disorders affecting 1 in every 400 to 1000 adults, while ARPKD is relatively rare, affecting 1 in 20,000 children worldwide. Management and treatment of ADPKD comes with a significant economic and societal cost, and despite several clinical studies only one medication has been approved to slow disease progression. Instead, dialysis is the only option for patients who ultimately develop renal failure. Thus, there is a need to develop novel drug targets and elucidate risk factors that could lead to early intervention with existing treatment plans. Mutations in the gene *PKD2,* which encodes polycystin-2 (PC2), have been linked with ADPKD. PC2 is a multi-pass membrane protein that functions as a cation permeable cation channel, and is involved in calcium transport and signaling in the cilia in renal epithelial cells. Previous studies uncovered associations between *PKD2* variants and ADPKD-relevant phenotypes (e.g., serum polyuria and hypertension). In this study, we furthered previous findings by exploring associations between variants of *PKD2* in whole exome sequence data from 470K individuals in the UK Biobank (UKBB) (Application ID: 51518) and ADPKD-relevant phenotypes. We utilized REVEAL, a bioinformatics platform developed by Paradigm4, to do large scale GWAS with PLINK, SAIGE, and REGENIE algorithms. We used ADPKD-relevant phecodes for more accurate case-control cohort selection, and tested hundreds of metabolomic biomarkers from the UKBB dataset. We also performed large scale linkage disequilibrium plus a burden test analysis between *PKD2* variants and related genes (e.g., *PKD1, GANAB*) to gain insights on network effects. In a complementary approach, based on our initial results, we have begun to select polymorphisms and examine the efficiency of the corresponding proteins to mature and function in yeast and mammalian cells. Preliminary data using the yeast model indicates that one polymorphism in the PC2 sequence identified in the UK Biobank affects PC2 biogenesis and/or activity. Ultimately, the objective of this work is to identify novel targets that impact PKD by employing a strategy that combines *in silico* genomic analysis with a functional candidate-based experimental yeast screen, along with studies in higher cells, that could lead to development of therapeutic strategies for PKD.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1479 Exploring the Gut Microbiome, Gut Metabolites, and Metabolic Pathways in Dyslipidemia; A Metagenomic Approach

Authors:

S. Raza[1], H-L. Kim[2], Y. Chang[3], S. Ryu[4], J. Cho[1], H-N. KIM[1]; [1]Samsung Advanced Inst. for Hlth.Sci. & Technology, Sungkyunkwan Univ., Seoul, Korea, Republic of, [2]Ewha Womans Univ, Seoul, Korea, Republic of, [3]Kangbuk Samsung Hosp., Seoul, Korea, Republic of, [4]Kangbuk Samsung Hosp., Sungkyunkwan Univ., Seoul, Korea, Republic of

Dyslipidemia is a complex metabolic disorder characterized by abnormal lipid profiles and an increased risk of cardiovascular diseases. Recent evidence suggests that the gut microbiome and its metabolites play a significant role in lipid metabolism and the development of dyslipidemia. In this study, we aimed to investigate the composition of the gut microbiome, gut metabolites, and associated metabolic pathways in individuals with dyslipidemia using a metagenomic approach. This cross-sectional study included 1384 adults, with 895 individuals in the dyslipidemia group and 489 healthy controls. Stool samples were collected from each participant, and metagenome sequencing was performed. Our findings revealed that four bacterial species, including *Bacteroides plebeius* and *Bacteroides stercoris*, were more enriched in the dyslipidemia group, while four species, including *Dorea longicatena* and *Bifidobacterium pseudocatenulatum*, were more abundant in the control group. Furthermore, we found 47 significantly different metabolites and 18 metabolic pathways between the two groups. Additionally, the 4 pathways were found to be more active in dyslipidemia cases, with *Bacteroides plebeius* contributing to the pyrimidine deoxyribonucleosides salvage pathway. The results of our study highlight the potential of *Bacteroides plebeius* and *Dorea longicatena* as major indicators of a distorted lipid profile. Their association with various gut metabolites and metabolic pathways could potentially play a pivotal role in unraveling the link between dyslipidemia and cardiovascular diseases. This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2023R1A2C2006416).

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1480 Expression analysis of circulating miRNAs and their target genes in the pathogenesis of chronic kidney disease patients of Pakistan.

Authors:

M. Khan[1], M. Ejaz[1], A. Liaquat[2], H. Salim[1], A. Dandare[1]; [1]COMSATS Univ. Islamabad, Islamabad, Pakistan, [2]Shifa Tameer-e-Millat Univ. Islamabad, Pakistan, Islamabad, Pakistan

Micro RNAs (miRNAs) are small non-coding sequences that play potent role in regulating the gene expression. The dysregulation of gene expression can disrupt many important biological processes and pathways fostering onset of several diseases including kidney diseases e.g., chronic kidney disease (CKD). The main aim of this research was to identify and explore the potential role of miRNAs as a biomarker in prognosis, diagnosis, and treatment of CKD during both early and late stages. The possible role of miRNA in CKD pathophysiology was explored through *in silico* analysis. Datasets of gene expression profiling by array were obtained through Gene Expression Omnibus (GEO), subjected for statistical analysis in GEO2R with threshold of $p<0.05$ to identify differentially expressed genes (DEGs). miRNA associated with CKD were acquired through an extensive literature scrutinization, followed by identification of miR-DEGs and functional enrichment analysis. miR-17-5p, miR-548c-3p, miR-132-3p, miR-145-5p, miR-143-3p and miR-21 indicated strong association with CKD regulating several genes involved in GTPase activity, cascade signalling, signal transduction and regulating transcription processes. Furthermore, we found significant association of these miRNA in several pathways promoting CKD pathogenesis. Next, the relative expression of all these miRNAs and their target genes was measured through qPCR in 100 healthy controls and 150 CKD Pakistani patients to validate the results of *in silico* analysis. The *in silico* and wet lab approaches employed in this research embodies inclusive role of ascertained miRNAs and their target genes as potential biomarkers for CKD. Further research can be carried out to develop miRNAs as biomarkers to modulate the dysregulated pathways and for diagnosis of CKD in early stages.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1481 Expression QTL-based phenome-wide association study identifies a link between *IL6ST* and polymyalgia rheumatica

Authors:

S. Lessard[1], F. Finngen[2], K. Klinger[1], E. de Rinaldis[1], S. Khader[1], J. Sloane Lazar[1], C. Chatelain[1]; [1]Sanofi, Cambridge, MA, [2]Univ. of Helsinki, Helsinki, Finland

Understanding the pleiotropic associations of a given therapeutic target can inform on potential indication expansion opportunities. Here, we propose to incorporate eQTL in a phenome-wide association study (PheWAS) context to identify shared associations within specific tissue and cell type contexts while informing on directionality. Focusing on genes related to the IL6 receptor (IL6-R), namely *IL6*, *IL6ST*, and *IL6R*, we performed MR and colocalization analysis between 2,405 disease GWAS from FinnGen Release 10 and eQTL from 28 datasets covering 118 tissues and cell types. For a given gene, we tested all phenotypes with at least one associated variant within 250kb (*P*-value $< 1x10^{-6}$). We report gene-disease pairs with MR q-value < 0.05 and colocalization posterior probability > 0.8. We identified putative causal links between increased *IL6* expression in CD16 monocytes and increased risk of varicose veins, ischemic heart disease, coronary atherosclerosis, and atrial fibrillation (MR beta > 0), but decreased risk of allergic asthma (MR beta < 0). Similarly, *IL6R* expression in artery, colon, and esophagus was associated with increased risk of coronary revascularization, coronary atherosclerosis, and abdominal aortic aneurysm, but lower risk of lower respiratory diseases and atopic dermatitis, Finally, increased *IL6ST* expression in T cells and whole blood is predicted to increase risk of rheumatoid arthritis, systemic connective tissue disorders, polyarthropathies, other arthritis, autoimmune diseases, and polymyalgia rheumatica. These associations were driven by rs7731626 (SuSIE fine-mapping probability >0.99). In summary, our analysis captured the known associations between the IL-6R pathway and cardiovascular and immune diseases and identified a novel association between *IL6ST* and polymyalgia rheumatica (rs7731626 $P=3x10^{-10}$, beta=-0.17). *IL6ST* encodes a protein responsible for signal transduction for the IL-6R. Sarilumab, an IL-6R inhibitor, has recently been approved in the US to treat polymyalgia rheumatica. Our results provide supporting genetic evidence that blockage of the IL6-R pathway can be beneficial to treat this condition.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1482 Extensive sex difference in regulatory elements in the developing human cortex and their plausible role in autism.

Authors:

D. Karunakaran[1], H. Berk-Rauch[1], K. Hennick[2], R. Leonard[2], J. Mitra[1], T. Nowakowski[3], A. Chakravarti[1]; [1]NYU Sch. of Med., New York, NY, [2]UCSF - Sch. of Med., San Francisco, CA, [3]

Autism spectrum disorder (ASD) is characterized by a sex ratio of 2-4 (male:female). Genomic studies have identified numerous genes with coding and noncoding variants contributing to ASD susceptibility, however, the causes of its sex difference are unknown. We hypothesize that ASD sex differences arise from enrichment of X-linked ASD gene variation, autosomal transcription factors (TFs) modulating X-linked ASD genes, and sex-biased cis regulatory element (CRE, enhancer) activity regulating ASD gene expression. Sex bias in the epigenome can arise in two ways: differential chromatin accessibility of CREs between sexes or interactions of ASD gene enhancers with cis acting sex hormone receptor (AR, ESR1, ESR2) binding sites. Note, that there is emerging evidence of variable and sex biased expression of AR, ESR1 and ESR2 during human fetal brain development.
In this study, we characterize sex bias in chromatin accessibility in multiple tissues of the developing fetal cortex and ganglionic eminences, brain regions which contribute to ASD phenotypes, from 18 individuals using ATAC-seq. We use machine learning methods to construct epigenomic reference maps from high-quality (post QC) data for each sex. Overall, we identified 252,099 index CREs of which 244,847 (97.1%), 6,813 (2.7%) and 459 (0.2%) were autosomal, X-linked and Y-linked, respectively. Consequently, there is a paucity of CREs on sex chromosomes relative to their size. We recognize that some, not all, Y chromosome CREs may be mapping artifacts. Within each sex, we identified 215,507 CREs in 9 females and 191,140 in 7 males, showing a small (13%) but significant excess of female CREs ($P < 10\text{-}3$), mostly autosomal. Our analyses of these sex-specific reference maps reveals that ~77.5% (157,797) of the CREs are shared between sexes but significant numbers are also sex-specific: 57,710 and 33,275 in females and males, respectively. These also demonstrate a female excess. Beyond these, we tested which CREs had significantly differential accessibility between sexes. We identified 2,181 CREs with a 4-fold difference and 21,166 CREs with a 2-fold difference (FDR <1%).
These results uncover extensive sex difference in the epigenome of the developing human cortex. We are currently identifying which specific genes these sex biased CREs modulate, the degree to which they affect gene expression and whether CRE sequence variants, in ASD families from the Simons Simplex Collection, are associated with ASD risk. Additionally, we are constructing cell type-specific sex-specific epigenome maps using 10X Chromium single cell Multiome ATAC + RNA analyses from 52 individuals ranging from mid-gestation to early childhood.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1483 Familial Hypobetalipoproteinemia caused by concurrent protein-truncating variants in *APOB* and *PCSK9.*

Authors:

A. Albuquerque[1], M. W. Lauria[1], C. Borlido Schuffner[1], N. K. Ribeiro de Almeida[1], E. Friedman[2], L. De Marco[3]; [1]Faculdade de Med., Univ.e Federal de Minas Gerais, Belo Horizonte, 30130-100, Brazil, Belo Horizonte, Brazil, [2]The Meirav High Risk Clinic, Sheba Med. Ctr., Ramat Gan, Israel, [3]Dept.s of Surgery and Med., Univ.e Federal de Minas Gerais, Belo Horizonte, Brazil

Introduction: Familial hypobetalipoproteinemia (FHBL) is a rare disease characterized by low LDL-C levels, lipid and lipid-soluble vitamin malabsorption, hepatic steatosis, and apparent protection from coronary artery disease. The full phenotype is associated with protein truncating variants (PTVs) in the *APOB* gene with a milder phenotype related to less deleterious sequence variants in the *APOB* and *PCSK9* genes. Here we report a patient with FHBL caused by concomitant loss of function (LOF) sequence variants in both the *APOB* and *PCSK9* genes. Subjects and methods: The proband, an asymptomatic 25-year-old Brazilian man was diagnosed with non-alcoholic steatohepatitis and fibrosis heralded by abnormal liver function tests. His family history was noted for two maternal great-uncles who died of non-alcoholic cirrhosis. Lipid profile revealed low LDL-C, HDL-C, and vitamin E. His daughter, brother, mother, and maternal grandmother also exhibited low LDL-C. Whole Exome Sequencing (WES) was performed on peripheral blood DNA from the proband. Results: WES revealed two heterozygous PTVs; one in *APOB* (p.Q1551*) and the other in *PCSK9* (C679*). The *APOB* variant was classified as Pathogenic and had a 0% allele frequency in databanks. The *PCSK9* variant was also Pathogenic and has been associated with low LDL-C. Its allele frequency in the Brazilian databank is 0.00085% despite high prevalence (0.79%) in African population (gnomAD). Conclusion: This patient exhibits two heterozygous PTV in genes involved in cholesterol metabolism, leading to low LDL-C levels and hepatic steatosis. The pathophysiology of FBHL involves ineffective packaging of LDL-C and lipoprotein reuptake due to PTVs in APOB, the ligand of the LDL receptor (LDLR). PCSK9 is a protease that degrades LDLR, hence LOF variants increase LDLR-mediated LDL-C uptake causing hypocholesterolemia. Both variants detected seemingly and in a biologically plausible manner synergistically lower LDL-C levels, accounting for the hepatic fatty acid buildup. While his very low LDL-C is protective against atherosclerotic-related diseases, it seems to cause a not yet treatable form of hepatic steatosis.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1484 Family study of the distribution pattern of polygenic risk scores in families with ulcerative colitis

Authors:

J. Radvanszky[1,2,3], I. Lojová[1,2,3], J. Minarik[1,3], Z. Pös[1,2], M. Hlavacka[4], A. Hurtuk[5], O. Kubicka[5], N. Forgacova[1,2,3], A. Zatkova[1], J. Gazdarica[2,6], J. Budis[2,6], T. Szemes[2,3]; [1]Inst. of Clinical and Translational Res., BioMed. Res. Ctr. of the Slovak Academy of Sci., Bratislava, Slovakia, [2]Comenius Univ. Sci. Park, Bratislava, Slovakia, [3]Dept. of Molecular Biology, Faculty of Natural Sci., Comenius Univ., Bratislava, Slovakia, [4]Dept. of Applied Informatics, Faculty of Mathematics, Physics and Informatics, Comenius Univ., Bratislava, Slovakia, [5]Inst. of Informatics, Information Systems and Software Engineering, Faculty of Informatics and Information Technologies of Slovak Univ. of Technology in Bratislava, Bratislava, Slovakia, [6]Slovak Ctr. of Scientific and Technical Information, Bratislava, Slovakia

Introduction: Genomic predispositions for inflammatory bowel disease (IBD), including ulcerative colitis (UC), may have complex patterns, including both monogenic, oligogenic and highly polygenic forms. We present our findings on: 1) the effect of four different selection strategies for choosing potentially relevant genomic positions for polygenic risk score (PRS) calculations; 2) the patterns of PRS distributions among UC affected and unaffected family members; and 3) the results of screening for potential monogenic pathogenic variants. **Material and Methods**: Whole-genome sequencing data were generated for 62 individuals, including 7 UC patients, 7 family members (three families), and 48 controls. Risk positions were extracted from GWAS Catalog. PRSs were calculated using the linear weighted sum of risk alleles. Monogenic causes were screened using a commercial annotation tool. **Results**: We proved the potential of PRSs in identifying at-risk individuals, but also in selecting patients for screening of monogenic/oligogenic UC. Considering the different variant sets tested, those obtaining the widest range of genomic positions (i.e. positions associated with traits IBD, including both UC and CD), were found to have the largest discriminatory power between patients and controls. Family members have not always increased PRSs. We also identified a novel variant/gene-phenotype association, possibly causing mono-/oligogenic UC. The significance of the identified ZBP1 variant (NM_030776.3:c.3G>A) will, however, require further studies. **Conclusion**: PRSs promise breakthroughs in prevention, diagnostics and therapy of complex diseases, especially when combined with screening for monogenic genetic causes. **Grants**: This work has been supported by the Operational Program Integrated Infrastructure within project ITMS:313021BUZ3, co-financed by the ERDF; and by the Scientific Grant Agency (VEGA_2/0146/23).

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1485 Filtering Signal From Noise: Finding Potential lncRNA biomarkers in IBD Disease Pathogenesis

Authors:

M. Kennedy Ng[1], S. Silverstein[1], C. Beasley[2], K. Darlington[1], M. Schaner[1], A. Jain[1], M. D. Long[1], E. L. Barnes[1], J. J. Hansen[1], M. Kapadia[1], J. Guillem[1], J. Stem[1], J. Calabrese[1], S. Sheikh[1], T. Furey[1]; [1]Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, [2]Emory Univ., Atlanta, GA

Background Crohn's Disease (CD) and Ulcerative Colitis (UC) are two types of inflammatory bowel disease (IBD) but with different clinical presentations; this includes location of the disease, contiguity of inflammation, and complications that arise. Overall, the molecular underpinnings that drive CD and UC are not well understood. Long non-coding RNAs (lncRNAs) have been shown to play numerous regulatory roles in modifying gene expression, making them an exciting area of study. Previous studies have identified IBD-associated lncRNAs, but the study of lncRNAs has been hampered because most lncRNAs are lowly expressed and are highly tissue and context specific, leading to an incomplete annotation and a lack of replicability. **Hypothesis** We hypothesize that by implementing a framework for lncRNA discovery and robustly quantifying both known and previously unannotated lncRNAs, we can better pinpoint lncRNAs that play a role in CD pathogenesis. Specifically, our aims are to (i) discover unannotated lncRNA transcripts; (ii) accurately quantify expression levels for previously annotated lncRNAs and novel lncRNAs; (iii) identify differential expression patterns across multiple CD patient cohorts. **Methods** We used genome-guided de-novo transcript assembly and filtration pipeline to discover and quantify well-supported unannotated lncRNAs. We then added these samples to our genome reference, and quantified these transcripts along with known lncRNAs, using RNA-seq data from uninflamed colonic mucosa tissue samples from adult CD patients and non-IBD controls across multiple cohorts. We refined these lists to target lncRNAs likely to be functionally relevant; these include lncRNAs with highly conserved promoter regions, non-spurious transcriptional activity, and nucleotide similarity to well-characterized lncRNA transcripts. **Results** We identified ~4500 well-supported unannotated lncRNAs. We also identified hundreds of unannotated and known lncRNAs that are differentially expressed between CD and compared to non-IBD patients in multiple datasets, as well as many lncRNAs that stratify patients by disease activity. Based on our refinement parameters, we have discovered a handful of lncRNAs (both known and unannotated) that are not characterized in the context of IBD and have characteristics that warrant future study. **Conclusion** This work will help us better identify and understand the molecular changes that underlie IBD.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1486 † Fine-Mapping Million Veterans Program gwPheWAS Reveals Novel Variants and Elucidates Genetic Architecture in Diverse Populations

Authors:

M. Conery[1], A. Verma[2,1], J. E. Huffman[3,4], A. A. Rodriguez[5], M. Liu[6], Y-L. Ho[7], Y. Kim[5], D. A. Heise[8], V. A. Panickan[9], H. Garcon[4], F. Linares[8], L. Costa[7], I. Goethert[8], J. Honerlaw[4], J. Moser[10], S. B. Whitbourne[7], J. Cohen[8], G. Tourassi[8], MVP PheWAS Working Group, S. Pyarajan[7], P. S. Tsao[11], C. J. O'Donnell[7], S. Muralidhar[10], J. Casas[4], A. G. Bick[4], S. F. A. Grant[1,12], T. Cai[9], B. F. Voight[2,1], K. Cho[4], J. M. Gaziano[7], R. K. Madduri[5], S. M. Damrauer[2,1], K. P. Liao[7]; [1]Univ. of Pennsylvania, Philadelphia, PA, [2]Corporal Michael Crescenz VA Med. Ctr., Philadelphia, PA, [3]Palo Alto VA Inst. for Res., VA Palo Alto Heath Care System, Palo Alto, CA, [4]MAVERIC, VA Boston Hlth.care System, Boston, MA, [5]Argonne Natl. Lab., Lemont, IL, [6]Columbia Univ.'s Mailman Sch. of Publ. Hlth., New York, NY, [7]VA Boston Hlth.care System, Boston, MA, [8]Oak Ridge Natl. Lab., Dept. of Energy, Oak Ridge, TN, [9]Harvard Med. Sch., Boston, MA, [10]Dept. of VA, Washington, DC, [11]VAPAHCS/Stanford Univ., Palo Alto, CA, [12]Children's Hosp. of Philadelphia, Philadelphia, PA

Though genome-wide association studies (GWAS) have accelerated our understanding of the phenotypic and biological consequences of genetic variation across the genome, a significant diversity gap exists in the characterization of genetic architecture due to historical overrepresentation of individuals genetically similar to European (EUR) reference populations. In contrast to most GWAS cohorts, the Department of Veteran Affairs (VA) Million Veteran Program (MVP) consists of nearly 30% representation from non-EUR populations. This study was performed as part of a larger effort to detect genetic associations across 2,070 phenotypes obtained from linked electronic health records (EHR) and questionnaires in 635,969 MVP participants genetically similar to African (AFR, n=121,177), Admixed American (AMR, n=59,048), East Asian (EAS, n=6,702), and EUR (n=449,042) reference superpopulations from the 1000 Genomes Project. Fine-mapping was performed on 1,257 phenotypes in which one or more significant loci were detected in a multi-population meta-analysis (P-value < 4.6×10^{-11}). Employing the Department of Energy's (DOE) Andes supercomputer, in-sample linkage disequilibrium (LD) reference panels, the Sum of Single Effects (SuSiE) framework, and a previously developed multi-population signal merging approach, we fine-mapped 57,601 signals, ~25% of which were fine-mapped to a single variant. Among the signals, we identified 2,069 variant-phenotype associations mapped with high confidence (posterior inclusion probability > 0.95) exclusively in non-EUR populations. These included novel variants, such as rs7338263 which associates with a lower white blood cell (WBC) count in AFR only, informing longstanding epidemiologic observations of lower WBC in AFR relative to EUR populations. We additionally observed high-confidence fine-mapped variants with different effects across populations; rs429358 at the *APOE* locus, for example, had a significantly larger effect on risk of dementia in EUR compared to the non-EUR populations, supporting findings from prior multi-population studies. Our findings represent a significant advance in understanding the genetic architecture of complex human traits in non-EUR populations and highlight the need and possible benefit of further expanding participation in GWAS efforts among these groups. Furthermore, this collaboration between the VA and DOE has significantly expanded the corpus of fine-mapped genotype-phenotype associations available to the global research community.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1487 First GWAS based replication of candidate genes association with Adolescent Idiopathic Scoliosis in Northwest Indian population.

Authors:

H. Singh[1], S. Kowra[2], M. Gupta[3], N. Gupta[4], G. Gupta[5], A. Pandita[6], R. Sharma[7], S. Pandita[8], V. Singh[9], B. Garg[3], E. Rai[10], S. Sharma[11]; [1]Shri Mata Vaishno Devi Univ., Katra, India, Jammu, India, [2]SMVDU, Jammu, India, [3]Dept. of Orthopaedics, All India Inst. of Med. Sci., New Delhi, New Delhi, India, [4]Consultant Orthopedic Surgeon, District Hosp. Poonch, Jammu and Kashmir, Poonch, India, [5]Dept. of Radiology, Shri Mata Vaishno Devi Narayana Superspeciality Hosp., Katra, Jammu and Kashmir, Katra, India, [6]Dept. of Orthopaedics, Accidental Hosp., Chowki Choura, Jammu, Jammu and Kashmir, Jammu, India, [7]Dept. of Radiology, Government Med. Coll., Jammu, Jammu and Kashmir, Jammu, India, [8]Dept. of Radiology, Chest Disease Hosp., Bakshi Nagar, Jammu, Jammu and Kashmir, Jammu, India, [9]Sch. of Biotechnology, Shri Mata Vaishno Devi Univ., Katra, Jammu and Kashmir, Katra, India, [10]Shri Mata Vaishno Devi Univ., Katra, India, [11]Ctr. for Molecular Biology, Central Univ. of Jammu, India, Samba, India

Scoliosis is the bending and twisting of the spine occurring at any age. Adolescent Idiopathic Scoliosis (AIS) is the most prevalent among all other types of scoliosis, with a worldwide prevalence of approximately 2-3%. Surprisingly, no genetic study on AIS was conducted in any Indian population.Blood/saliva samples were collected from 113 clinically confirmed (through anteroposterior X-ray) AIS cases, and 500 non-AIS controls were recruited in the study after obtaining ethical clearance from the Institutional Ethics Review Board at Shri Mata Vaishno Devi University and All India Institute of Medical Sciences in New Delhi. All the samples were genotyped in the population of North-west India using Infinium Global Screening Array-24 v3.0 chip. A total of 1163 SNPs were genotyped in 14 previously reported candidate Genes/regions (*AJAP1, MATN1, PAX3, VDR, SOX9, BCL2, DOT1L, PAX1, LBX1, COL6A3, FBN1, FBN2, COL11A2,* and *CALM1*). The Bonferroni correction threshold was set at 0.000043 based on the total number of variants. The statistical analysis was performed using PLINK v1.07, 1.09, R programming, IBM SPSS v 23, and Linux shell scripting. The functional role of the genes/variants was further evaluated using FUMA GWAS, MSigDB, and Cytoscape v 3.9.1 bioinformatic tools and databases.This genetic study has highlighted an association of newly identified variants of developmental and bone morphogenetic pathway genes, namely *AJAP1, BCL2, SOX9, COL6A3,* and *FBN2*, with AIS risk in North-west India. The present study found an association between the *SOX9* variant; rs74898711, rs9302936, and rs9900249, *AJAP1* variants; rs141899425, rs11591221, and rs7517857, *BCL2* variant; rs11591221, *COL6A3* variant rs77009782, and *FBN2* variant rs17697567, was found to pose an increased risk to AIS among the Northwest Indian population. All the associated variants followed the Hardy-Weinberg Equilibrium (HWE) within the control population dataset. The variants were found to lack functional significance at the protein or gene expression level. While the associated genes were found to play a crucial role in the musculoskeletal system. The gene enrichment analysis has demonstrated the role of *SOX9, FBN2,* and *COL6A3* in bone formation, mineralization, and endochondral ossification. The findings suggest that *AJAP1, BCL2, SOX9, COL6A3,* and *FBN2* genes are associated with increased AIS risk predisposition in the Northwest Indian population. Therefore, it can be inferred that these genes play a significant role in the pathogenesis of AIS in this population, and this study also strengthens the hypothesis of genetic heterogeneity among different populations.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1488 Fish oil supplementation modifies the genetic potential for blood lipids and cardiovascular disease

Authors:

Y. Sun, T. McDonald, A. Baur, H. Xu, N. Bateman, Y. Shen, K. Ye; Univ. of Georgia, Athens, GA

Dyslipidemia is well-known to be a risk factor for cardiovascular disease (CVD), which has been the leading cause of mortality worldwide. Although habitual intake of fish oil has been implicated in offering modest cardioprotective effects through triglyceride reduction, the interactions of fish oil with the genetic predisposition to dysregulated lipid profiles and CVD risk remain elusive. A total of 441,985 participants of predominantly European ancestry with genetic and complete data from the UK Biobank were included and prospectively followed up to 2023. Polygenic risk scores (PRS) were calculated in participants of diverse ancestries, and a higher PRS indicates a higher genetic susceptibility to elevated lipid levels or CVD risk. Multivariable linear regression models or multivariable Cox proportional hazards regression models were used to assess associations with adjustment for relevant risk factors. Fish oil supplementation mitigated genetic susceptibility to elevations in total cholesterol, low-density lipoprotein cholesterol (LDL-C), and triglycerides, while it amplified genetic susceptibility to increased high-density lipoprotein cholesterol (HDL-C) among 424,090 participants of European ancestry ($P_{\text{interaction}} < 0.05$). Consistent significant findings were obtained using PRS calculated based on multiple genome-wide association studies or alternative PRS methods. We also detected that fish oil significantly attenuated genetic predisposition to high triglycerides among African ancestry participants. During an average of 13.4 years of follow-up, there were 1,854 coronary artery disease (CAD) deaths, 3,577 CVD deaths, 23,327 incident CAD cases, and 28,417 incident CVD cases. Although fish oil was significantly associated with lower CAD and CVD mortality and incidence rates, it did not alter genetic predisposition to CAD and CVD mortality or incidence. In categorical analyses, the multivariable adjusted hazard ratios (HR) for CAD mortality among individuals with higher PRS (greater than/equal to the median score) were 0.71 (95% confidence interval 0.62-0.81) for those consuming fish oil versus non-users and 0.86 (0.73-1.01) among those with lower PRS; multivariable HR for CVD incidence were 0.94 (0.91-0.97) for higher PRS and 0.99 (0.95-1.03) for lower PRS. Fish oil supplementation attenuated the genetic potential for elevated blood levels of total cholesterol, LDL-C, and triglycerides, while accentuating genetic susceptibility to higher HDL-C. Moreover, we detected that there might be a more pronounced protective effect of fish oil supplementation on CAD mortality and CVD incidence in people at high genetic risk.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1489 Functional dissection of putative heart enhancers implicated in left ventricle function and cardiomyopathy

Authors:

Z. Weber[1], T. Monroe[2], E. McNally[3], M. Nobrega[1]; [1]Univ. of Chicago, Chicago, IL, [2]Northwestern Univ., Chicago, IL, [3]Northwestern Univ, Chicago, IL

Inherited cardiomyopathies have long been thought to be Mendelian disorders, driven by highly penetrant, rare coding mutations; albeit with variable clinical expressivity. Recently, with biobank-scale genetic cohorts, genome-wide association studies (GWAS) have suggested a more oligogenic trait liability underlying the inheritance and expressivity seen in the clinical environment. Although several novel loci have been implicated in GWAS of cardiomyopathies, most of these loci have not been experimentally validated, and the relatively rare nature of these disorders has limited the power of GWAS to further resolve candidate causal variants. As cardiomyopathies tend to manifest as extreme ends of diastolic and systolic heart dysfunction, we leveraged a well-powered cardiac MRI cohort in the UK Biobank to investigate loci implicated in inherited cardiomyopathies. To this end, we conducted GWAS in three left ventricle phenotypes: End-Diastolic Volume (LVEDV), End-Systolic Volume (LVESV), and Ejection Fraction (LVEF) and jointly meta-analyzed these trait associations using Multi-Trait Analysis of GWAS (MTAG). Results of the MTAG analysis and subsequent functionally informed fine-mapping prioritized two loci of interest, whose signals appear to result from variation in cardiac enhancers. The first signal appears to intersect a large enhancer complex downstream of WNT7a and upstream of SLC6A6 at the chr3p25.1 locus. This locus has been implicated in previous GWAS of dilated cardiomyopathy, but has yet to be validated experimentally. Furthermore, rare variants in the gene encoding taurine transporter SLC6A6 have been reported to cause childhood onset cardiomyopathy and retinal degeneration. We also identified a novel genome-wide association signal in a putative upstream enhancer of BAG3 at the chr10q26.1 locus. Rare variants in BAG3 are also known to cause inherited dilated cardiomyopathy -suggesting that this signal may be a promising candidate for further investigation. To experimentally interrogate these loci we developed a variant to function strategy for testing these enhancers for their potential effects on gene expression and myocardial function in iPSC-derived cardiomyocytes and engineered heart tissues. Using haplotype-resolved reporter assays and CRISPR engineered models, we can not only test whether enhancer variation is the likely source of these association signals, but can also probe potential epistatic effects between enhancer variants and rare-coding variants, and examine pleiotropic effects across cardiac development, contributing to our evolving understanding of complex trait liability in cardiomyopathies.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1490 Functionally informed fine-mapping of glycaemic trait loci improves fine-mapping resolution compared to an agnostic approach.

Authors:

J. Soenksen[1], J. Chen[1], A. Varshney[2], S. Martin[1], S. Parker[3], J. Asimit[4], A. Morris[5], I. Barroso[1], MAGIC; [1]Exeter Ctr. of Excellence for Diabetes Res. (EXCEED), Univ. of Exeter Med. Sch., Exeter, United Kingdom, [2]Univ. of Michigan, Canonsburg, PA, [3]Univ. of Michigan, Ann Arbor, MI, [4]MRC Biostatistics Unit, Cambridge, United Kingdom, [5]Univ. of Manchester, Manchester, United Kingdom

Introduction: Previously, the Meta-Analysis of Glucose and Insulin-related traits consortium (MAGIC) reported 242 loci associated ($P<5x10^{-8}$) with glycaemic traits 2h-Glucose (2hGlu), fasting glucose (FG), fasting insulin (FI) and glycated haemoglobin (HbA1c). However, at most locus-trait associations the underlying causal variant(s) remained unknown. To help narrow down the number of potential causal variants at these loci, we previously employed multi-trait fine-mapping with flashfm. This improved fine-mapping resolution in 99/116 locus-trait associations compared to single trait fine-mapping with FINEMAP. The median 99% credible set (99CS) size reduced from 50 to 19 SNPs ($p=2.2x10^{-16}$).
Aim: To further improve fine-mapping resolution by incorporating functional annotations.
Methods: We used fGWAS v.0.3.6 to build a model of enriched annotations by testing 28 static annotations, including conserved regions and 32 cell-type specific annotations such as stretch enhancers in islets. We used the resulting enrichments to calculate prior probability of causality for each SNP, and then used FINEMAP v1.4 to calculate posterior probabilities and credible sets accounting for 99% of the posterior probability of being causal (99CS).
Results: In agreement with previous results, we found that static annotations such as coding or conserved regions were enriched across multiple glycaemic traits (FG, 2hGlu and HbA1c) whilst cell-type specific annotations varied across traits. FG and 2hGlu results were enriched for islet stretch enhancers, while FI was enriched for adipose-specific stretch enhancers. The median 99CS size improved from 40 SNPs in functionally-agnostic fine-mapping to 30 in functionally-informed fine-mapping ($P=3.6x10^{-13}$) when looking at loci predicted to have a single causal variant. The number of locus-trait associations with 10 or less SNPs in the 99CS improved from 32 to 39, and the number of locus-traits with 1 SNP in the 99CS improved from 3 to 6.
Conclusion: Incorporating functional information into fine-mapping can reduce 99CS size compared to agnostic approaches. In the future, we will incorporate these priors and take advantage of the correlation between traits in multi-trait fine-mapping with flashfm to further improve fine-mapping resolution.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1491 *G6PD* genetic variants prevalent in British Africans and British South Asians: Under-identification of G6PD deficiency leads to delayed diagnosis of type 2 diabetes and reduced time to diabetes complications

Authors:

S. Martin, J. Chen, T. J. McKinley, K. G. Young, H. Green, G. Hawkes, R. N. Beaumont, A. R. Wood, I. Barroso; Univ. of Exeter, Exeter, United Kingdom

Glucose-6-phosphate dehydrogenase (G6PD) deficiency is an X-linked genetic disorder where a mutation of the protein produced by the *G6PD* gene causes haemolytic anaemia and leads to reduced red blood cell lifespan. The *G6PD* variant rs1050828, common in African Americans, is known to be associated with a corresponding reduction in glycated haemoglobin (HbA1c) levels independently of blood glucose. We aimed to determine how prevalent this and other *G6PD* variants are in the UK population, how many carriers are recorded as having G6PD deficiency, and the effect these variants have on age of type 2 diabetes diagnosis as well as time to diabetes complications.

We used whole-exome sequencing data from the UK Biobank to identify pathogenic variants for the *G6PD* gene, determining carrier status of these for 7394 individuals genetically identified as British African (AFR) and 9541 British South Asian (SAS), with 42% and 53% males respectively. Using electronic health records, we identified how many carriers were diagnosed with G6PD deficiency and/or haemolytic anaemia. HbA1c and fasting glucose (FG) were compared between carriers and non-carriers for individuals with no diagnosis of diabetes. Next, we identified individuals with a new diagnosis of type 2 diabetes since 2011, when HbA1c became the standard diabetes screening assay in the UK, and compared age of diagnosis and time to diabetic retinopathy and/or nephropathy between carriers and non-carriers. A dominant model was assumed in all analyses.

We identified 141 missense and 2 loss-of-function *G6PD* variants that we defined as pathogenic, 63 of which were singletons. The prevalence of carriers for at least one *G6PD* variant was 17% and 30% in AFR men and women respectively, and 4% and 7% in SAS. Despite this prevalence, 91% of AFR men and women are not diagnosed with either G6PD deficiency or haemolytic anaemia, with 100% and 96% in SAS men and women. In individuals without diabetes, HbA1c was on average 6.4mmol/mol lower in carriers compared to non-carriers in AFR and SAS ($p<5E-18$), while no difference was observed in FG. Carriers were on average diagnosed with type 2 diabetes 7.4 years later than non-carriers in AFR men ($p=0.02$), and diabetes complications arose 1.7 years earlier in SAS men ($p=0.04$). Our results were consistent when using only the rs1050828 variant in AFR.

Our results show that despite *GP6D* variants being common in British Africans and South Asians, the majority of carriers are not identified as having G6PD deficiency. Such misidentification may lead to delayed type 2 diabetes diagnosis with faster progression to diabetes complications, further increasing existing health inequities in these population groups.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1492 Genealogy-based Investigation of Epilepsy Genetics

Authors:

J. BOULBES[1], C. Moreau[1], P. Cossette[2,3], H. Vézina[1,4], S. Girard[1,5]; [1]UQAC, Chicoutimi, QC, Canada, [2]Université de Montréal, Montréal, QC, Canada, [3]Ctr. de recherche du CHUM, Montréal, QC, Canada, [4]Projet BALSAC, Chicoutimi, QC, Canada, [5]Ctr. de recherche CERVO, Québec, QC, Canada

Epilepsy is a neurological disorder characterized by temporary dysfunction of the brain, resulting in abnormal, excessive, or synchronized neuronal activity. It is estimated that 30% of epilepsies have a genetic cause, and there are over 500 potentially involved genes in the onset of epilepsy. However, a large portion of the heritability of these genes is missing or unexplained. Our objective is to develop a method for identifying and analyzing genetic changes related to the transmission and genetics of epilepsy at the scale of Quebec and to fill in this gap in information regarding heritability. We have access to a Quebec cohort of 348 epilepsy patients, their genotypes, and their deep genealogies. The current population of Quebec is the result of founder effects that have greatly influenced its genetic structure. By studying genotypes in conjunction with genealogical data, we can investigate the transmission of complex diseases with greater precision. An analysis of the number of recent common ancestors (MRCA) among pairs of patients and by region has allowed us to identify genealogical clusters. Genealogical kinship allows us to group families with affected individuals and determine the genealogical proximity between these patients which would not be possible using only 2-3 generations' pedigrees. A differential MRCA matrix of patients/controls enables us to identify a difference in population structure between the population and patient families. Finally, the genotype data will help us identify the identity-by-descent segments (IBD) transmitted within the identified family clusters. This study will enhance our understanding of the transmission of epilepsy-related genes at the individual, family, and regional levels in Quebec, thereby limiting bias related to population structures and identifying new responsible genes while furthering knowledge of the heritability of a complex disease, namely epilepsy.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1493 Gene-Environment Interactions and Their Impact on Nonalcoholic Fatty Liver Disease in Mexican Americans: Insights from south Texas.

Authors:

E. Manusov[1], V. P. Diego[2], X. Mao[2], M. Almeida[2], A. Leandro[2], S. Laston[2], J. Blangero[2], S. Williams-Blangero[2]; [1]One West Univ. Boulevard, Brownsville, TX, [2]Univ. of Texas Rio Grande Valley, Brownsville, TX

Background: Nonalcoholic fatty liver disease (NAFLD) is a chronic liver condition influenced by genetic, cellular, and environmental factors. Depression is associated with NAFLD, suggesting an interaction between genetic factors and depression in developing hepatic fibrosis. We investigated the association between ADAMTS7 gene variations and Beck Depression Inventory-II (BDI-II) scores, Alanine Aminotransferase (ALT), Aspartate Aminotransferase (AST), and Controlled Attenuation Parameter (CAP) values in Mexican Americans with NAFLD. Previous studies have linked ADAMTS7 gene variations to an increased risk of NAFLD and hepatic fibrosis, making it a potential genetic marker for identifying at-risk individuals, particularly among Mexican Americans. **Methods**: We included 279 Mexican American participants from a south Texas community. We measured hepatic fibrosis (Vibration Controlled Transient Elastography-VCTE), liver steatosis (Controlled Attenuation Parameter-CAP), Alanine Aminotransferase (ALT), Aspartate Aminotransferase (AST), and depression (Beck Depression Inventory-BDI-II). RNA sequencing was performed using the Illumina Novaseq 6000 platform. Gene-by-Environment (GxE) decomposition analysis was conducted to estimate heritabilities and GxE interactions. Transcript identification was made using Kallisto software against the UCSC hg19 reference transcriptome. **Results**: We found significant associations between ADAMTS7 gene variations and BDI-II scores, ALT, AST, and CAP values. These findings support the role of genetic factors interacting with depression in influencing hepatic fibrosis in Mexican Americans with NAFLD. **Discussion**: ADAMTS7, a gene involved in inflammation and Extracellular Matrix (ECM) regulation, emerges as a candidate gene implicated in liver fibrosis, lipid metabolism, coronary artery disease, and depression. The GxE interactions identified in this study provide insight into the complex interplay between NAFLD, depression, and genetic factors. **Conclusion**: Our study highlights the importance of GxE interactions in NAFLD and depression among Mexican Americans. The associations between ADAMTS7 gene variations, depression, and liver disease provide valuable insights into the inflammatory theory of depression and its impact on NAFLD. Further research is needed to elucidate the underlying molecular mechanisms and pathways and identify additional candidate genes contributing to these interactions.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1494 Generalizability of Alzheimer's disease plasma biomarkers of phospho-Tau and beta amyloid in individuals of diverse genetic ancestries

Authors:

A. Griswold[1], F. Rajabli[1], T. Gu[1], C. Garcia-Serje[1], J. Arvizu[1], C. Golightly[1], P. Whitehead[1], K. Hamilton-Nelson[1], L. Adams[1], K. Celis[1], J. Sanchez[1], S. Tejada[1], P. Mena[1], T. Starks[2], M. Cornejo-Olivas[3], M. Illanes-Manrique[3], C. Silva-Vergara[4], M. Cuccaro[1], J. Vance[1], B. Feliciano-Astacio[4], G. Byrd[2], G. Beecham[1], J. Haines[5], M. Pericak-Vance[1]; [1]Univ. of Miami, Miami, FL, [2]Wake Forest Sch. of Med., Winston-Salem, NC, [3]Inst. Natl. de Ciencias Neurologicas, Lima, Peru, [4]Univ. Central del Caribe, Bayamon, PR, [5]Case Western Reserve Univ, Cleveland, OH

Background: Plasma concentrations of phosphorylated threonine-181 of Tau (pTau181) and the ratio of amyloid beta isoforms Aβ42/Aβ40 are biomarkers for differential diagnosis and preclinical detection of Alzheimer disease (AD). However, measurement of these biomarkers is mostly from individuals of non-Hispanic, European ancestry. Given differences in AD risk, generalizability of these findings is not assured in individuals of diverse ancestry. Here we evaluate the utility of plasma pTau181 and Aβ42/Aβ40 in discriminating clinically diagnosed AD from cognitively intact, age-matched controls in ancestrally diverse, admixed cohorts. **Method:** We measured plasma pTau181 and Aβ42/Aβ40 with Simoa chemistry using the pTau181 AdvantageV2 and NEUROLOGY 3-PLEX A assays, respectively. Our cohorts consisted of: 642 African Americans (162 AD and 480 cognitively intact (CI)), 906 Puerto Ricans (385 AD and 521 CI), 149 Peruvians (49 AD and 100 CI), 60 Cubans (26 AD and 34 CI), and 246 non-Hispanic, European ancestry (22 AD and 224 CI). Linear mixed-effect regression models adjusted for age, sex, population substructure and relatedness followed by Bonferroni correction was applied to identify differences across AD status. Diagnostic performance and construct receiver operator characteristic (ROC) curves were created from logistic regression models. **Result:** Plasma pTau181 concentrations were increased in individuals with AD compared to CI ($p < 2x10^{-16}$) taking into account all individuals and in each cohort separately (African Americans, $p = 1.2x10^{-9}$; Puerto Ricans, $p=7.6x10^{-9}$; Peruvians, $p=0.02$), and European ancestry, $p=2.2x10^{-8}$) except for the Cubans where there was a trend. There was no significant difference in the plasma Aβ42/Aβ40 ratio, however there was a trend towards a decreasing concentration in AD. Using the area under the ROC, pTau181 was more accurate at predicting status than the Aβ42/Aβ40 ratio, but the classification improved when both biomarkers were combined. The accuracy varied widely over the individual cohorts with AUC from 0.845 for African Americans to 0.683 for Peruvians. **Conclusion:** These results suggest AD biomarkers are generalizable across ancestries, though the predictive value may differ depending on specific ancestral backgrounds. Ultimately, combining genomic and biomarker data from diverse individuals will increase understanding of genetic risk and refine clinical diagnoses in individuals of diverse ancestries

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1495 Genes and Cognition study, a recallable cohort to study Dementia/Alzheimer's disease

Authors:

M. Rahman[1], S. Hill[2], B. Tom[1], P. Chinerry[1]; [1]Univ. of Cambridge, Cambridge, United Kingdom, [2]Univ. of Manchester, Manchester, United Kingdom

Background: Decades of research have not resulted in a cure for dementia/Alzheimer's disease (AD). Most treatments have failed to benefit patients in clinical trials, potentially due to treatment being provided too late when brain damage is irreversible. There is an urgent need to understand the disease mechanism at the preclinical and prodromal stages of AD, which requires early identification of participants at risk of AD.
Method: We attempt to address this need by establishing an open cohort named Genes and Cognition (GC), nested within NIHR Bioresource (https://bioresource.nihr.ac.uk/) and comprising 21,051 healthy people aged 17-85 who consented to be recalled for follow-up studies. Participants took 11 cognitive tests (CTs) covering various domains of cognition. We examined the association between CT scores and demographic characteristics, explored phenotypic and genotypic correlations between CTs, and estimated SNP-based heritability. In addition, AD polygenic risk (high vs low) was determined in 10,038 participants using 20 single nucleotide polymorphisms (SNPs) from Lambert et al. (PMID:24162737) to identify the earliest age with a noticeable score difference in CTs.
Result: CT scores (higher score indicates poorer performance) were associated with age and gender differences were significant for each CT score. Significant linear trends were observed between CTs and educational attainment. Genetic correlations between tests were stronger than the phenotypic correlations. SNP-heritability for CTs and general cognitive ability ranged from 8-28%. We observed that three CTs (Reaction Time, Stroop Box, Stroop Ink) began deviating around age 55 between high ($\geq 96^{th}$ percentile) and low ($\leq 95^{th}$ percentile) AD polygenic risk groups, although not significant. Similar deviations in some CTs were observed among *APOE* E4 allele carriers compared to E3/E3 carriers. However, AD risk groups determined without including the *APOE* region indicated such deviation after age 65.
Conclusion: We presented preliminary findings for the GC study cohort. Our results suggest that AD polygenic risk measured with or without *APOE* region may be less beneficial than the two *APOE* SNPs (*rs429358* and *rs7412*) alone for the early identification of AD. Repeated CT measures would shed more light on this observation; therefore, we aim to investigate this with CTs repeated at two years intervals.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1496 Gene-Sleep Duration Interactions in the Genetic Architecture of Blood Pressure

Authors:

H. Wang[1], P. Nagarajan[1], R. Noordam[2], A. Bentley[3], A. Giri[4], H. Chen[5], W. Gauderman[6], A. Hung[4], T. Winkler[7], S. Redline[1], CHARGE Gene-Lifestyle Interactions Working Group; [1]Brigham and Women's Hosp., Harvard Med. Sch., Boston, MA, [2]Leiden Univ. Med. Ctr., Leiden, Netherlands, [3]NIH, Bethesda, MD, [4]Vanderbilt Univ. Med. Ctr., Nashville, TN, [5]The Univ of Texas Hlth.Sci. Ctr. at Houston, Houston, TX, [6]Univ of Southern California, Los Angeles, CA, [7]Univ. of Regensburg, Regensburg, Germany

Sleep duration may affect blood pressure traits, involving dysregulation of the sympathetic nervous system, endothelial dysfunction, or inflammation. To test the modification effect of sleep duration on the genetic variants for blood pressure, we conducted large-scale gene-by-short- and long-sleep duration interaction analyses on 3 blood pressure (BP) traits (systolic BP, diastolic BP, and pulse pressure [PP]). Analyses included a sample size of 810,724 individuals (6% African [AFR], 4% Hispanic/Latino [HIS], 6% East Asian [EAS], 83% European [EUR], 1% South Asian [SAS]) from 37 cohorts including the Million Veteran Program and the UK Biobank. Cohort-specific genome-wide interaction analyses were meta-analyzed using a 2 degree of freedom (df) joint test of main and interaction effects and a 1df interaction test within each population group. Significant BP loci were identified using genome-wide significance criteria (P_{2df} or $P_{1df} \leq$5e-9). Variants with marginal significance for the main genetic effect (P≤10e-5) were further followed up to identify significant interaction effects (Bonferroni $P_{1df} \leq 0.05$ accounting for the number of independent SNPs). Our preliminary results identify 940 previously reported BP loci and 10 novel loci, including and 8 novel loci in HIS (*HTR1F* and *SDK1* for SBP; *CTNNA3* and *CRBN* for DBP; *PAK5, KCNJ3, WWOX* and *PRMT6* for PP), 1 novel locus in AFR (*BRINP3* for PP), and 1 novel locus in EUR (*RPL39L* for PP) We expect to identify additional novel findings in the cross-population meta-analysis and will interpret findings through bioinformatics. While further work is ongoing, these preliminary results support the use of gene-by-sleep duration interactions to identify novel BP loci.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1497 Genetic analyses of inflammatory polyneuropathy and chronic inflammatory demyelinating polyradiculoneuropathy identified candidate genes.

Authors:

Z. Du, S. Lessard, T. Iyyanki, M. Chao, T. Hammond, D. Ofengeim, K. Klinger, E. de Rinaldis, S. Khader, C. Chatelain; Sanofi US Services Inc., Cambridge, MA

Objective: Chronic inflammatory demyelinating polyneuropathy (CIDP) is a rare, immune-mediated disorder in which an aberrant immune response causes demyelination and axonal damage of the peripheral nerves. Genetic contribution to CIDP is unclear and no genome-wide association study (GWAS) has been reported so far. In this study, we aimed to identify CIDP-related risk *loci*, genes, and pathways. Methods: To boost power, we first included all patients with a diagnosis of inflammatory polyneuropathy (IP) as cases. We performed a GWAS study using Finngen R10 individual data and combined the results with GWAS summary statistics from UK biobank (UKBB) using a fixed-effect meta-analysis. A total of 1,261 cases and 823,730 controls were included in the analysis. The second GWAS focused on CIDP patients and a total of 516 cases and 403,545 controls were included in the analysis. Stratified analyses by gender were also performed for both IP and CIDP. We performed gene-level analyses using transcriptome-wide mendelian randomization (TWMR) analysis, colocalization analysis, transcriptome-wide association study (TWAS) using S-PrediXcan and MAGMA to identify genes associated with IP and CIDP. Gene-set analyses were conducted using MAGMA to identify pathways that are related to IP and CIDP. Results: In GWAS study, we identified one genome-wide significant *loci* for CIDP risk among women. TWMR, colocalization and S-PrediXcan analyses identified four candidate pathogenic genes for IP; three candidate pathogenic genes for CIDP; three candidate pathogenic genes for IP among males; and two candidate pathogenic genes for IP among women. MAGMA gene-set analyses identified a total of 18 pathways related to IP or CIDP. Conclusion: Our study identified suggestive risk genes and pathways for IP and CIDP. Functional analysis will be conducted to further confirm these associations.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1498 Genetic and embryonic transcriptomic analyses revealed the molecular basis of Mayer-Rokitansky-Küster-Hauser syndrome.

Authors:

X. Cheng[1], S. Zhao[2], N. Chen[1], H. Zhao[1], Y. Zhang[3], Z. Zhao[1], G. Li[1], K. Xu[1], J. Cai[1], Q. Li[1], N. Wu[1], L. Zhu[1]; [1]Peking Union Med. Coll. Hosp., Beijing, China, [2]Baylor Coll. of Med., Houston, TX, [3]Wenzhou Med. Univ., Wenzhou, China

Background: Mayer-Rokitansky-Küster-Hauser syndrome (MRKHS), a congenital anomaly affecting the female reproductive system, is characterized by the aplasia of the uterus, cervix, and upper portion of the vagina. The etiology of MRKHS could be attributed to genetic perturbation during uterine development. Despite its clinical significance, a comprehensive understanding of MRKHS remains limited with the absence of single-cell transcriptomic analysis of the embryonic uterus, as well as the lack of systematic genetic association study.
Methods: Gene-level and pathway-level burden analyses were conducted based on exome sequencing/genome sequencing data from 727 probands with MRKHS and 2504 female control individuals. To gain more insight into the role of MRKHS-associated genes in the developing uterus, we performed single-cell transcriptome sequencing (scRNA-seq) on the metanephros of human and mouse embryos at different embryonic stages. Integration analyses of genetic and transcriptome data were employed to elucidate the intricate processes underlying uterine development.
Results: On the gene level, we identified five genes at an exome-wide significance (false discovery rate [FDR] < 0.05), including three previously reported genes (*PAX8*, *BMP7*, *GREB1L*) and two novel genes (*PAN2*, *AGPAT2*). In the human metanephros, scRNA-seq data indicated that the expression of MRKHS-associated genes ($P < 0.01$) was significantly enriched in the w8 uterine epithelium and the w11 Wolffian duct (WD) epithelium. Consistently, the expression in mouse metanephros was marginally enriched in the E14.5 WD and E18.5 Müllerian duct (MD) epithelium. Enrichment analysis and hierarchical clustering aided the prioritization of suboptimal genetic signals (e.g. *MT1F*). On the pathway level, two top hits were 'regulation of apoptotic process involved in development' (GO:1904748, P = 1.14E-06) and 'mesenchymal-to-epithelial transition involved in metanephros morphogenesis' (GO:0003337, P = 7.91E-06), consistent with their essential roles in the formation and morphogenesis of the female reproductive tract. Further analysis of cell type-specific driver genes revealed that dynamic driver genes of MD epithelium harbored the most significant mutational burden in patients. Notably, the digenic combination analysis identified three combinations reaching exome-wide significance, indicating potential digenic inheritance modes of MRKHS with co-expression evidence.
Conclusion: Our study uncovered the complex genetic factors associated with MRKHS and delineated their biological relevance in the context of the single-cell transcriptome of the embryonic uterus.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1499 Genetic and phenotypic association analyses of cardiometabolic traits in diverse African samples with whole-genome sequencing data.

Authors:

D. Hui[1], M. Hansen[1], D. Harris[1], M. McQuillan[1], D. Ju[1], A. Platt[1], W. Beggs[2], W. Sununguko Mpoloka[3], G. Mokone[3], G. Belay[4], T. Nyambo[5], S. Chanock[6], M. Yeager[7], TOPMed Consortium, G. Sirugo[1], M. Ritchie[8], S. Williams[9], S. Tishkoff[10]; [1]Univ. of Pennsylvania, Philadelphia, PA, [2]Univ Pennsylvania, Philadelphia, PA, [3]Univ. of Botswana, Gaborone, Botswana, [4]Addis Ababa Univ., Addis Ababa, Ethiopia, [5]Muhimbili Univ. of Hlth.and Allied Sci., Dar es Salaam, Tanzania, United Republic of, [6]Natl. Cancer Inst, Rockville, MD, [7]FNLCR/NCI, Rockville, MD, [8]Univ. OF PENNSYLVANIA, Philadelphia, PA, [9]Case Western Reserve Univ, Cleveland, OH, [10]Univ. of Pennsylvania Sch. of Med., Penn Valley, PA

African populations demonstrate exceptional genetic and phenotypic diversity, due in part to their varied environments, lifestyles, and demographic history. We conducted genetic and phenotypic association analyses in 6,965 geographically and ethnically diverse Sub-Saharan African individuals (6,280 with whole-genome sequences from the NIH TOPMed consortium and 685 with genotypes from Illumina arrays), using 15 cardiometabolic phenotypes (range 686-6,854 individuals/trait). Each phenotype had at least one ethnicity with significantly differing mean values compared to the remaining cohort, such as short stature in the Baka rainforest hunter-gatherers of Cameroon, and high adiposity in the Herero pastoralists of Botswana. An analysis of ethnicity-sex interactions revealed several ethnic groups with significant sexual dimorphism for at least one cardiometabolic phenotype, such as Herero women having markedly higher body mass index than men. Comparison between the African cohort and African ancestry UK Biobank (UKBB) individuals showed the latter have higher mean values than any of the 53 African ethnic groups for multiple cardiometabolic measurements, including low density lipoprotein cholesterol (LDL), body fat percentage (BFP), and systolic blood pressure. We also found that phenotype-phenotype correlations differ between the UKBB and African cohort, as well as between African ethnicities. For example, BFP and LDL had low correlation in the UKBB (r=0.04) but showed a range of correlation among African groups, from r = 0.00 in the Maasai pastoralists of eastern Africa to r = 0.43 in the Agaw agriculturalists of Ethiopia. Genome-wide association analyses identified 76 significantly associated loci ($p<5.0x10^{-8}$), with 14 passing a more stringent empirical threshold ($p<3.0x10^{-9}$), including *APOE* and *APOC1* loci for various blood lipids, *PCSK9* for LDL, and *CETP* for high density lipoprotein cholesterol (HDL), as well as novel loci. Set-based rare variant analyses for loss-of-function variants found 12 gene-phenotype associations replicating known associations with *PCSK9* and *APOE* for LDL and total cholesterol and uncovering several novel gene-trait associations for adiposity traits and HDL. Ongoing analyses include phenotype associations with subsistence and genetically inferred ancestry, replication of genetic associations, and gene-set enrichment. In total, these results offer insights into the genetic and phenotypic landscape of cardiometabolic traits in African populations. This work was supported by grant numbers: ADA 1-19-VSN-02, NIH grants 1R35GM134957, R01DK104339, and R01AR076241, and 1X01HL139409-01.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1500 Genetic architecture of occupational creativity and extensive genetic overlap with psychiatric disorders revealed from genome-wide association analyses of 241,736 individuals.

Authors:

H. Kim[1], Y. Ahn[1], J. Yoon[2], S. Kim[1], I. Shim[3], T. Park[4], H. Ko[5], S-H. JUNG[1], J. Kim[1], S. Park[1], D. Lee[1], B. Kim[1], M. Cho[1], H. Cho[1], D. Kim[1], H. Ihm[2], W-Y. Park[6], H. Bakhshi[7], K. O`Connell[8], O. Andreassen[9], K. Kendler[10], W. Myung[2], H-H. Won[1]; [1]Sungkyunkwan Univ., Seoul, Korea, Republic of, [2]Seoul Natl. Univ. Bundang Hosp., Seongnam, Korea, Republic of, [3]Sungkyunkwan Univ., Seongnam-si, Gyeonggi-do, Korea, Republic of, [4]Hallym Univ. Coll. of Med., Hwasung, Korea, Republic of, [5]Seoul Natl. Univ., Seoul, Korea, Republic of, [6]Sungkyunkwan Univ. Sch. of Med., Seoul, Korea, Republic of, [7]Nesta, London, United Kingdom, [8]Oslo Univ. Hosp., Oslo, Norway, [9]Oslo Univ. Hosp., Oslo, Oslo, Norway, [10]Virginia Commonwealth Univ, Richmond, VA

†These individuals contributed equally to this work as co-first authors.
*These individuals contributed equally to this work as co-corresponding authors.

Although creativity is heritable and often has a familial aggregation with psychiatric disorders, its genomic basis and genetic links with psychiatric disorders are mainly unexplored. By using a machine learning-based phenotyping of creativity based on occupation, we conducted a genome-wide association study (GWAS) of occupational creativity in individuals of European ancestry from the UK Biobank (n = 241,736) and identified 25 creativity-associated loci. The genetic association, polygenic risk score, and MiXeR analyses revealed extensive genetic overlap of occupational creativity with psychiatric disorders such as schizophrenia, major depression, bipolar I disorder, attention deficit/hyperactivity disorder, and anorexia nervosa. The condFDR and conjFDR investigations discovered additional loci for occupational creativity and psychiatric disorders, as well as shared genetic loci. Our GWAS showed similar results with GWASs conducted with traditionally creative occupations and GWASs adjusted for educational achievement. Our findings elucidate the genetic architecture of occupational creativity and shed light on its polygenic associations with psychiatric illnesses.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1501 Genetic association of ubiquitin specific peptidases (USPs) with the risk of Alzheimer's disease

Authors:

A. Cheema[1], K-H. Fan[1], M. Aslam[1], V. Acharya[1], E. Feingold[2], M. Kamboh[2]; [1]Dept. of Human Genetics, Sch. of Publ. Hlth., Univ. of Pittsburgh, Pittsburgh, PA, [2]Dept. of Human Genetics, Sch. of Publ. Hlth., Univ of Pittsburgh, Pittsburgh, PA

Taupathy is a major pathological hallmark of Alzheimer's disease (AD). Tau aggregation and clearance are controlled by multiple post translational events, including ubiquitination, which is the final modification step that controls proteasome-mediated clearance and degradation of tau. Ubiquitination is regulated by deubiquitinase (DUB) enzymes that remove ubiquitin groups. The human genome consists of ~100 DUBs, of which 50% are USPs, and 22 of them are expressed in the brain. We hypothesized that genetic variation in the brain USPs might alter the risk of AD by directly affecting the ubiquitin-linked clearance and degradation of tau. We examined this hypothesis in the International Genomics of Alzheimer's Project (IGAP) sample comprising ~72, 000 AD cases and controls. A total of 200 variants with nominal significance ($P<0.05$) were observed. After applying the M_{eff} method (effective number of independent tests) for multiple-testing correction, 62 variants remained significant in five genes, all in non-coding regions. The greatest number of significant associations were found in the USP8 gene (n=57), with both risk and protective variants. The most significant variant was protective (β= - 0.0726; P= 1.64E-05) followed by a risk variant (β= 0.0627; P= 4.80E-05). Based on RegulomeDB functional annotation, both variants had a RegulomeDB score of 1f, indicating that both have regulatory functions as they can affect transcription binding and expression of a gene target. In addition to SNPs, we also observed seven 2-5 bp insertions associated with both AD risk (n= 3) and protection (n= 4). The two top protective insertions were TACC (β= - 0.0697; P= 7.79E-05), and AAAAC (β= - 0.0695; P= 9.14E-05), and the top risk insertion was GTA (β= 0.0574; P= 5.81E-04). In conclusion, our study indicating that genetic variation in the brain USPs may affect the risk of AD, potentially highlights a tau-linked ubiquitination-deubiquitination pathway. Future genetic and associated functional studies may help to better understand this pathway.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1502 Genetic Associations with lupus in Sundanese population

Authors:

K. Patasova[1], S. Elbagir[2], N. Mohammed[3], M. A. Nur[4], E. M. Elagib[5], A. Elshafie[2,6], I. Gunnarsson[1], E. Svenungsson[1], J. Rönnelid[2], L. Diaz-Gallo[1]; [1]Karolinska Inst.t, Stockholm, Sweden, [2]Uppsala Univ., Uppsala, Sweden, [3]Al Neelain Univ., Khartoum, Sudan, [4]Alribat Univ. Hosp., Khartoum, Sudan, [5]Military Hosp., Omdurman, Sudan, [6]Linköping Univ. Hosp., Linköping, Sweden

Background: Published genome-wide studies have identified close to ~120 loci associated with systemic lupus erythematosus (SLE). However, current genetic findings are based mainly on Europeans. Investigations of SLE in East-Asian populations revealed distinct genetic associations with SLE manifestations across ancestries. We join the growing efforts to study the SLE genetic risk factors in other ancestry groups. **Aims:** This study aimed to evaluate genetic associations with SLE in the Sundanese population. To evaluate whether the HLA alleles associated in this cohort are similar to the reported ones in SLE. **Methods:** The study sample included 494 Sundanese participants; 92 were diagnosed with SLE, classified according to the 1982 revised (ACR) criteria. Genotyping was performed using Infinium® Expanded Multi-Ethnic Genotyping Array (MEGAEX). *HLA* alleles and amino acids were imputed against a modified 1000G African panel (5) in SNP2HLA. Genetic markers with MAF < 1%, genomic missingness >5% and imputation quality < 75% were removed; following quality control procedures, 453 unrelated samples, and 1,183.339 variants were advanced for statistical analyses. Genetic associations with SLE were evaluated in logistic regression model adjusted for age, sex, and the first two principal components, using PLINK2.0 (7). Associations with HLA alleles and amino-acids were assessed in separate analyzes and were considered significant if their probabilities were below Bonferroni correction threshold of 0.001 (p-value=0.05/44) and $5.55x10^{-05}$ (p-value=0.05/900), respectively. We applied conventional genome-wide significance threshold (p-value=$5x10^{-08}$) to genetic associations outside MHC region. **Results:** The strongest association within MHC region was detected with *HLA-DRB1*03* allele (frequency=12%) conferring an increased risk of SLE (OR=1.95, 95% CI=1.19,3.20; p-value=0.008). At higher allele resolution the *HLA-DRB1*0301* allele was associated with the two-fold increased risk of SLE (OR=2.00; 95%CI: 1.21, 3.33; p=0.007). The top amino acid associated were AK at position 71 of the *HLA-DRB1* gene (OR= 1.77; 95%CI: 1.18,2.66; p= 0,006). We also identified a suggestive novel association with rs12953472 (OR= 5.64, 95% CI= 2.86,11.09; p=$5.55x10^{-07}$) marker located within intronic region of *ZNF236-DT* gene, which encodes ferroptosis-related lincRNA. **Conclusions:** We have conducted the first GWAS assessing genetic determinants of SLE in the Sundanese population. We observed that HLA-DRB1*03 is also associated in SLE in this population and we report a novel suggestive signal within *ZNF236-DT* gene which is associated with a marked increase in the risk of SLE.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1503 Genetic Basis of Dilated Cardiomyopathy in Indian population

Authors:

S. Singh[1], A. Pandey[2], N. Garg[1], A. Srivastava[3]; [1]Sanjay Gandhi Postgraduate Inst. of Med. Sci. Lucknow, Lucknow, India, [2]Sanjay Gandhi post graduate Inst. of Med. Sci.s lucknow,, Lucknow, India, [3]Sanjay Gandhi Postgraduate Inst. of Med. Sci., Lucknow, India

Introduction- Dilated cardiomyopathy (DCM) is an autosomal dominant disease characterized by enlargement and dilation of one or both the ventricles along with impaired contractility.The genetic architecture of DCM continues to evolve rapidly due to its diverse and complex nature. In the present study, we sought to determine the genetic basis of DCM in Indian population. **Methods-**. We performed whole-exome sequencing (WES) on 50 individuals presented with the clinical features of dilated cardiomyopathy.Among them, 16 (32%) were familial. Sanger sequencing validated the pathogenic variants. Minigene assay was performed to functionally characterize the novel splice variant. **Results and discussion:** We identified pathogenic variants in 6 (12%) of the DCM-affected individuals. *ACTC1* c.309C>A, *DSP* c.478C>T, *PSEN2* c.250G>A and *TNNT2* c.547C>A. were identified as known whereas *ACTN2* c.616-2A and *FLNC* c.5897C>A were identified as novel variants. Minigene reporter assay confirmed that c.616-2A variant results in the skipping of exon 7 and expression of mutant protein.While DCM exhibits a Mendelian, monogenic architecture in the families cases, our data suggests that 4 (8%) of DCM-affected individuals have a digenic basis.Variant of unknown significance was identified in 33 individuals while remaining 17 (34%) were WES negative. **Conclusion:** Our study identified genes directly related to Z disc, sarcomere, ion channel, cytoskeleton and desmosome that highlights the complex genetic scenario of DCM. Furthermore, we show the digenic basis of DCM where two rare variants from different, unlinked loci determine the clinical features and the advantage of analysing population-specific sequences variants in the underrepresented populations.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1504 Genetic determinants of host IgG response to COVID-19 vaccination.

Authors:

S. Bian, X. Guo, X. Yang, Y. Wei, Z. Yang, X. Du, Y. Shu, S. Liu; Sch. of Publ. Hlth.(Shenzhen), Shenzhen Campus of Sun Yat-sen Univ., Shenzhen, China

Vaccination is an effective strategy for safeguarding vulnerable populations and reducing the severity of COVID-19 infections during the ongoing pandemic. However, there is substantial variation among individuals in the duration and strength of immune response following vaccination, and the underlying reasons for this variability remain unclear. Here, we explore the extent to which host genetic predisposition may contribute to the IgG responses following COVID-19 vaccination. We conducted a genome-wide association study (GWAS) on the qualitative IgG serostatus of individuals who had received either a single dose (N=54,066) or two doses (N=46,232) of the COVID-19 vaccine which use the spike protein of virus as antigen, and had no prior SARS-CoV-2 infection from UK Biobank. We found significant associations between common variants in the *LAT* gene (16q13) and HLA class II region with serostatus following a single-dose vaccination. Notably, we identified four independent alleles, including *HLA_DRB1*13:02*, *HLA_DQA1*01:01*, *HLA_DPB1*04:01*, and *HLA_DQB1*02:01*, which were consistently associated with serostatus after a single vaccine dose. Particularly, we observed an association between the amino acid residue at position 71 on DRβ1 and its electrostatic properties of the binding pocket and the serostatus among individuals who received only one vaccine dose. These findings underscore the role of host genetic factors in influencing the responses to COVID-19 vaccines and highlight the importance of considering the impact of the genetic background when formulating and implementing vaccination strategies to maximize protection within different populations.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1505 Genetic Determinants of Placental Growth Factor Levels and their Association with Coronary Artery Disease Risk

Authors:

J. Hartiala[1], J. R. Hilser[1], X. Huang[1], Y. Han[1], P. Huang[1], Q. Jia[1], S. L. Hazen[2], W. Tang[2], H. Allayee[1]; [1]USC Keck Sch. of Med., Los Angeles, CA, [2]Cleveland Clinic, Cleveland, OH

Introduction: Recent studies show that elevated placental growth factor (PlGF) concentrations predicted higher risk of cardiovascular disease (CVD) independent of known risk factors and were associated with adverse cardiovascular outcomes (death or future MI) in patients with acute coronary syndromes (ACS) or with chronic kidney disease (CKD). Although PlGF is an important cytokine involved in angiogenesis, cardiac remodeling and inflammation, it is not known whether the association between PlGF levels and coronary artery disease (CAD) outcomes is causal. To address this question, we used a large-scale genome-wide association study (GWAS) approach combined with Mendelian randomization (MR) and mediation MR analyses. **Methods:** A GWAS analysis was carried out in subjects from the GeneBank (n=3,455) and UK Biobank (n=52,096), followed by a weighted Z-score meta-analysis with four additional cohorts including AGES (n=5,368), SCALLOP (n=19,188), Fenland (n=10,708), and INTERVAL (n=3,301). To investigate the causal association of PlGF levels with cardiovascular risk, we performed MR analyses for CAD and MI with our lead SNPs that were not associated with cardiovascular-related traits as instrumental variables (IVs). Additionally, we performed mediation MR analyses using systolic blood pressure (SBP) as a potential mediator. **Results:** A meta-analysis for PlGF levels with 94,166 subjects and 9,723,360 overlapping SNPs identified 1,176 genetic variants distributed across 13 loci that exceeded the genome-wide significance threshold ($P<5.0x10^{-8}$). Of these, 12 loci were novel and not previously associated with PlGF. To investigate the causal relationship between PlGF levels and CAD, we used variants near the *FLT1*, *PGF*, and *BANF2* (rs9319427, rs91144 and rs17793602) for MR analyses with risk of CAD and MI. These analyses provided evidence that PlGF levels may be causally associated with decreased risk of CAD (OR=0.75, 95% CI 0.71-0.79, $P=2.5x10^{-23}$) and MI (OR=0.80, 95% CI, 0.71-0.91, $P=6.5x10^{-4}$), and that SBP was not mediating these associations. **Conclusions:** Our results suggest that higher levels of PlGF are associated with decreased risk of CAD and MI. This observation is surprising since higher PlGF levels have been clinically associated with increased cardiovascular risk. While increased levels of PlGF may precede progression and development of CAD, it is also plausible that PlGF levels may increase as a function of CAD as a protective response. Future studies will be needed to further explore the directionality of this relationship and to better understand the underlying biological mechanisms for this apparent paradoxical causal association.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1506 Genetic etiologies of sex hormones differ between female late premenopause and postmenopause stages.

Authors:

T. Waller[1], S. S. Faubion[2,3], J. M. Biernacka[1,4], S. J. Winham[1]; [1]Dept. of Quantitative Hlth.Sci., Mayo Clinic, Rochester, MN, [2]Div. of Gen. Internal Med., Mayo Clinic, Jacksonville, FL, [3]Ctr. for Women's Hlth., Mayo Clinic, Rochester, MN, [4]Dept. of Psychiatry and Psychology, Mayo Clinic, Rochester, MN

Steroid sex hormones regulate many aspects of physiology, and the concentrations of these hormones differ between males and females and change over the lifespan. In addition to variants on the sex chromosomes, broad differences in the autosomal genetic etiologies of sex hormones regulate their respective concentrations in both females and males. We hypothesized that genetic factors might also contribute to changes in sex hormone regulation over the lifespan, especially during the menopause transition in females. Using data from the UK Biobank, which included ~500,000 males and females over the age of 40, we compared genetic etiologies of hormone levels across stages of life. We stratified data by male age (tertiles) and by female menopause stage (late-pre-, peri-, post-) and ran genome-wide association studies (GWAS) on levels of testosterone (T), bioavailable T (BT), estradiol (E), bioavailable E (BE), and sex hormone binding globulin (SHBG) with adjustment for hormone-influencing factors such as the menstrual cycle, previous full-term pregnancies, and use of hormonal therapies. Using these GWAS summary statistics, we estimated the genetic correlation (rg) for each hormone level between males and females, age tertiles in males, and menopausal stage in females; rg<1 suggests differences in the contribution of genetic variants to hormonal phenotypes between groups. Consistent with previously published findings, we observed differences (i.e. rg<1) by sex for T (**rg: 0.09; se: 0.05**), BT (**rg: -0.05; se: 0.03**), E (**rg: 0.18; se: 0.11**), BE (**rg: 0.71; se: 0.08**), and SHBG (**rg: 0.88; se: 0.03**). Between pre- and postmenopause females, we observed differences for T (**rg: 0.65; se: 0.08**), BT (**rg: 0.74; se: 0.07**), and BE (**rg: -0.17; se: 0.21**). We also observed differences for BE (**rg: 0.38; se: 0.15**) between peri- and postmenopause females. We did not observe differences across the lifespan in males or for E and SHBG in females. These results confirm that contributions of common autosomal genetic variants to steroid sex hormones differ by sex, and they also suggest that genetic contributions vary based on menopause stage in females. Female reproductive stage is an important consideration in studies of genetic predisposition for levels of steroid sex hormones.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1507 † Genetic Factor Structure of the Posttraumatic Stress Disorder Checklist (PCL-17) in the Million Veteran Program

Authors:

C. Overstreet[1], D. Levey[1], R. Pietrzak[2], J. Deak[3], M. Galimberti[3], K. Adhikari[1], H. Zhou[4], K. Harrington[5], R. Quaden[5], J. Gaziano[6], M. Stein[7], J. Gelernter[8]; [1]Yale Univ., West Haven, CT, [2]VA Connecticut Hlth.care Ctr., West Haven, CT, [3]Yale Univ., New Haven, CT, [4]Yale Univ., Orange, CT, [5]Massachusetts Veterans Epidemiology Res. and Information Ctr. (MAVERIC), VA Boston Hlth.care System, Boston, MA, [6]VA Boston Hlth.care System, Roxbury Crossing, MA, [7]Univ. of California San Diego, San Diego, CA, [8]Yale Univ Sch. of Med., West Haven, CT

Posttraumatic stress disorder (PTSD) is a heterogenous disorder and summing items reflecting symptoms into a single sum score may obscure our ability to discern the biological underpinnings of the disorder. Attempts have been made to refine the trait definition by parsing items into phenotypic subdomains of re-experiencing, avoidance, and hyperarousal, per the DSM-IV; however, the genetic subdomains of PTSD may vary from phenotypic subdomains. No known study has performed item-level genome-wide association studies (GWAS) of the PTSD Checklist (PCL-17) or completed genetic modeling with GWAS results. We conducted (1) item-level GWAS of the PCL-17 in veterans of European (EUR) ancestry in the Million Veteran Program (MVP) and (2) genomic structural equation modeling (gSEM) to determine the best-fitting model of PTSD symptoms.
The sample consisted of ~230,000 EUR MVP participants. GWAS were conducted on the PCL-17 items and FUMA was used for mapping and annotation. Eigenvalues were examined to determine the number of factors to extract. To complete the modeling, gSEM was used. Exploratory factor analyses (EFA) were conducted using odd chromosomes and confirmatory factor analyses (CFA) were conducted using even chromosomes. The following conventions were used to evaluate model fit, with CFI>=0.90 and SRMR<=0.08 suggesting good-to-adequate fit. Genetic correlations across the 17 items ranged from 0.68 to 0.98. The number of independently significant SNPs ranged between 1 and 19. GWAS results across multiple items revealed SNPs associated with genes previously identified in other PTSD studies (e.g., *MAD1L1*). Examination of eigenvalues demonstrated two eigenvalues > 1 so an EFA was conducted using a two-factor model. The EFA results indicated that 11 items loaded onto factor one encompassing ‘threat’ symptoms, including 5 items assessing re-experiencing (e.g., emotional reactivity to trauma cues), 3 avoidance (e.g., avoiding reminders of the trauma), and 3 hyperarousal (e.g., hypervigilance) symptoms. Factor two items included symptoms encompassing ‘dysphoric’ symptoms, including 4 assessing avoidance (e.g., loss of interest) and 2 hyper-arousal (i.e., sleep disturbance) symptoms. A CFA was then conducted and results indicated adequate fit (χ^2=639.25, p=1.48e-72, CFI=0.92, SRMR=0.027).
Although all strongly genetically correlated, there was also some genetic variability across PCL-17 items. Modeling results demonstrated that a two-factor model best fit the data. Additional analyses will be conducted using the two factors derived from gSEM to further examine genetic variability and interrogate post-GWAS findings.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1508 Genetic heterogeneity in type 1 diabetes risk and age-at-onset using diverse ancestry populations in the Type 1 Diabetes Genetics Consortium (T1DGC).

Authors:

D. Michalek[1], C. Tern[1], W-M. Chen[1,2], W. Zhou[3,4,5], S. Onengut-Gumuscu[1,2], S. Rich[1,2], T1DGC; [1]Ctr. for Publ. Hlth.Genomics, Univ. of Virginia, Charlottesville, VA, [2]Dept. of Publ. Hlth.Sci., Univ. of Virginia, Charlottesville, VA, [3]Analytic and Translational Genetics Unit, Dept. of Med., Massachusetts Gen. Hosp., Boston, MA, [4]Stanley Ctr. for Psychiatric Res., Broad Inst. of MIT and Harvard, Cambridge, MA, [5]Program in Med. and Population Genetics, Broad Inst. of Harvard and MIT, Cambridge, MA

Background: Type 1 diabetes (T1D) is a complex autoimmune disease characterized by the destruction of the pancreatic β-cells, resulting in life-long requirement for insulin. The etiology of T1D includes both genetic and unknown environmental components. To date, GWAS meta-analyses and fine mapping studies reported over 100 loci associated with risk of T1D, with the majority of associated SNPs enriched in enhancers, although specific alleles in HLA class I and class II genes account for ~50% of T1D genetic susceptibility.
Objective: Use a multi-ancestry GWAS approach to discover novel genetic variants, HLA genes and amino acids, that are associated with T1D risk and age-at-onset.
Methods: The T1DGC data included 3,222 affected sib-pair/trio families (12,213 individuals, largely of European ancestry (EUR)), 891 unrelated African ancestry (AFR, 409 T1D cases, 482 controls) and 308 unrelated Admixed individuals (AMR, 153 T1D cases, 155 controls). All samples were genotyped using the Illumina CoreExome BeadChip array. Genotypes were imputed to the Trans-Omics for Precision Medicine (TOPMed) reference panel. HLA imputation used the multi-ancestry HLA reference panel (HLA-TAPAS). Imputation generated 13,965,323 variants (MAF > 0.01), including 20,425 HLA region variants (MAF > 0.005). T1D risk and HLA analysis was performed with logistic mixed models (SAIGE), while age-at-onset of T1D used survival models (GATE).
Results: In meta-analysis, seven T1D-associated loci attained genome-wide significance: 1p13.2 (*PTPN22*), 6p21.32 (*HLA-DQA1*), 10p15.1 (*IL2RA*), 10q23.31 (*RNLS*), 11p15.5 (*INS*), 12q13.2 (*IKZF4-RPS26-ERBB3*), and 12q24.12 (*SH2B3*), all known. Four of these loci were associated with age-at-onset of T1D (*PTPN22*, *HLA-DQB1*, *INS* and *ERBB3*). In meta-analysis of AFR and AMR, the *NRP1* locus (rs722988, OR = 1.61, P = 1.10×10^{-8}) was associated with both T1D risk and age-at-onset, *NRP1* was not significant in EUR; *PTPN22* was only significant in EUR. Fine mapping in the HLA region revealed *HLA-DRB1*03:01-DQA1*05:01-DQB1*02:01* as the most significant haplotype in AFR and AMR and *HLA-DRB1*04:01-DQA1*03:01-DQB1*03:02* in EUR. In conditional analysis, *HLA-DRB1*08:02-DQA1*04:01-DQB1*04:02* was protective in AMR population (OR = 0.39, P = 2.9×10^{-2}); in EUR, *HLA-DRB1*08:01-DQA1*04:01-DQB1*04:02* was associated with increased risk for T1D (OR = 1.81, P = 5.5×10^{-5}).
Conclusions: The inclusion of subjects with diverse genetic ancestry identified NRP1 with a stronger effect on T1D risk and age-at onset in non-EUR populations. The *NRP1* locus and specific HLA haplotypes, in addition to *PTPN22* distinguish T1D risk in EUR and non-EUR ancestries.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1509 Genetic Identification of New Chronic Pain Associations with the Central and Peripheral Nervous System.

Authors:

E. UZOCHUKWU[1], Z. Cader[2], C. Sandor[3], C. Webber[4]; [1]CARDIFF Univ., CARDIFF, United Kingdom, [2]Univ. of Oxford, Oxford, United Kingdom, [3]Univ Oxford, Oxford, United Kingdom, [4]Cardiff Univ., Cardiff, United Kingdom

Chronic pain, defined as pain persisting or reoccurring for 3 months or longer, is a leading cause of disability and represents a significant unmet medical need. Hence, there is an urgent need to identify novel drug targets for chronic pain, especially non-opioid analgesics, to support the development of new analgesic therapies and improved long-term pain outcomes. Current -omic studies of chronic pain often focus on clinical diagnoses of specific pain disorders. We hypothesize that there are genetic mechanisms that are shared across chronic pain disorders that are not specific to any single disease. To identify 'chronic pain', we utilized recurring analgesic prescription usage, defined as a minimum duration of 3-months of continuous use, (in lieu of specific pain diagnosis) to identify chronic pain phenotypes in the UK Biobank that we could then leverage in novel genome-wide association studies (GWAS). We generated n=13 phenotypes and conducted n=39 individual GWAS (both sexes combined and sex-stratified) to understand the genetic architecture of chronic pain. Subsequently, we identified cell specific pathways specific associated with these genetic pain traits by integrating the information of multiple PNS and CNS single-cell RNA sequencing atlases. We found that the cell/pathway varied according to type of analgesic, gender, and time on/off prescription. The identification of these distinct pathway is critical in the future development of successful therapeutic strategies for chronic pain.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1511 Genetic insights into platelet counts and Gestational Thrombocytopenia during pregnancy: A Genome-wide association study of 100,186 Chinese pregnancies.

Authors:

Z. Yang, Y. Gu, Y. Wei, Y. Liu, H. Zheng, X. Guo, S. Liu; Sch. of Publ. Hlth.(Shenzhen), Sun Yat-sen Univ., Shenzhen, China

Platelet counts decrease throughout pregnancy, and 5-12% of pregnant women are diagnosed with gestational thrombocytopenia (GT), characterized by a reduced platelet count during pregnancy. However, the biological mechanism of altered platelet counts and GT during pregnancy is still unclear. Here, we employed the human genetics approach and performed the first genome-wide association studies (GWAS) to investigate genetic factors for platelet counts at five stages of pregnancy (the first, second, and third trimesters, delivery, and the postpartum period), as well as two statuses of GT (GT: platelet count $< 150 \times 10^9$/L and severe GT: platelet count $< 100 \times 10^9$/L), using whole-genome sequencing data from non-invasive prenatal sequencing (NIPT) among 100,186 Chinese pregnancies. We identified 138 loci that showed genome-wide associations ($P < 5\times10^{-8}$), accounting for 3.2% of the phenotypic variation, and eight loci were newly discovered. Notably, we observed two loci in *PEAR1* and *CBL* with changing genetic effects in the five pregnancy periods. In the time-dependent GWAS across pregnancy, *PEAR1* was associated with platelet count variation during pregnancy.
Furthermore, *PEAR1* and *TUBB1* were found to be the most significant loci associated with GT and severe GT. We provided the first insights into the genetic basis of platelet counts and GT in pregnancy, and our findings indicate that *PEAR1* is the most crucial gene in the decrease of platelet counts and GT during pregnancy. We further propose the hypothesis that women carrying specific variants associated with a decrease in platelet count during pregnancy experience a more pronounced decline in platelet count during pregnancy, thereby contributing to the occurrence of GT. Our research provides insights for subsequent exploration of the biological mechanisms of GT.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1512 Genetic loci implicated in a meta-analysis of body shape in Africans

Authors:

M. Nakabuye, TACG Research Group; MRC/UVRI & LSHTM Uganda Res. Unit, Entebbe, Uganda

Background & Significance: Obesity is one of the leading causes of non-communicable diseases (NCD). Thus, NCD risk varies in obese individuals based on the location of their fat depots; while subcutaneous adiposity is protective, visceral adiposity increases NCD risk. Although, previously, anthropometric traits have been used to quantify body shape in low-income settings, there is no consensus on how it should be assessed. Hence, there is a growing interest to evaluate body shape derived from the principal component analysis (PCA) of anthropometric traits; however, this is yet to be explored in individuals of African ancestry whose body shape is different from those of Europeans. We set out to capture body shape in its multidimensional structure and examine the association between genetic variants and body shape in individuals of African ancestry. **Methods:** We performed a genome-wide association study (GWAS) for body shape derived from a PCA analysis of anthropometric traits in the Ugandan General Population Cohort (GPC, n = 6407) and the South African Zulu Cohort (SZC, n = 2595), which was followed by a GWAS meta-analysis to assess the genetic variants associated with body shape. **Results:** We identified variants in *FGF12*, *GRM8*, *TLX1NB* and *TRAP1* to be associated with body shape. These genes were different from the genes that have been associated with BMI, height, weight, WC and waist-hip ratio in continental Africans. Notably, we also observed that a standard deviation change in body shape was associated with an increase in blood pressure and blood lipids. **Conclusion:** Variants associated with body shape, as a composite variable might be different for those of individual anthropometric traits. Larger studies are required to further explore these phenomena.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1513 Genetic pleiotropy for kidney function and soluble receptor for advanced glycation end-products (sRAGE) using correlated meta-analysis in genome-wide association study and genome-wide transcriptome analysis: The Long Life Family Study (LLFS).

Authors:

M. Feitosa[1], B. Thyagarajan[2], S. J. Lin[1], S. Acharya[3], M. Wojczynski[1], A. Kuipers[4], A. Kulminski[5], G. Patti[6], K. Christensen[7], J. Zmuda[4], M. Brent[3], M. Province[1]; [1]Div. of Statistical Genomics, Dept. of Genetics, Washington Univ. Sch. of Med., St. Louis, MO, [2]Dept. of Lab. Med. and Pathology, Sch. of Med., Univ. of Minnesota, Minneapolis, MN, [3]Dept. of Computer Sci. and Engineering, Washington Univ., St. Louis, MO, [4]Dept. of Epidemiology, Graduate Sch. of Publ. Hlth., Univ. of Pittsburgh, Pittsburgh, PA, [5]Biodemography of Aging Res. Unit, Social Sci. Res. Inst., Duke Univ., Durham, NC, [6]Dept. of Chemistry, Washington Univ. in St. Louis, St. Louis, MO, [7]Unit of Epidemiology, Biostatistics and Biodemography, Dept. of Publ. Hlth., Southern Denmark Univ., Odense, Denmark

The estimated glomerular filtration rate (eGFR) is used for detecting and managing kidney disease. Patients with chronic kidney disease (CKD) present increased oxidative stress and chronic inflammation, which may escalate the production of advanced glycation end-products (AGE). Low eGFR measured by creatinine (eGFRcr) and cystatin C (eGFRcys) and high circulating soluble receptor for AGE (sRAGE) are associated with aging and CKD. In the LLFS, eGFRcr and eGFRcys correlated negatively with sRAGE (r= -0.25, p= $1.4x10^{-37}$ and r= -0.30, p= $2.7x10^{-53}$, respectively). To identify whether pleiotropic genetic variants share effects on kidney function and sRAGE, we conducted GWAS for eGFRcr, eGFRcys, and sRAGE, adjusting for age, sex, relatedness, and principal components, followed by correlated meta-analysis (CMA) on GWAS p-values from whole-genome sequence variants in 4182 subjects. Of the 59 loci identified (p< $5x10^{-8}$), 42 were novel discoveries, and 17 were previously reported for eGFR. Several novel locus variants (e.g., *CDKN2C*, *ZNF326 / SNORD3G*, *STK32B*, *LINC01987*) were reported with CKD, urea nitrogen, and creatinine levels but not with eGFR. Four locus variants have minor alleles with eGFRcys protective effects. For example, the *CASC17 / ROCR* rs528757227-A frequency is ~4.2 times higher in LLFS (A= 0.0131, β= 0.105 ± 0.029) than in ALFA-Europeans (A=0.003), suggesting better kidney function in the health-aging LLFS members than the general population. In addition, several variants (e.g., in *SH3GLB1 / SELENOF* and *CENPP*) are eQTLs in the glomeruli and tubule kidney (Human Kidney eQTL Atlas), and different tissues (GTEx), and promoter and enhancer histone marks, DNAse hypersensitive sites, and protein regulatory binding sites (HaploReg). To identify signatures of kidney function with expression protein-coding genes and lincRNAs, we performed TWAS for eGFRcr, eGFRcys, and sRAGE on blood RNA-seq and then CMA in a 1209 subsample. We identified 4 genes from TWAS eGFR and 13 genes from CMA between eGFR and sRAGE (Bonferroni p< $2.73x10^{-6}$). Among the 17 genes, the Kidney eQTL Atlas reports eQTLs for *DAAM2*, *TOPORS*, and *LSP* from meta-analyses. From kidney samples, the Atlas indicates that *TEC*, *CCN1*, *ERGIC1*, *SNTB1*, *LSP1*, and *TOPORS* harbor eQTLs with expression levels in the glomeruli or tubule kidney. Our findings demonstrate that the CMA enhances the association significance by combining GWAS or TWAS, allowing the discovery of novel variants and genes and replicating others for eGFR traits. Adding, genomic, transcriptomic, and bioinformatics results demonstrate compelling evidence that some novel variants and genes have functional regulatory features in the kidney.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1514 Genetic predictors of Venous Thromboembolism in Europeans

Authors:

F. Thibord, INVENT consortium, Million Veteran Program; NIH/NHLBI, Jamaica Plain, MA

Background: Venous thromboembolism (VTE) is a cardiovascular event resulting from an imbalance of hemostasis toward coagulation, and presents clinically as deep vein thrombosis or pulmonary embolism. In order to better characterize VTE etiology at the molecular level, two independent large-scale efforts recently performed VTE genome-wide association studies (GWAS) meta-analyses: Ghouse *et al.* (N=81,190 VTE cases of European ancestry) and Thibord *et al.* (N=81,669 VTE cases, including 71,771 Europeans).
Aims: To validate novel findings from both studies by cross-replication, and to perform an expanded meta-analysis to discover additional novel gene loci influencing VTE risk.
Methods: The Ghouse *et al.* meta-analysis involved only participants of European ancestry, thus only European ancestry individuals from the Thibord *et al.* study were included in this work. For the cross-replication analysis, novel loci that were significantly associated with VTE risk in only one study were investigated in the other study, after removal of cohorts shared between the two meta-analyses. Associations from a one-tailed test reaching a Bonferroni-corrected P-value threshold were deemed successfully replicated. We then performed a meta-analysis of all cohorts from the two studies which included a total of (71,771 + 37,250 =) 109,021 cases and over a million controls. Variants that exceeded the significance threshold ($P < 5 \times 10^{-8}$) were called significant.
Results: Out of 23 significant novel loci in the Ghouse *et al.* study, signals at the *F2RL3* and *PTPRR* loci were successfully replicated ($P < 0.0043$) in the Thibord *et al.* study after removal of overlapping cohorts (N=25,258 cases). Reciprocally, out of 28 novel loci from the Thibord *et al.* study, associations at the *AK5* and *RGS18* loci were replicated ($P < 0.0036$) in the Ghouse *et al.* meta-analysis, shared cohorts excluded (N=37,250 cases). Finally, the meta-analysis of the two studies revealed significant associations at 129 genetic loci, of which 21 have not been reported by previous efforts, including notably a missense variant (rs61749613, p.K349E) in *VCAN*, a gene encoding an extracellular matrix protein with suggested roles in wound healing and inflammation.
Conclusion: With this work, we added four loci to the list of 68 VTE loci replicated in independent studies that were identified by previous efforts. These four loci were part of 129 signals significantly associated with VTE risk in our meta-analysis, which also included 21 novel genetic variants influencing VTE risk. These results will offer novel insights in the mechanisms involved in VTE progression as well as potential opportunities for new therapeutic approaches.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1515 Genetic Predisposition of DRD4 and Associated Metabolic Enzymes with Psychiatric Felonies in Pakistani Prison Inmates

Authors:

R. Qadeer[1], S. Hasnain[1], J. J. Mann[2]; [1]Inst. of microbiology and Molecular Genetics., Lahore, Pakistan, [2]Molecular Imaging and Neuropathology Div., Columbia Univ./New York State Psychiatric Inst., New York, NY

The postsynaptic action of monoamine dopamine is known to be influenced by DRD4 and its metabolising enzyme genes. These are the most studied genes in connection with the pathogenesis of aggression and criminal delinquency. A total of 729 subjects were recruited, comprising 370 violent prison inmates convicted of first-degree murder(s) from different prisons in Punjab, Pakistan, and 359 control men without any history of aggression or criminal tendency from the general population. All the subjects were genotyped for six genetic polymorphisms in genes of dopaminergic 48 bp VNTR polymorphism of the *DRD4* gene, genetic variants of TPH1 (rs1800532), *TPH2* (rs7305115), neurotrophic (*BDNF* rs6265), and neurotransmitter metabolism (*COMT* rs4680, and *MAOA* VNTR polymorphisms) systems. The overall study results demonstrate a significant association of dopaminergic DRD4 (OR = 3.62, p = <0.0005 and OR = 11.02, p = <0.0005, respectively), and no association was evident for TPH genetic variants or genetic variants of neurotransmitter metabolism system genes (BDNF, COMT, and MAOA) genetic variants with aggression and criminal behaviour. In addition, some of these candidate genetic variants were also significantly associated with verbal abuse (*BDNF* polymorphisms) and physical abuse (*COMT* and *BDNF* polymorphisms) Also, the carriers of risk genotypes /alleles of *DRD4* polymorphisms had significantly higher mean aggression scores on the STAXI subscale as compared to wild-type genotypes/alleles in murderers. Taken together, these results suggest a major role for dopaminergic (*DRD4*) polymorphism and a modest contribution of neurotransmitter metabolism (*COMT*) and neurotrophic (*BDNF*) genetic variants in determining susceptibility to self-reported aggression and criminal behaviour in this sample subset of violent murderers. Using the Braineac and GTEx databases, significant eQTL-based functional effects for 48 bp VNTR in DRD4 and other neurotransmitter metabolism polymorphisms were analysed in different brain regions and peripheral tissues. In conclusion, these findings implicate the 48-bp VNTR of DRD4 as a major genetic determinant associated with criminal aggression. Future studies are needed to replicate these findings in other populations of murderers and extend them to other forms of violence and lesser degrees of aggression.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1516 Genetic profiling and polygenic risk score performance evaluation in the multi-ethnic population of the Kahn-Sagol-Maccabi Biobank.

Authors:

O. Sabik[1], W. Kim[1], A. Madar[1], B. Peterson[1], M. Cohn Adulamy[1], W. Gordon[1], A. Peretz[2], T. Patalon[2]; [1]Valo Hlth., Boston, MA, [2]Kahn Sagol Maccabi (KSM) Res. & Innovation Ctr., Maccabi Hlth.care Services, Tel Aviv, Israel

Coupling rich electronic health record data with the ability to sequence patients of interest is key to the identification of molecular signatures of specific populations and their application to drug discovery. The collaboration between Valo Health and Kahn-Sagol-Maccabi (KSM), the Research and Innovation Center of Maccabi Healthcare Services, Israel's leading HMO, unifies Valo's Opal Computational Platform with KSM's Biobank, one of the most comprehensive and rich biomedical datasets globally. Containing more densely measured medical history, driven by Israel's adoption of universal socialized healthcare, linked pharmacy claims, and more frequent measures of a broader panel of laboratory tests than comparable biobanks, such as the UK Biobank or FinnGen, KSM's Biobank provides an unparalleled opportunity to identify patient cohorts using deep longitudinal data and leverage Israel's unique population history to drive genetic discovery. Beginning with tens of thousands of sequenced patients, this collaboration is uniquely positioned to combine clinical and genetic data for drug discovery. As an example, computing genome-wide polygenic risk scores (PRS) for diseases of interest and mediating endophenotypes in conjunction with rich longitudinal clinical data can help in the identification of high-risk populations (Khera et al, Nat Gen, 2018) and may indicate new drug target hypotheses. In the pilot study, population ancestry was evaluated, identifying predominantly individuals of Ashkenazi ancestry (~60%), but also representation from Eastern Mediterranean, Middle Eastern, and Eastern European ancestry, including many admixed individuals. This ancestry structure poses some challenges; most human genetics tools and data are biased towards western European ancestry and suffer reduced performance in admixed individuals (Ding et al, Nature, 2023). To calibrate the extent of performance degradation and identify effective analysis approaches in non-European populations, PRS performance was evaluated for the prediction of gross phenotypes and endophenotypes, as was the portability of predominantly European-derived PRS weights to this diverse and admixed population. Future work will include tuning PRS using transfer learning (Zhao et al, AJHG, 2022) for more effective application of existing PRS weights to the cohort. Additionally, the bottlenecked nature of the Israeli population is uniquely positioned to have increased rare variant discovery power due to higher local frequencies of globally rare genetic variants. Measures of identify by descent will be used to identify cohorts for rare genetic discovery.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1517 † Genetic risk and changes in metabolomic risk over repeated measures.

Authors:

S. Kerminen, Nightingale Health Biobank Collaborative Group; Nightingale Hlth.Plc, Helsinki, Finland

Polygenic scores (PGS) are shown to improve risk prediction of complex diseases and could be used as a tool to motivate interventions and lifestyle changes. As genetic risk stays the same throughout life, it will also be important to complement genetic risk prediction with measures responsive to changes in lifestyle to encourage positive behavior. Metabolomic scores, measured from blood and analogous to PGS, have recently been shown to provide impressive risk prediction for many diseases. However, there is yet little understanding on how changes in metabolomic risk are associated with genetic and overall future risk. Here, we measure blood biomarkers using nuclear magnetic resonance for 255,000 individuals from UK Biobank. For 15,000 participants, we also measured a second blood sample 5 years after the baseline visit. Using half of the baseline samples without repeat sample and Cox proportional hazards model with Lasso and 5-fold cross validation, we developed metabolic risk scores for type 2 diabetes (T2D), myocardial infarction (MI) and chronic obstructive pulmonary disease (COPD). We calculated genetic risk for T2D, MI and COPD using previously published PGS. First, by fitting a joint risk model with baseline and follow-up metabolomic scores, we show that both time points were significantly associated with 10-year risk suggesting that both person's current metabolomic risk, and how long they have lived in a state of elevated risk, contribute information about future risk. Furthermore, we found that leaving the highest risk decile showed a dramatic reduction in future risk of diabetes (HR 3.1, p=3e-6) and COPD (HR 2.4, p=0.002), and consistent but non-significant difference for MI (HR 1.5, p=0.2). We replicated the results in 5000 Estonian Biobank participants with two metabolomic measurements for T2D (HR 4.5, p=0.002) which was the only disease with enough cases to test. Second, we tested associations of change in metabolomic score with PGS. For T2D, we observed that individuals with PGS at the highest decile were 1.3 times more likely (p=0.004) to switch from average metabolomic risk group into the highest decile 5 years later than other participants who started with average metabolomic risk. Analogously for COPD, we observed that among individuals who started with high metabolomic risk, individuals with high PGS were 0.51 times less likely (p=0.003) to reduce their metabolomic risk than the rest. Our results demonstrate that metabolomic risk scores dynamically capture risk over time and can track changes in an individual's risk profile. We also show it is more difficult to maintain good metabolomic risk profiles over time for individuals with high genetic risk.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1518 Genetic Risk for Body Mass and Mortality in Chronic Obstructive Pulmonary Disease

Authors:

J. Zhang[1], M. Moll[2], B. Hobbs[2], P. Bakke[3], E. Regan[4], H. Xu[5], J. Dupuis[6], E. Silverman[2], G. O'Connor[1], M. Cho[2]; [1]Boston Univ. Sch. of Med., Boston, MA, [2]Brigham and Women's Hosp., Boston, MA, [3]Univ. of Bergen, Bergen, Norway, [4]Natl. Jewish Hlth., Denver, CO, [5]Boston Univ. Sch. of Publ. Hlth., Boston, MA, [6]McGill Univ., Montreal, QC, Canada

Background: Low body mass index (BMI) is associated with increased COPD mortality, but the underlying mechanism is unclear. BMI is partially determined by genetics. We hypothesized that BMI genetic risk is associated with COPD mortality and has differential effects on respiratory and cardiovascular death in COPD. Methods: To obtain a summative measure of genetic risk for BMI, we developed a polygenic risk score (PRS_{BMI}, with higher scores associated with higher BMI) using lassosum based on the summary statistics from the largest BMI GWAS to date (~700,000 general population participants of European ancestry). We calculated the PRS_{BMI} for COPDGene non-Hispanic white (NHW) and African American (AA), ECLIPSE, and Framingham Heart Study (FHS) participants who had moderate-to-severe COPD (FEV_1/FVC <0.7 and FEV_1 % predicted <80%). We used penalized splines to test for non-linear effects of the PRS_{BMI} on mortality. We used Cox proportional hazards models to examine the association between the PRS_{BMI} and all-cause mortality. For respiratory and cardiovascular mortality, we used cause-specific hazard models to account for competing risks. Models were adjusted for age, sex, current smoking, pack-years of smoking, genetic ancestry, and FEV_1 % predicted. We conducted meta-analysis to combine effect estimates among cohorts. Results: We included 2811 COPDGene NHW, 753 COPDGene AA, 1708 ECLIPSE, and 751 FHS participants with moderate-to-severe COPD. The PRS_{BMI} contained approximately 439,000 variants, and accounted for 9.9% of the BMI in COPDGene NHW, 1.0% in COPDGene AA, 7.5% in ECLIPSE, and 18.3% in FHS participants. We did not observe a significant non-linear effect of the PRS_{BMI} on all-cause, respiratory, or cardiovascular mortality. In meta-analysis, a one standard deviation increase in the PRS_{BMI} was associated with an increased hazard for cardiovascular mortality (HR=1.30, 95% CI=1.12-1.51), but not with respiratory mortality (HR=1.01, p=0.80) or all-cause mortality (HR=1.04, p=0.052). We did not observe significant heterogeneity in effects of the PRS_{BMI} among cohorts (I^2=0%, p for the Q statistic >0.60 for all outcomes). Conclusions: In individuals with COPD from smoker-enriched cohorts and a population-based cohort, higher BMI genetic risk is associated with higher cardiovascular mortality, but not with respiratory mortality.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1519 Genetic risk stratification for primary open-angle glaucoma in Japanese individuals

Authors:

M. Akiyama[1], G. Tamiya[2], K. Fujiwara[1], K. Sato[2], Y. Kamatani[3], T. Ninomiya[1], K-H. Sonoda[1], T. Nakawaza[2]; [1]Kyushu Univ., Fukuoka, Japan, [2]Tohoku Univ., Sendai, Japan, [3]The Univ. of Tokyo, Minato-ku, Japan

Primary open-angle glaucoma (POAG) is a complex disease. Although the recent studies suggested that genetic risk scores (GRSs) could estimate the susceptibility of POAG, the most studies had been conducted in non-East Asian populations. To assess the utility of genetic risk stratification for POAG in Japanese individuals, we generated GRSs based on the summary statistics of the genome-wide association study (GWAS) of Japanese POAG (3,979 cases vs 30,278 controls). Three types of GRSs were constructed: 1) polygenic risk scores (PRSs) using pruning and thresholding (P+T) procedure (N = 24), 2) PRSs constructed by LDpred software (N = 4), and 3) GRS using variants significantly associated with POAG in the GWAS of International Glaucoma Genetics Consortium (IGGC; Gharahkhani et al. *Nat Commun* 2021). A total of 3,535 Japanese individuals including 1,191 POAG cases and 2,344 controls were analyzed to select the model with highest discriminative accuracy based on area under the receiver operative characteristics curves (AUCs). Among the 29 constructed models, the GRS based on 98 variants which were reported by IGGC showed the best discriminative accuracy (AUC = 0.65). In this case-control dataset, the proportion of POAG in individuals at the top decile of GRS was significantly higher than that at bottom decile (54.4% vs16.2%, odds ratio = 6.15, 95% confidence interval = 4.35-8.71). Associations of the selected GRS with prevalence of POAG was investigated in a dataset of general Japanese population (Hisayama study; N = 1,900). In the Hisayama study, we confirmed that the GRS was significantly associated with the prevalence of POAG (*P* for trend = 0.01). In conclusion, GRS showed moderate discriminative accuracy for POAG in the Japanese population which was in line with those in the European and African populations.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1520 Genetic support for interleukin-1 inhibition in the treatment of pericarditis.

Authors:

R. B. Thorolfsdottir[1], A. B. Jonsdottir[1], G. Sveinbjornsson[1], H. M. Aegisdottir[1], A. Oddsson[1], O. A. Stefansson[1], G. H. Halldorsson[1], S. Saevarsdottir[1,2], L. Stefansdottir[1], G. Thorleifsson[1], O. B. Pedersen[3,4], C. Erikstrup[5,6], D. O. Arnar[1,2,7], DBDS Genomic Consortium, K. U. Knowlton[8,9], L. D. Nadauld[10,11], S. Gretarsdottir[1], A. Helgadottir[1], S. R. Ostrowski[4,12], D. F. Gudbjartsson[1,13], I. Jonsdottir[1,14], H. Bundgaard[4,15], H. Holm[1], P. Sulem[1], K. Stefansson[1,2]; [1]deCODE genetics/Amgen, Inc., Reykjavik, Iceland, [2]Faculty of Med., Sch. of Hlth.Sci., Univ. of Iceland, Reykjavik, Iceland, [3]Dept. of Clinical Immunology, Zealand Univ. Hosp., Køge, Denmark, [4]Dept. of Clinical Med., Faculty of Hlth.and Med. Sci., Univ. of Copenhagen, Copenhagen, Denmark, [5]Dept. of Clinical Immunology, Aarhus Univ. Hosp., Aarhus, Denmark, [6]Dept. of Clinical Med., Hlth., Aarhus Univ., Aarhus, Denmark, [7]Dept. of Med., Landspitali-The Natl. Univ. Hosp. of Iceland, Reykjavik, Iceland, [8]Intermountain Med. Ctr., Intermountain Heart Inst., Salt Lake City, UT, [9]Sch. of Med., Univ. of Utah, Salt Lake City, UT, [10]Precision Genomics, Intermountain Hlth.care, Salt Lake City, UT, [11]Sch. of Med., Stanford Univ., Stanford, CA, [12]Dept. of Clinical Immunology, Copenhagen Univ. Hosp., RigsHosp.et, Copenhagen, Denmark, [13]Sch. of Engineering and Natural Sci., Univ. of Iceland, Reykjavik, Iceland, [14]Dept. of Immunology, Landspitali, The Natl. Univ. Hosp. of Iceland, Reykjavik, Iceland, [15]Dept. of Cardiology, The Heart Ctr., Copenhagen Univ. Hosp., RigsHosp.et, Copenhagen, Denmark

Drugs inhibiting interleukin-1 cytokines are a promising new treatment option for recurrent pericarditis, but their use is based on scarce biological evidence and clinical trials of modest sizes. Our aim was to use human genetics and multiomics to study the pathogenesis of pericarditis. We performed a GWAS meta-analysis of 4,894 pericarditis cases and 1,457,822 controls from five countries; Iceland (deCODE genetics), Denmark (Copenhagen Hospital Biobank Cardiovascular Study/Danish Blood Donor Study, CHB/DBDS), the UK (UK Biobank), the USA (Intermountain), and Finland (FinnGen). We identified two independent sequence variants at the interleukin-1 locus on chromosome 2q14 that associate with pericarditis; rs12992780[T] (effect allele frequency, EAF = 31-40%, OR = 0.83, $P = 6.7\times10^{-16}$) and rs7575402[A or T] (EAF = 45-55%, adjusted OR = 0.89, adjusted $P = 9.6\times10^{-8}$). Rs12992780 has a stronger effect on recurrent pericarditis than the acute form (OR = 0.76 vs 0.86, *P*-het = 0.030). Available transcriptomics and proteomics datasets did not reveal plausible causative genes for the associations. Rs7575402 associates with CpG methylation overlapping binding sites of four transcription factors known to regulate interleukin-1 production, PU.1 (encoded by *SPI1*), STAT1, STAT3, and CCAAT/enhancer-binding protein beta (encoded by *CEBPB*). The genetic associations substantiate the involvement of interleukin-1 in the pathogenesis of pericarditis and have the potential to solidify the use of interleukin-1 inhibitors in the treatment of the disease.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1521 Genetic variants associated with premature peripheral artery disease and surgical revascularization.

Authors:

j. hu[1], D. Alameddine[2], C. I. Ochoa Chaar[2], A. Dewan[1]; [1]Yale Sch. of Publ. Hlth., New Haven, CT, [2]Yale Sch. of Med., New Haven, CT

Peripheral artery disease (PAD) affects over 200 million people worldwide and impacts mostly the elderly. Even though advanced stages of the disease can lead to major lower extremity amputations, most patients are asymptomatic. Identification of patients at risk of progression is crucial to develop strategies for early treatment and risk factor modification. Multiple lines of evidence have suggested that there is a significant genetic component contributing to the development of PAD, and a recent genome-wide association study (GWAS) has identified and replicated 19 independent single nucleotide polymorphisms (SNPs) associated with the diagnosis of PAD, but the relationship between these variants and PAD severity has not been established. This study explored the association between these 19 variants and PAD severity using 6,815 PAD patients and 401,872 controls of White European ancestry from the UK Biobank. We have demonstrated in prior work that the common age for patients to present for surgical lower extremity revascularization was 60-80, and that patients with premature PAD were at increased risk of major amputation. Thus, early age of onset (<55) and surgical revascularization for PAD are indicators of increased severity. We classified PAD cases into 4 groups with increasing levels of severity: non-premature (>=55), non-surgical PAD (N=3,589); non-premature, surgical PAD (N=2,115); premature (<55), non-surgical PAD (N=710); and premature, surgical PAD (N=401). Genetic association analysis was conducted using REGENIE, to compare each severity group with the common control subjects. Each analysis was adjusted for age, sex and 10 principal components. Association results were then extracted for the 19 variants. A linear test for trend was conducted to test whether the effect of each SNP increased or decreased with severity. Three variants (rs6025, rs11066301, and rs10851907) showed increasing odds ratios (ORs) with increasing severity and each had a significant trend test ($p<0.0026$ [0.05/19]). Among them, rs6025 had the clearest trend: OR=1.13 (95% CI: 0.89, 1.32) for non-premature, non-surgical PAD, 1.22 (95% CI: 0.83, 1.48) for non-premature, surgical PAD, 1.36 (95% CI: 0.74, 1.93) for premature, non-surgical PAD, and 1.45 (95% CI: 0.7, 2.33) for premature, surgical PAD. Ten of the 19 SNPs showed larger ORs for surgical compared to non-surgical PAD in both premature and non-premature PAD suggesting that these SNPs are associated with surgical intervention regardless of age of onset. Our findings demonstrate that SNPs associated with PAD development may also be associated with PAD severity as assessed by surgical intervention and prematurity.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1522 Genetic variants associated with type 2 diabetes mellitus among Filipinos

Authors:

E. Cutiongco De La Paz[1,2], J. Nevado[1], E. Paz-Pacheco[3], G. Jasul[3], A. Aman[1,4], D. Francisco[3]; [1]Inst. of Human Genetics, NIH, Univ. of the Philippines Manila, Manila, Philippines, [2]Philippine Genome Ctr., Univ. of the Philippines, Quezon City, Philippines, [3]Section of Endocrinology, Dept. of Med., Philippine Gen. Hosp., Univ. of the Philippines Manila, Manila, Philippines, [4]Philippine Council for Hlth.Res. and Dev., Dept. of Sci. and Technology, Taguig City, Philippines

Background: Type 2 diabetes mellitus (T2DM) affects a significant number of Filipinos and leads to debilitating complications and poor quality of life. Genetic variability contributes to as much as 30% to70% of T2DM risk. The determination of genomic variants related to susceptibility to T2DM can help in early identification of at-risk individuals. However, it has been recognized that interethnic variability in T2DM genetic susceptibility exists. This study aimed to identify the variants that may be associated with type 2 diabetes mellitus in the Filipino population.**Methodology:** Using a case-control design frequency matched for age and sex, and compared 66 adult, unrelated Filipinos with T2DM with 188 without T2DM. Genotyping was done using a candidate gene approach derived from various literature and database sources in relation to T2DM and related conditions. Analyses involved allelic association and genotypic association studies with correction for multiple testing.**Results:** The study found eight (8) significant variants associated with T2DM in Filipinos (implicated gene): rs7119 (*HMG20A*), rs7766070 (*CDKAL1*), rs708272 (*CETP*), rs12150053 (*SERPINF1*), rs2383208 (*CDKN2B-AS1*), rs391300 (*SRR*), rs659366 (*UCP2*), and rs10497721 (*TMEFF2*). In particular, the highest ORs were rs7119, rs7766070, and rs708272 at 31.06 (95% CI: 4.18, 230.52), 20.03 (95% CI: 6.15, 65.26), and 12.80 (95% CI: 3.87, 42.34), respectively. Notably, both rs7119 and rs708272 had high risk allele frequency at 0.77 and 0.66, making them potentially good markers for screening for T2DM. As expected, most variants are from genes related to glucose and/or energy metabolism.**Conclusion:** The study found eight significant SNPs associated with T2DM susceptibility, with a variant of HMG20A (rs7119) conferring the highest level of risk. The data generated from this study can be valuable towards the development of genetic prediction models for type 2 diabetes relevant to the prevention and diagnosis of T2DM among Filipinos.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1523 Genetic Variants Impacting Brain Structure Changes in Cognitive Decline: A Genome-wide Association Study.

Authors:

Y. Jin[1], A. Scott[1], C. Rochet[1], P. Drineas[2], P. Paschou[1]; [1]Purdue Univ., West Lafayette, IN, [2]Purdue Univeristy, West Lafayette, IN

Individuals with cognitive decline exhibit distinct brain structure changes, highlighting the importance of understanding the genetic factors that may influence these alterations. We performed the first genome-wide association study aiming to identify common genetic variants that are associated with rates of brain change in the context of cognitive decline. We analyzed longitudinal magnetic resonance imaging (MRI) data from three cohorts, including the ADNI, AddNeuroMed, and NACC (1627 individuals). We initially assessed correlations between the rate of change in the Mini-Mental State Examination and changes in different brain structures. Subsequently, we performed a genome-wide association study (GWAS) and discovered common genetic variants that had a significant impact on the rates of brain change associated with cognitive decline. The identified variants were found to influence the rates of change in 16 distinct brain regions, including the bilateral hippocampus, lateral ventricles, temporal lobe, frontal lobe, occipital lobe, parietal lobe, left inferior lateral ventricles, left putamen, right cerebellum cortex and white matter hyperintensities. Notably, our study revealed a noteworthy association between the IRS1 gene and changes in the left temporal lobe in our gene-based GWAS. Previous research has linked IRS1 to anti-amyloid effects in mouse models of Alzheimer's disease. We also investigated the pleiotropy and concordance of genetic risk factors associated with changes in brain structure in the context of cognitive decline, examining the influence of various factors such as smoking, drinking, obesity, diabetes, high cholesterol, and head injuries on alterations in brain structure. Our findings revealed that all these risk factors could affect the rate of brain change in cognitive decline through genetic pleiotropy, concordance, or a combination of both mechanisms. Gene-set analysis highlighted the involvement of vascular dilation, inflammation, and nucleus protein stabilization processes specifically in the rates of brain changes associated with cognitive decline. To further explore the genetic findings, we performed exploratory Transcriptome-wide Association Studies (TWAS) and phenome-wide association studies (pheWAS), which allowed us to systematically identify mRNA expression of genome-wide significant genes and analyze their associations with various phenotypes. By identifying genetic variants associated with structural brain changes in individuals with cognitive decline, our study sheds light on the biological pathways underlying dysfunctional brain development in this population.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1524 Genetically elevated abdominal fat and non-fasting triglycerides: implications for cardiovascular risk.

Authors:

M. R. Christiansen[1], G. D. Carrasquilla[1], T. O. Kilpeläinen[1,2]; [1]Novo Nordisk Fndn. Ctr. for Basic Metabolic Res., Univ. of Copenhagen, Copenhagen, Denmark, [2]Novo Nordisk Fndn. Ctr. for Genomic Mechanisms of Disease, Broad Inst. of MIT and Harvard, Cambridge, MA

Background: Elevated fasting triglyceride (TG) levels are a risk factor for cardiovascular disease (CVD). While most previous studies have focused on the association between fasting TG levels and CVD risk, non-fasting TG levels may be more predictive of CVD, as they reflect changes in remnant cholesterol levels. Recently, we showed that abdominal obesity is associated with increased postprandial TG levels after a high-fat meal, independent of fasting TG levels. These differences may reflect the relatively high lipolytic activity of visceral fat, compared to lower-body fat that stores excess fatty acids after a meal, acting as a "metabolic sink". However, it is unclear whether these findings are generalizable to the broader population and whether they have implications for CVD risk.
Objective: Using the UK Biobank, we aimed to study how abdominal obesity affects non-fasting and fasting TG levels and assess the impact on incident CVD risk.
Methods: We utilized data from the UK Biobank, including TG levels for 382,203 non-fasting individuals (<8 hours since last meal) and 15,175 fasting individuals (≥8 hours since last meal). A genetic risk score for waist-hip ratio adjusted for BMI (WHR_{adjBMI}) was used to determine the effect of abdominal obesity on non-fasting and fasting TG levels to minimize the influence of confounding and reverse causation. Furthermore, the association between the WHR_{adjBMI} GRS and incident CVD in non-fasting and fasting individuals was assessed by Cox regression models, controlling for age, sex, hypertriglyceridemia treatment, assessment center, genotype chip, and 15 genetic principal components.
Results: The genetic score for WHR_{adjBMI} showed a 25% greater effect on non-fasting TG levels (beta = 0.005 SD per allele) than fasting TG levels (beta = 0.004 SD per allele). For every SD unit increase in the WHR_{adjBMI} genetic score, there was an increased CVD risk in the non-fasting group (Hazard Ratio = 1.003, 95% CI 1.002; 1.004), but not in the fasting group (Hazard Ratio = 1.000, 95% CI 0.995; 1.006). The effect of the WHR_{adjBMI} score on CVD was reduced after adjusting for non-fasting TG levels.
Conclusion: Genetic predisposition to abdominal obesity increases non-fasting TG levels more than fasting TG levels. Furthermore, the association between the genetics of abdominal obesity and incident CVD is stronger in the non-fasting than in the fasting state.
Grants: Novo Nordisk Foundation (NNF18CC0034900, NNF20OC0063707, NNF17SA0031406, and NNF21SA0072102), Danish Diabetes Academy, European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie (846502), UKBB study application number 32683.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1525 Genetically informed type 2 diabetes risk prediction: leveraging machine learning for personalized approaches

Authors:

A. Khattab[1,2], S-F. Chen[1,2], A. Torkamani[2,1]; [1]Scripps Res., La Jolla, CA, [2]Scripps Res. Translational Inst., La Jolla, CA

Background: Early detection and prevention of type 2 diabetes (T2D) are crucial considering its global prevalence. This study proposes comprehensive person-centered risk prediction models for T2D, integrating environmental, clinical, and genetic factors using machine learning (ML) techniques. Methods: Utilizing data from the UK Biobank, we identified a prevalent T2D group at baseline (8,969 cases and 107,981 controls) and an incident T2D group over a 10-year follow-up (21,470 cases and 258,485 controls). Our models incorporated three types of risk factors: environmental, clinical (including biomarkers and diseases with known diabetes risk like CVD), and genetic factors in the form of polygenic risk scores (PRS) related to both clinical risk factors and T2D pathogenesis. Decision-tree-based ensemble ML algorithms were employed to train separate models for each group. The importance of features was assessed using SHAP values, and the top predictors were selected to build the final model. Results: Serving as a baseline model, the T2D prevalence model, consisting of 395 features, exhibited an ROC-AUC of 0.99, an area under the precision-recall curve (PR AUC) of 0.96, and an F1 score of 0.9. The T2D incidence model, incorporating the same 395 features, achieved an ROC-AUC of 0.97, a PR AUC of 0.87, and an F1 score of 0.78. Conclusions: While this study is still in progress, our interim results highlight the potential of ML in identifying relevant risk factors. The metrics of our comprehensive ML models, including the ROC-AUC, PR AUC, and F1 score, indicate their higher performance compared to existing literature. These models have the capacity to assist clinicians in making more precise and individualized decisions regarding T2D screening and treatment, ultimately aiming to prevent the onset of T2D. Future validation will be conducted on the All of Us dataset, which is known for its greater diversity, to further solidify our findings.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1526 Genetically predicted telomere length is associated with adverse clinical and pathological outcomes of Alzheimer's disease and dementia

Authors:

M. Lynch[1], J. Jin[1], M. Taub[2], J. Farfel[3], J. Yang[3], N. Chatterjee[4], D. Bennett[3], R. Mathias[5]; [1]Johns Hopkins Univ., Baltimore, MD, [2]Johns Hopkins, Baltimore, MD, [3]Rush Med. Coll., Chicago, IL, [4]Johns Hopkins Univ., Gaithersburg, MD, [5]Johns Hopkins Univ, Baltimore, MD

Telomere Length (TL) attrition is an established hallmark of aging and shorter TL has been associated with aging-related phenotypes including Alzheimer's disease (AD) and mild cognitive impairment (MCI). However, the interplay between TL and pathologic measures of AD has not been as well established. This study examined the association of genetically predicted TL (gTL) with dementia, MCI, level of and change in cognition, and AD-related pathologies. We leveraged existing genotype array data on N=2057 European ancestry samples from the Religious Orders Study (ROS) and the Rush Memory and Aging Project (MAP). In the combined ROSMAP sample, 37.6.% were diagnosed with clinical dementia, 22.9% with MCI, and 64.9% were diagnosed with pathologic AD among the n that came to postmortem. The mean age was 78.8 years old at blood draw. The polygenic risk score (PRS) previously derived for TL based on multi-ancestry TOPMed data using a ME-Bayes SL approach was used to compute gTL in ROSMAP. The approach uses a Bayesian framework to first model the genetic correlation in SNP effect size across ancestries, followed by a super learning step to combine PRS across various tuning parameter settings and all ancestries. Next, we tested for associations between gTL and four clinical outcomes and 12 neuropathologies, adjusting for age at death and sex. Education was also included as a covariate when assessing clinical outcomes. In ROSMAP, higher gTL is associated with lower odds of dementia (OR= 0.86, p=0.0034) and MCI (OR=0.86, p=0.0036), higher global cognition (Beta=0.32, p=0.0100), less severe cognitive decline (Beta=0.02, p=0.0332), and less global AD pathology (Beta=-0.17, p=0.0268) and tau tangle density (Beta=-0.44, p=0.0061).Our findings provide evidence linking shorter telomere length with adverse clinical and pathological outcomes of AD. We are currently extending this research to include a more diverse ancestral representation, utilizing data from both the Alzheimer's Disease Sequencing Project and the Health and Retirement Study cohorts.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1527 Genetically regulated gene expression influences chronotype and sleep duration in diverse populations.

Authors:

G. Chilton[1], K. Bozeman[1], M. Patel[1], H. Wheeler[2]; [1]Loyola Univ. Chicago, Chicago, IL, [2]Loyola Univ Chicago, Chicago, IL

Sleeping habits are integral components of daily life and circadian misalignment can increase cardiometabolic disease risk. Two important sleeping traits are chronotype, defined as the inclination to be a morning person, and sleep duration. Heritability estimates are 50% for chronotype and 46% for sleep duration. Populations with African genetic ancestries benefit from more genetic variation and smaller linkage blocks, increasing potential to discover new underlying mechanisms of complex traits. However, around 80% of GWAS are done on European populations, resulting in very little diversity of results and insufficient predictive power across diverse populations. Here, we focused on how genetically regulated gene expression may affect sleep phenotypes in diverse populations. We used Pan-UK Biobank GWAS summary statistics and Genotype Tissue Expression (GTEx) Project multivariate adaptive shrinkage gene expression prediction models implemented in S-PrediXcan to perform transcriptome-wide association studies (TWAS) on the chronotype and sleep duration phenotypes. For chronotype, there were 498,781 total participants from 6 diverse populations (META) and 5,561 participants from the African ancestries population (AFR). For sleep duration, there were 501,504 META participants with 6,382 participants from AFR. We performed S-PrediXcan on the META and AFR populations for each phenotype for 13 brain tissues. In the META analysis, 89 genes significantly associated with chronotype and 63 were significant in more than one tissue ($p < 5.0e-8$). Additionally, of the 22 genome-wide significant genes for sleep duration, 10 were significant across several tissues. The most significant association was increased *RGS16* expression with the “morning person” chronotype in hypothalamus and 8 other tissues (P = 4.7e-46) in the META analysis. *RGS16* (regulator of G protein signaling 16) is a highly plausible association because the protein is involved in the formation and functioning of retinal ganglion cells by sending external light signals to the circadian clock region of the hypothalamus known as the suprachiasmatic nucleus. In the AFR analysis, the most significant associations were decreased *DOK5* expression with the “morning person” chronotype in 6 tissues (P = 9.0e-5) and increased *ATP1A3* expression with increased sleep duration in 13 tissues (P = 6.6E-6), both including hypothalamus, potentially implicating these genes in sleep phenotypes for the first time. These results are important in determining how an individual’s genotype may affect their sleeping habits and demonstrating how expanding the diversity of genetic research can increase discovery.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1528 Genetically-predicted gene expression identifies novel candidate genes for Polycystic Ovary Syndrome

Authors:

H. Rao[1], N. Josyula[2], L. Fernandez-Rhodes[3], A. Justice[4]; [1]Pennsylvania State Univ., State College, PA, [2]Geisinger Hlth.System, Danville, PA, [3]Penn State Univ., State College, PA, [4]Geisinger, Danville, PA

Polycystic Ovary Syndrome (PCOS) is a multifactorial and polygenic condition that affects metabolism, endocrine function, adiposity, menstrual cycle, fertility, and mental illness. Genome wide association studies have found single nucleotide polymorphisms (SNPs) associated with PCOS, most of which are located in non-coding regions and with limited knowledge on any functional role in PCOS pathogenesis. To identify the genes that may be relevant for PCOS-associated genetic loci, we conducted a Transcriptome-wide Association study (TWAS) for PCOS in female participants of the Geisinger MyCode Community Health Initiative Study (MyCode). Our analyses included 63,806 females with genetically-linked electronic health record data. Ninety percent of the study self-identified as of white/European American, 3% as black/African American, 4% as Hispanic/Latino, and <1% as Asian/Asian American. There were 2,466 females with PCOS in our study (prevalence= 3.82%), based on ICD billing codes and Rotterdam criteria to diagnose PCOS. Genetically regulated gene expression was estimated for PCOS-relevant tissues (ovary, uterus, brain, adipose, pancreas, whole blood, pituitary, and thyroid) using PrediXcan software and GTEx v8 reference data and used in subsequent TWAS analyses. Analyses were adjusted for age, self-identified ancestry, and first 10 genetic ancestry principal components. TWAS identified a significant ($P< 0.05/22,077$ genes tested, 2.26×10^{-6}) association for a long non-coding RNA (lncRNA), Lnc-APOL1-16 (Ensemble id: ENSG00000279652), on chromosome 22 with PCOS in 5 tissues (Adipose Subcutaneous, Adipose Visceral Omentum, Pituitary, Uterus, and Testes). LncRNAs can affect chromatin structure, RNA splicing, stability, and translation; however, there is no prior evidence linking this lncRNA to PCOS. Genes *IL4R* and *RBFOX2* achieved suggestive significance ($p<2.26 \times 10^{-5}$), in visceral omentum adipose tissue, and thyroid gland, respectively. *IL4R*, interleukin 4 receptor gene, is a protein coding gene that regulates IgE and TH2 cell production and is implicated in the insulin resistance pathway. *RBFOX2* (also known as *RBM9*) is an RNA binding protein known to regulate estrogen-receptor-1 transcriptional activity. We identified tissue-specific expression of two candidate genes (*IL4R* and *RBFOX2*) with strong evidence of a relationship with PCOS risk and comorbidities. Additionally, we identify a novel association between *Lnc-APOL1-16* and PCOS in several relevant tissues. Future analyses are needed to evaluate the utility of these newly identified genes as potential biomarkers or therapeutic targets for PCOS.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1529 Genome Wide Association Study accounting for local ancestry reveals new loci associated with Parkinson Disease in a Latin Americans

Authors:

T. Peixoto Leal[1], M. Gouveia[2], V. Borda Pua[3], M. Inca Martinez[1], E. Mason[1], A. Horimoto[4], D. Loesch[5], E. Sarihan[1], M. Cornejo-Olivas[6], L. Torres[6], P. Mazzetti-Soler[6], C. Cosentino[6], E. Sarapura-Castro[6], A. Rivera-Valdivia[6], A. Medina[7], E. Dieguez[8], V. Raggio[8], A. Lescano[8], V. Tumas[9], V. Borges[10], H. Ferraz[10], C. Rieder[11], A. Schumacher-Schuh[12], B. Santos-Lobato[13], C. Velez-Pardo[14], M. Jimenez-Del-Rio[14], F. Lopera[14], S. Moreno[14], P. Chana-Cuevas[15], W. Fernandez[16], G. Arboleda[17], H. Arboleda[16], C. Arboleda-Bustos[16], D. Yearout[18], C. Zabetian[19], T. Thornton[20], T. O'Connor[3], I. Fernandez Mata[1]; [1]Cleveland Clinic, Cleveland, OH, [2]NIH, Bethesda, MD, [3]Univ. of Maryland Sch. of Med., Baltimore, MD, [4]Brigham and Women's Hosp., Framingham, MA, [5]Univ. of Maryland Baltimore, Baltimore, MD, [6]Inst. Natl. de Ciencias Neurologicas, Lima, Peru, [7]Univ. Natl. del Altiplano, Puno, Peru, [8]Univ. de la República, Montevideo, Uruguay, [9]Univ.e de São Paulo, Ribeirão Preto, Brazil, [10]Univ.e Federal de São Paulo, São Paulo, Brazil, [11]Univ.e Federal de Ciências da Saúde de Porto Alegre, Porto Alegre, Brazil, [12]Hosp. de Clínicas de Porto Alegre, Porto Alegre, Brazil, [13]Univ.e Federal do Pará, Belem, Brazil, [14]Univ. de Antioquia (UdeA), Medellín, Colombia, [15]Univ. de Santiago de Chile, Santiago, Chile, [16]Univ. Natl. de Colombia, Bogotá, Colombia, [17]Univ. Natl. de Colombia, Bogota, Colombia, [18]VA Puget Sound Hlth.Care System, Seattle, WA, [19]VA Puget Sound Hlth Care Sys, Seattle, WA, [20]Univ. of Washington, Atlanta, GA

Genome Wide Association Studies (GWAS) have identified more than 100 loci associated with Parkinson Disease (PD). However, previous studies were mostly focused on European and Asian populations, leading to limited understanding of the genetic architecture of PD in Latin Americans. Recently we performed the first GWAS of PD in Latin Americans using 1,502 samples from Brazil, Chile, Colombia, Peru and Uruguay in the Latin American Research consortium on the GEnetics of Parkinson's Disease (LARGE-PD). To evaluate the impact of admixture in LARGE-PD GWAS, we performed local ancestry-based analysis (TRACTOR) using 1,450 unrelated LARGE-PD samples. LARGE-PD samples were imputed using TOPMed imputation server and local ancestry was inferred using RFMix v1. We identified 19 PD risk intergenic variants near *NRROS, SNCA, CTNND2, YTHDF1P1, SLC15A1, and ZNF264* and variants in *MAML3*, *SBSPON*, and *TRAV27*. Importantly, we replicated the association in *SNCA* (rs356182), which is one the most common genes implicated in Parkinson's disease. Additionally, the *NRROS* was found to be suggestively associated in our previous LARGE-PD GWAS, but significantly associated in our analysis accounting for local ancestry. The *NRROS* signal is Native-American ancestry specific and it is the only statistical significant p-value when we calculate the p-value based in a single ancestry. According FUMA analysis, our dataset has variants associated or linked with variants reported on GWAS catalog, which we found variants associated Schizophrenia (*SNCA*) and PD (*SNCA*), and in linkage with Educational attainment (*SNCA*) and Smoking initiation (*MAML3*). Notably, Mendelian Randomization analysis showed that Educational attainment is a risk factor while smoke is a protective factor for PD. Overall, after accounting for local ancestry, we discovered new ancestry specific loci associated with PD (*NRROS*) while confirming loci from previous conventional GWAS (*SNCA*). These results highlight the importance of account for local ancestry in GWAS using admixed populations such as Latin Americans.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1530 Genome wide association study identifies novel risk loci and provides new insights into genetic architecture of severe obesity.

Authors:

M. Krishnan[1], M. Yaser[2], A. Justice[3], G. Chittoor[3], H-H. Chen[4], R. Smit[5], M. Preuss[6], E. Parra[7], M. Cruz[8], Q. Hui[9], P. Wilson[9], Y. Sun[9], W. Zhao[10], J. Smith[10], L. Lange[11], M. Meyer[12], C-T. Liu[13], L. Yanek[14], R. Loos[5], P. Below[4], K. North[15], M. Graff[15]; [1]Univ. of North Carolina, Chapel Hill, NC, [2]Univ. of North Carolina Chapel Hill, Chapel Hill, NC, [3]Geisinger, Danville, PA, [4]Vanderbilt Univ., Nashville, TN, [5]Icahn Sch. of Med. at Mount Sinai, New York, NY, [6]Mount Sinai Sch. of Med., New York, NY, [7]Univ. of Toronto, Toronto, ON, Canada, [8]Inst. Mexicano del Seguro So, Mexico D.f. C.p 06725, Mexico, [9]Emory Univ., Atlanta, GA, [10]Univ. of Michigan, Ann Arbor, MI, [11]Univ Colorado Denver, Aurora, CO, [12]Univ. of Colorado Anschutz Med. Campus, Denver, CO, [13]Boston Univ. SPH, Boston, MA, [14]Johns Hopkins Univ, Baltimore, MD, [15]Univ North Carolina, Chapel Hill, NC

Severe obesity (SevO; BMI≥40kg/m^2) is an understudied global health issue, limiting knowledge about its mechanisms and impacts. To provide novel insights into the genetic architecture and mechanisms of SevO, we performed GWAS in 120,676 individuals across 11 ancestrally diverse population-based studies, generated a polygenic score (PRS) to quantify susceptibility to SevO, and assessed PRS applicability in phenome-wide association studies (PheWAS) to assess associations with chronic conditions. Ancestry and sex specific GWAS and meta-analysis were conducted with three different classes of SevO traits; 95th percentile [cases defined as the upper 5th percentile and controls as the 5th-50th percentile], Obesity Class 3 ([BMI≥40kg/m^2] vs. BMI≤25 kg/m^2) and Obesity Class 4 ([BMI≥50kg/m^2] vs. BMI≤25 kg/m^2). All analyses were adjusted for age, age^2, 10 principal components of ancestry and study specific covariates. We assessed replication of our novel findings in a follow-up sample of 479,213 individuals across 6 ancestrally diverse studies. We derived a weighted class 3 SevO PRS in the UK BioBank (UKBB) and performed a PheWAS of 19 clinical conditions. Participants were stratified according to PRS deciles and lifestyle behaviors including physical activity, diet, alcohol, and sleep were compared for those in the bottom (<10th percentile) with the top (>90th percentile) deciles. Across the three SevO traits we mapped 12 independent variants (P<5×10^{-8}) not previously associated with obesity related traits. Six variants located in *UBXN4, BHLHE40-AS1, TENM2, PLCL2, ZNF184, PSMD9* were replicated (Bonferroni P<0.004). Over 200 (21%) known signals for BMI were significantly associated with SevO, demonstrating a strong overlap in their genetic architecture. The class 3 SevO PRS in UKBB explained 11.4%, 6.0% and 14.4% of the phenotypic variance for Europeans, Africans, and South Asian participants, respectively and was associated with 37% of phenotypes considered (max 1668); many are known biological sequalae of obesity and some are novel (reticulocyte traits). SevO was present in 4.9% of those >90th percentile for PRS compared to 0.55% of those in <10th percentile, corresponding to a 10-fold increased risk of SevO. Within the 90th percentile, 20% of individuals had BMI<25kg/m^2 and reported significantly healthier lifestyle behaviors when compared to those with SevO. Our study expands the number of identified SevO signals, demonstrates a strong overlap in the genetic architecture of SevO and BMI, and reveals a remarkable impact of SevO on the clinical phenome, affording new opportunities for clinical prevention and mechanistic insights.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1531 Genome wide association study of Vitamin D deficiency in the Middle East with a relevant characterization of the novel*SDR42E1*gene

Authors:

N. Hindi[1], O. Albagha[2], P. M. Ericsson[1], G. M. Nemer[2]; [1]HBKU, Doha, Qatar, [2]Hamad Bin Khalifa Univ., Doha, Qatar

Vitamin D deficiency is highly prevalent in Middle Eastern countries, yet studies on Vitamin D polymorphisms and polygenic models have primarily focused on European populations. Interestingly, genomic research has identified a nonsense variant in the *SDR42E1* gene as a potential contributor to Vitamin D deficiency, but its role in Vitamin D metabolism remains understood.Here, we conducted the first genome-wide association study (GWAS) on Middle Eastern populations to identify genetic determinants of Vitamin D levels using genome and exome sequencing approaches on 6,047 and 199 Qatari and Lebanese subjects, respectively. We also functionally and structurally characterized the *SDR42E1* through CRISPR/Cas9 genome editing technique on cutaneous HaCat and intestinal HCT116 human cells.We discovered a novel variant, rs2298850 (*P* value = 1.71×10^{-08}), in the *GC* gene among Qataris. We further identified two novel suggestive variants, rs141064014 in *MGAM* (*P* value = 4.40×10^{-06}) and rs7036592 in *PHF2* (*P* value = 8.43×10^{-06}), in Lebanese individuals. A meta-analysis combining our Middle Eastern with the largest European GWAS confirmed the replication of numerous known variants. We also demonstrated that European-derived polygenic scores for Vitamin D exhibited lower performance than European estimations yet effectively predicted hypovitaminosis D.Significant associations between *SDR42E1* and genes involved in Vitamin D pathways were identified, including *ALPP*, *ABCC1*, and *SLC7A5*. Gene regulators of steroid and lipid biosynthesis, cellular senescence, and cancer prognosis were significantly affected in *SDR42E1* HaCat and HCT116 knockout models. We observed significant alterations in Vitamin D metabolites, including 24R,24,25 Dihydroxyvitamin D, in the *SDR42E1* HaCat knockout cells. In conclusion, our research underscores the diversity in genetic architecture and its impact on precision medicine across different populations. The study of the novel *SDR42E1* gene contributes to a comprehensive understanding of Vitamin D metabolism and related health conditions, laying a foundation for future research and clinical precision medicine applications.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1532 Genome-wide analyses of machine-learning predicted brain aging identifies new genomic loci and causal cross-phenotype associations

Authors:

P. Jansen[1,2], H. E. M. Abaddy[2], S. C. de Lange[2], T. A. Gausvik[2], I. A. Dekkers[3], J. E. Savage[2], M. van den Heuvel[2], D. Posthuma[2]; [1]Amsterdam UMC, Amsterdam, Netherlands, [2]VU Univ., Amsterdam, Netherlands, [3]Leiden Univ. Med. Ctr., Leiden, Netherlands

Introduction: Brain aging is a complex multifaceted process with profound implications on health, impacting cognitive function, neurological disorders, and overall lifespan. The specific genetic factors and biological mechanisms that influence this complex process remain to be fully understood. Here, we conduct a genome-wide association study (GWAS) on brain-age derived from machine learning in the largest single-site imaging dataset of the UK Biobank.
Methods: In UK Biobank we ran multiple ML algorithms to predict brain-age and calculate brain-age gaps (difference between predicted and actual brain-age), using structural and diffusion tensor magnetic resonance imaging (DTI) data. We perform GWAS analyses of brain-age in N=29,030 individuals to map genetic variants to brain aging, and extensive follow-up analyses using FUMA and other bioinformatic tools. We explore genetic overlap, and causality with possible risk factors of accelerated brain aging using LD score regression (LDSC) and Mendelian Randomization (MR).**Results:** XGBoost performed well in predicting brain age (mean gap: ~4.5 years). We observe a high SNP heritability (~25%) of BAG, with enrichment in conserved regions. We discover 9 lead variants in 7 genomic loci, and 153 mapped genes. We show significant genetic overlap in LDSC analyses with various phenotypes (depression, alcohol use). Mapped genes contain previously involved in (monogenic) neurodegenerative disease (MAPT, GFAP). Causality analyses using inverse-variance weighted MR highlights lifestyle factors (smoking) and chronic disease (type 2 diabetes) as causal factors for brain aging.
Conclusions: Our results illustrate the possibilities of advanced machine learning algorithms to derive novel, heritable brain phenotypes from multimodal brain MRI datasets. Combined with large-scale genetic data, we uncover the cross-trait association of brain-age with risk factors that shed new light on the causes of brain aging.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1533 Genome-wide analysis identifies genetic associations for multiple deep-learning-derived brain imaging endophenotypes.

Authors:

B. Guo[1], Z. Xie[2], S. Islam[1], A. Gottlieb[3], H. Chen[4], D. Zhi[5]; [1]The Univ. of Texas Hlth.Sci. Ctr. at Houston, Houston, TX, [2]Univ. of Texas Hlth.Sci. Ctr. at Houston, Houston, TX, [3]UTHlth., Houston, TX, [4]The Univ of Texas Hlth.Sci. Ctr. at Houston, Houston, TX, [5]U Texas Hlth.Sci. Ctr. at Houston, Houston, TX

Investigating the genetic architecture of brain structure from imaging data can be challenging, partly because of the difficulties involved in developing accurate and impartial brain morphology descriptors. Previously, we have created a 128-dimensional brain imaging endophenotype (ENDOs) using an unsupervised deep representation learning approach, and identified multiple loci from single-phenotype genome-wide association studies (GWAS) for each of the ENDO dimensions, using data from the UK Biobank. However, this approach may miss genetic associations where one single nucleotide polymorphism (SNP) is weakly or moderately associated with multiple ENDO dimensions. Here we develop a computationally efficient method, Joint Analysis of multiple GWAS (JAGWAS), and applied it on those ENDOs. Compared to existing multi-phenotype GWAS methods, JAGWAS can take related individuals and is free of the need for reference panels or permutations. Multi-phenotype GWAS of the ENDOs in discovery (n = 22,948 T1/T2) and replication cohorts (n = 12,839/11,708 for T1/T2) identified 22,552/21,109 significant replicated variants from 215/183 independent loci, respectively for T1 and T2. Out of the loci discovered, 51/39 cannot be found by taking the minimum p-value across single-phenotype GWAS of 128 ENDOs. Out of the genes mapped, 27 out of 46 T1 genes and 23 out of 34 T2 genes had significant transcription factor-expression associations in at least one of 47 tissues from the GTEx database. Out of those, 9 genes from T1 and 9 genes from T2 were from brain tissues, 3 of them came from both. Meta-analysis of the discovery and replication cohorts identified 159,769/150,076 variants which clustered into 906/848 independent association loci for T1/T2, at the genome-wide significance threshold of 5×10^{-8}, and 516/471 loci overlapped with expression quantitative trait loci from brain tissues in the GTEx database. We identified the optimal linear combination of ENDO dimensions that was most significantly associated with each SNP, and used a perturbation-based decoder interpretation (Eigen-PerDI) approach to identify brain signatures corresponding to the SNP. We showed that Eigen-PerDI identified relevant brain regions of a SNP more interpretable than individual ENDO dimensions. Our results showed that multivariate GWAS using unsupervised deep learning derived ENDOs, can boost the discovery of new genetic markers relevant to Alzheimer's disease (AD) which may provide insights into the genetic architecture of the brain.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1534 Genome-wide association analysis and construction of polygenic risk score model for predicting paroxysmal atrial fibrillation using machine learning models.

Authors:

M. Shiomi[1], Y. Nagata[1,2], T. Sudo[3], T. Ichikawa[2], K. Ihara[4], K. Asada[5], Y. Tanaka[6], Y. Yamauchi[7], T. Sasaki[8], H. Hachiya[9], Y. Imai[10], H. Fujita[11], T. Sasano[12], T. Furukawa[4], T. Tanaka[1,2]; [1]Dept. of Human Genetics and Disease Diversity, Graduate Sch. of Med. and Dental Sci., Tokyo Med. and Dental Univ., Tokyo, Japan, [2]Bioresource Res. Ctr., Tokyo Med. and Dental Univ., Tokyo, Japan, [3]Inst. of Ed., Tokyo Med. and Dental Univ., Tokyo, Japan, [4]Dept. of Bio-informational Pharmacology, Med. Res. Inst., Tokyo Med. and Dental Univ., Tokyo, Japan, [5]Cancer Translational Res. Team, RIKEN Ctr. for Advanced Intelligence Project, Tokyo, Japan, [6]Cardiovascular Ctr., Yokosuka Kyosai Hosp., Kanagawa, Japan, [7]Dept. of Cardiology, Japan Red Cross Yokohama City Bay Hosp., Kanagawa, Japan, [8]Dept. of Cardiology, Natl. Hosp. Organization Disaster Med. Ctr., Tokyo, Japan, [9]Cardiovascular Ctr., Tsuchiura Kyodo Hosp., Ibaraki, Japan, [10]Div. of Clinical Pharmacology, Dept. of Pharmacology, Jichi Med. Univ., Tochigi, Japan, [11]Div. of Cardiovascular Med., Saitama Med. Ctr., Jichi Med. Univ., Saitama, Japan, [12]Dept. of Cardiovascular Med., Tokyo Med. and Dental Univ., Tokyo, Japan

Atrial fibrillation (AF) is one of the most common heart rhythm disorders, often developing as paroxysmal atrial fibrillation (PAF) and progressing to persistent AF. Diagnosis of PAF is challenging due to its intermittent occurrence. It is important to design a method to stratify the risk of PAF. Polygenic risk scores (PRS) facilitate the identification of individuals at high risk of a disease and can be clinically used for risk stratification. This study aimed to construct PRS-based models to predict PAF by conducting a comparative analysis of the genomic DNA isolated from patients with and without PAF. For model construction, we conducted a genome-wide association study (GWAS) of 1,208 subjects, and the results were used as summary statistics. We tested the constructed models using the data of 600 subjects who were not in part of the GWAS. Based on several parameters, including PRS and clinical factors, the Light Gradient Boosting Machine framework, an ensemble model based on decision trees for classification and regression prediction, was used to construct PAF-predicting models. We performed 5-fold cross-validation and evaluated model performance by determining the area under the curve (AUC). This study identified 57 significant PAF-associated SNPs by GWAS ($P < 5 \times 10\text{-}8$). To obtain the optimal model, SNPs were selected per the associated resultant *p*-values of the GWAS, and multiple R2 and *p*-value thresholds were considered to obtain candidate models that could predict PRS. The performance of these models was evaluated using test and validation datasets with a randomly split cohort of 600 subjects, and models with AUCs greater than 0.7 were selected. The investigation of the utility of clinical factors in identifying patients at risk of PAF revealed that AUC further increased upon including risk factors of PAF and other clinical factors in the model. These results suggest that predictive models including PRS and clinical factors may help in identifying PAF cases.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1535 Genome-wide association analysis and Evolutionary Action Equation for the Fitness Effect of Missense Mutations of Systemic Sclerosis identifies new susceptibility locus and mechanistic pathway genes

Authors:

S. Ketkar[1], H. Dai[1], B. Dawson[1], L. Burrage[1], D. Murdock[2], J. Asmussen[1], K. Wilhelm[1], O. Lichtarge[1], B. Skaug[3], M. Kerick[4], J. Martin[4], S. Assassi[3], M. Mayes[3], B. Lee[1]; [1]Baylor Coll. of Med., Houston, TX, [2]The Univ. of Texas Hlth.Sci. Ctr. at Houston, Houston, TX, [3]Univ. of Texas McGovern Med. Sch., Houston, TX, [4]Inst. of Parasitology and Biomedicine López-Neyra, CSIC, Granada, Spain

Systemic Sclerosis (SSc) has considerable morbidity and the highest mortality of all the systemic autoimmune diseases. SSc is a chronic, autoimmune connective tissue disorder that is primarily characterized by fibrosis of the skin and internal organs associated with a vasculopathy and exhibiting wide clinical heterogeneity. The causes of SSc are unknown and it is generally considered to result from a combination of external triggers operating in the context of genetic susceptibility. Several genome-wide association studies (GWAS) have shed light on the genetic component and estimated that it contributes 30% to the risk of developing SSc.

We conducted an extensive study using the exome sequencing (ES (performed by Regeneron in an academic-industry partnership)) from the participants of SSc Registry to elucidate etiology and pathogenesis of SSc, to identify novel SSc susceptibility loci, and to identify causative pathways and prioritize variants that stratify disease severity by examining the genotype-phenotype relationships which shape health and population fitness.

We first performed GWAS of 2,865 Caucasian cases and 1,043 controls. Genotype imputation of ES data with the 1,000 Genome Project Phase 3 panel yielded 5,522,961 variants for the analysis. Case-control association analysis was done using PLINK v1.9 logistic regression model of additive effects, including gender and first two principal components as covariates.

We identified a novel susceptibility locus at *MICB* (rs2516497, $P = 8.42 \times 10^{15}$) within the HLA region. Additionally, we confirmed and firmly established the role of *HLA-DQA1* (rs1048372, $P = 2.95 \times 10^{21}$) and *HLA-DQB1* (rs2647032, $P = 1.32 \times 10^{16}$) gene regions as SSc genetic risk factors. Replication testing in an independent case-control set of European ancestry (949 SSc patients, 998 controls) validated association of *MICB* SNP (rs2516497, $P = 7.17 \times 10^{9}$) to SSc and confirmed it as an independent event, i.e., not in linkage disequilibrium with other HLA class genes.

Next, we applied an evolutionary action method which predicts the impact of missense variants on protein function by measuring the fitness effect based on phylogenetic distances and substitution odds in homologous gene sequences. Using the missense variants from the ES data, this method corroborated genes associated with SSc and identified additional genes in the interferon pathway among others. *IFI44L* and *IFIT5* are interferon genes and are shown to be dysregulated in SSc blood and skin.

Using a GWAS analysis and evolutionary action method, we have identified a novel locus and missense mutations in the interferon pathway that may contribute to SSc pathogenesis.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1536 Genome-wide association analysis of sarcopenia-related traits in the CLSA cohort

Authors:

T. SASAKO[1,2], Y. Ilboudo[1], S. Yoshiji[1,3,4,5], T. Lu[1,6,7], K. Y. H. Liang[1,6], Y. Chen[1,3], B. Richards[1,3,7,8,9]; [1]Lady Davis Inst. for Med. Res., Jewish Gen. Hosp., Montreal, QC, Canada, [2]Tanaka Diabetes Clinic Omiya, Saitama, Japan, [3]Dept. of Human Genetics, McGill Univ., Montreal, QC, Canada, [4]Kyoto-McGill Intl. Collaborative Program in Genomic Med., Graduate Sch. of Med., Kyoto Univ., Kyoto, Japan, [5]Japan Society for the Promotion of Sci., Tokyo, Japan, [6]Quantitative Life Sci. Program, McGill Univ., Montreal, QC, Canada, [7]5 Prime Sci. Inc, Montreal, QC, Canada, [8]Dept. of Epidemiology, Biostatistics and Occupational Hlth., McGill Univ., Montreal, QC, Canada, [9]Dept. of Twin Res., King's Coll. London, London, United Kingdom

Background: Growing attention has been focused on sarcopenia, an aging-related disease of skeletal muscle characterized by low muscle mass, low muscle strength, and low physical performance. Although it is a common disease in the elderly which is associated with various comorbidities and mortality, the mechanisms underlying its development are not well understood. **Aim:** To elucidate the genetic drivers of sarcopenia, we performed genome-wide association studies (GWASs) using data from the Canadian Longitudinal Study on Aging (CLSA) cohort, in which data of grip strength, skeletal muscle index, a parameter of lean mass, and gait speed, a parameter of physical performance, are all available. **Results:** We assessed 24,655 subjects of European ancestry, with a mean age of 63.0 years where 50.3% were females. We first performed a GWAS of grip strength and identified three genome-wide significant variants: *EFCAB8* (rs2377735) and *CABLES1* (rs28589524) are novel loci for grip strength, and *UQCC1* (rs2425063) is close to *GDF5* which was reported to be associated with low grip strength and lean mass. We also performed a GWAS of skeletal muscle index, identifying three loci: *FTO* (rs1558902) and *RNU4-17p*/*MC4R* (rs35361355) are close to those which were reported to be associated with lean mass and grip strength, and *CCDC26* (rs4568584) is a novel locus for lean mass. No genome-wide significant findings were found for gait speed. Moreover, we performed a principal component (PC) analysis of the three traits and found that just a single variant showed a significant association with the first PC, a locus near *GDF5-AS1*/*GDF5* (rs224329) which is close to *GDF5* mentioned above. Regression analyses of the first PC showed that the first PC was associated with type 2 diabetes, osteoporosis, osteoporotic fracture, Alzheimer's disease, and other aging-related diseases. **Conclusions:** This is the first study to examine the genetic determinants of the three sarcopenia-related traits to the best of our knowledge, in contrast with previous reports to focus on just one of them. Further analysis of the data is expected to provide insights into mechanisms underlying the development of sarcopenia, which could lead to the identification of potential therapeutic targets and biomarkers.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1537 Genome-wide association and HLA fine-mapping analysis of Hunner-type interstitial cystitis identify predisposing class II HLA variants.

Authors:

K. Sonehara[1,2,3], Y. Akiyama[1], D. Maeda[4], H. Katoh[1], T. Naito[2,3], K. Yamamoto[2,5], Biobank Japan, T. Morisaki[6], S. Ishikawa[1], T. Ushiku[1], H. Kume[1], Y. Homma[7], Y. Okada[1,2,3,5]; [1]Graduate Sch. of Med., The Univ. of Tokyo, Tokyo, Japan, [2]Osaka Univ. Graduate Sch. of Med., Suita, Japan, [3]RIKEN Ctr. for Integrative Med. Sci., Yokohama, Japan, [4]Graduate Sch. of Med. Sci., Kanazawa Univ., Kanazawa, Japan, [5]Immunology Frontier Res. Ctr. (WPI-IFReC), Osaka Univ., Suita, Japan, [6]The Univ. of Tokyo, Inst Med Sci, Tokyo, Japan, [7]Faculty of Med., Kyorin Univ., Tokyo, Japan

Hunner-type interstitial cystitis (HIC) is a rare, chronic inflammatory disease of the urinary bladder, with an extremely low prevalence of 10-50 cases per million population worldwide. HIC is among the most intractable, devastating bladder disorders, with severe impacts on patients' quality of life. Although recently generated evidence suggests its nature of the immune-mediated disease, the etiology and genetic background of HIC remain still enigmatic. Here, employing approximately 10% of the estimated HIC cases in Japan, we perform an initial genome-wide association study of 144 patients with HIC and 41,516 control participants, both of Japanese ancestry. Thanks to the largest case sample size ever reported for this rare disease, we identify a genome-wide significant association in the major histocompatibility complex (MHC) region on chromosome 6 (lead variant, rs1794275; risk allele frequency = 0.27; $P = 3.4 \times 10^{-9}$; odds ratio = 2.32 [95% CI 1.75-3.06]). The association is replicated in an independent dataset of 26 cases and 1,026 controls ($P = 0.014$). To fine-map the association, we perform an HLA imputation analysis using a high-resolution population-specific reference panel, pinpointing HLA-DQβ1 amino acid positions 71, 74, and 75 as the most significant associations ($P = 5.0 \times 10^{-8}$; OR, 1.94; 95%CI, 1.53-2.44). These amino acid variants are in complete linkage disequilibrium in the Japanese population ($r^2 = 1.00$). The completely linked haplotype of Thr71-Glu74-Leu75 shows a protective effect against HIC. Other than HLA-DQβ1 variants, we identify HLA-DPβ1 amino acid position 178 surpassing the region-wide significance threshold, which perfectly tags HLA-DPB1*04:02 ($P = 7.5 \times 10^{-8}$; OR, 2.35; 95%CI, 1.77-3.12). The three HLA-DQβ1 amino acid positions of 71, 74, and 75 are located together at the peptide binding groove, suggesting their functional importance in antigen-presentation. Our study provides the first evidence of genetic determinants of HIC risk, which may be attributed to class II MHC molecule antigen-presentation.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1538 Genome-wide association studies for alloimmunization among patients with sickle cell disease.

Authors:

Q. Sun[1], M. S. Karafin[1], M. Garrett[2], M. Telen[2], A. Ashley-Koch[2], Y. Li[1]; [1]UNC-Chapel Hill, Chapel Hill, NC, [2]Duke Univ., Durham, NC

Sickle cell disease (SCD) is a severe inherited blood disorder affecting about 100K US individuals, and blood transfusion is a key component in the clinical management of SCD patients. However, 8-76% SCD patients unfortunately become alloimmunized, many of which cause accelerated donor red blood cell destruction. Patients show substantial variability in their predisposition to alloimmunization, where genetic variability is one proposed component. Although several genetic association studies have been conducted for alloimmunization, the results have been inconsistent, and the genetic determinants of alloimmunization remain largely unknown. In an effort to advance genetic studies of alloimmunization, we performed a genome-wide association study (GWAS) in 236 African American (AA) SCD patients from the Duke OMG-SCD cohort, which is part of TOPMed, with whole genome sequencing data available. We applied logistic mixed models adjusting for sample relationship matrix and covariates, including presence of autoantibodies, hemoglobin genotype categories and sex. Despite the small sample size, we identified one genome-wide significant locus on chr12 ($p = 3.1e-9$) with no evidence of genomic inflation (lambda = 1.003). We also performed sensitivity analyses adjusting for additional covariates (Fy^b antigen and the amount of blood transfusion) and applying different sample grouping strategies based on the number of alloantibodies patients developed (>=2 v.s. 0 and >=2 v.s. <=1). These analyses consistently revealed several additional suggestive loci. Furthermore, these suggestive loci are supported by both eQTL evidence from GTEx whole blood and/or Jackson Heart Study PBMC RNA-seq data, and 3D chromatin conformation information derived from HiC data in spleen and/or K562 cells. In conclusion, we identified several novel genetic variants that are significantly associated with alloimmunization among SCD patients and linked those variants to genes leveraging functional annotation information. We call for the community to collect additional alloantibody information within SCD cohorts to further the understanding of the genetic basis of alloimmunization in order to improve transfusion outcomes.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1539 Genome-wide association studies identify a susceptibility locus to pelvic organ prolapse in the Japanese

Authors:

M. Matsunami[1], M. Imamura[2], A. Ashikari[1], X. Liu[3], N. Miyagawa[4], K. Tomizuka[4], K. Hikino[5], P. The Biobank Japan[6], K. Matsuda[6], C. Terao[7], M. Miyazato[1], S. Maeda[8]; [1]Univ. of the Ryukyus, Nishihara-cho, Japan, [2]Univ. of the Ryukyus, Nishihara, Okinawa, Japan, [3]RIKEN, Kawasaki, Japan, [4]RIKEN, Yokohama, Japan, [5]RIKEN Ctr. for Integrative Med. Sci., Yokohama, Japan, [6]The Univ. of tokyo, Tokyo, Japan, [7]RIKEN IMS, Yokohama, Japan, [8]Univ. of the Ryukyus, Nishihara, Japan

Pelvic organ prolapse (POP) is a one of *common female health* issues, characterized by a descent of pelvic organs into the vaginal cavity. Although contribution of genetic factors is suggested by a twin study, its genetic predisposition has been poorly understood. While a few genome-wide association studies (GWAS) for POP have been conducted in European populations, there is no report of GWAS for POP using Asian populations, including the Japanese. To identify genetic loci associated with susceptibility to POP in the Japanese, we performed GWAS for POP using two Japanese datasets, Okinawa Bioinformation Bank (OBi) 324 POP cases and 2,733 controls and BioBank Japan (BBJ) 447 POP cases and 74,561controls. Genotypings were performed using Illumina Asian Screening Array for OBi participants, or Illumina Infinium Omni Express, Human Exome, Infinium Omni Express Exome v1.0, and Infinium Omni Express Exome v1.2, for BBJ participants. Genotype imputation was performed by minimac4 using 1KGP3+JEWEL_3k_v1.2 (1000 genomes phase 3 + 3k Japanese) as a reference. It has been known that there are two distinct genetic groups in the Japanese based on the results of principal component analysis (PCA), namely Hondo and Ryukyu. The former is a major Japanese group and the later includes individuals originally from the Ryukyu Archipelago, southernmost islands of Japan. Considering these genetic backgrounds, we divided participants of each dataset into these two groups, and analyzed the association of ~9.6 million imputed SNP dosages (RSQ > 0.3) with POP by SAIGE for each study group (OBi-Ryukyu, OBi-Hondo, BBJ-Ryukyu and BBJ-Hondo). A Meta-analysis of GWAS for OBi-Hondo and BBJ-Hondo identified *WT1* locus as susceptibility to POP (rs10742277, Chromosome 11, $p = 2.43 \times 10^{-8}$, Odds ratio (OR) = 1.54, 95% Confidence Interval (CI) 1.32-1.80, n = 68,973). A same trend of association of this locus was also observed in a meta-analysis of GWAS for OBi-Ryukyu and BBJ-Ryukyu ($p = 5.92 \times 10^{-2}$, OR = 1.29, 95% CI 0.99 - 1.68, n = 9,092), and integration of all 4 studies further strengthen the association of this locus with POP ($p = 7.67 \times 10^{-9}$, OR = 1.47, 95% CI = 1.29-1.68, n = 78,065). In conclusion, we have identified *WT1* as a genetic locus susceptible to POP in Japanese women.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1540 Genome-wide association studies of blood pressure phenotypes in 4,819 participants from Samoa and American Samoa.

Authors:

T-A. Yapp[1], J. Z. Zhang[1], M. Krishnan[1], S. Liu[1], B. M. Blobner[1], H. Cheng[2], T. Naseri[3], M. S. Reupena[4], S. Viali[5], J. Tuitele[6], R. Deka[2], N. L. Hawley[7], S. T. McGarvey[3], D. E. Weeks[1], R. L. Minster[1], J. C. Carlson[1]; [1]Univ. of Pittsburgh, Pittsburgh, PA, [2]Univ. of Cincinnati, Cincinnati, OH, [3]Brown Univ., Providence, RI, [4]Lutia i Puava ʻae Mapu i Fagalele, Apia, Samoa, [5]Natl. Univ. of Samoa, Apia, Samoa, [6]Government of American Samoa, Pago Pago, American Samoa, [7]Yale Sch. of Publ. Hlth., New Haven, CT

Exploring the genetic determinants of blood pressure (BP) phenotypes in Polynesian individuals could help address inequity in research with marginalized populations and has the potential to provide additional insight to what is known from other populations around the globe about the biological foundations of such traits. In the current study, we performed genome-wide associations studies of five traits—systolic blood pressure (SBP), diastolic blood pressure (DBP), mean arterial pressure (MAP), pulse pressure (PP), and hypertension (HTN)—in 4,819 research participants across five subsamples from Samoa and American Samoa. Genotypes were measured via two methods: 895,103 variants were genotyped in a subsample of 3,053 individuals from Samoa with the Affymetrix 6.0 array and 669,569 variants were genotyped in a subsample of 1,766 individuals from Samoa and American Samoa with the Illumina Global Screening Array. We then imputed up to 7,408,897 variants using a haplotype reference panel based on 1,285 participants from Samoa that were whole-genome sequenced by the TOPMed Program. We performed association testing of inverse-normally transformed traits in each subsample separately using linear or logistic mixed models, adjusting for fixed effects of age, sex, polity, principal components of ancestry and a kinship random effect. The subsamples were meta-analyzed using the *p* values and effect direction with weighting by the subsample size. No variants were observed to be genome-wide significant; variants at fourteen unique loci associated with BP phenotypes were observed at $p < 1 \times 10^{-7}$. Several of these loci have been associated with BP phenotypes before: *FGF5*, associated here with SBP and MAP, has been associated with HTN previously; *KIF15*, associated with SBP here, has been associated with PP; *KCNK3*, here with SBP, with SBP, DBP, and MAP; *DOT1L*, here with DBP and MAP, with SBP and PP previously; and *PDILT*, here with HTN, has been associated with SBP, DBP, and HTN. *PLPPR2*, associated in this study with SBP and MAP, however, has not been, to our knowledge, been associated with BP phenotypes before. Little is known about the function of *PLPPR2*; it is predicted to be involved in phospholipid processing and signal transduction. In this analysis, both known and novel genetic loci associated with BP phenotypes were observed, suggesting that the genetic architecture of blood pressure and hypertension in individuals from Samoa and American Samoa is affected by both universal and unique genetic determinants. Further investigation and validation of these results will be necessary to determine the biological mechanisms of these associations in cardiovascular health.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1541 Genome-Wide Association Study (GWAS) of Sepsis in BioMe Cohort

Authors:

M. Stapylton[1,2], A. Biji[3,2], S. Cullina[4,2], E. Kenny[3,2,5,6], S. Asgari[3,2,5,6]; [1]Mount Sinai, New York, NY, [2]Inst. for Genomic Hlth., Icahn Sch. of Med. at Mount Sinai, New York, NY, [3]Icahn Sch. of Med. at Mount Sinai, New York, NY, [4]Icahn Sch. of Med. at Mt. Sinai, New York, NY, [5]Dept. of Med., Icahn Sch. of Med. at Mount Sinai, New York, NY, [6]Dept. of Genetics and Genomics, Icahn Sch. of Med. at Mount Sinai, New York, NY

Background: Sepsis is a life-threatening dysregulation of immune response to infections. Susceptibility to sepsis is partly explained by human genetic factors. Most previous genetic studies of sepsis have used relatively small patient populations with European ancestry, been poorly replicated and their relevance for non-European ancestry populations remains unclear. To fill this gap, we conduct a large, multi-ancestry genome-wide association study (GWAS) of sepsis that can identify the full spectrum of sepsis-associated genetic risk variants.
Methods: We identified 944 individuals in BioMe biobank (ages 18-95, median=62) with the sepsis ICD codes: A40, A41, R65.20, R65.21. Following imputation, we filtered out variants with a minor allele frequency (MAF) < 0.01, genotyping rate < 0.02, minor allele count (MAC) < 20, or Hardy-Weinberg p-value < 1e-8. Individuals with a genotyping rate < 0.1 were also excluded. BioMe participants were divided using self-reported ancestry into: Europeans N=8771, African-American/Africans N=7136, Hispanic/Latin-Americans N=10679 and Others N=4357. We then performed a GWAS of the sepsis phenotype in each subgroup using SAIGE after accounting for genetic-relatedness, population structure, age, sex, and 5 genotyping principal components. Lastly, we meta-analyzed results from each population along with results from UK Biobank (UKBB) A41 sepsis GWAS using the inverse-variance method implemented in METAL.
Results: In the meta-analysis, 242 variants reached the suggestive genome-wide association threshold (p < 10e-5). We did not observe any inflation in our meta-analysis results, (genomic inflation factor = 1.007). The odds ratio and MAF was consistent between BioMe Europeans and UKBB Europeans (1.19 (18%) vs 1.15 (18%)). The loci with the lowest p-value (p=6.5e-8) overlapped chr1:168951141 (1q24.2). This locus included four SNPs in high linkage disequilibrium (r2 > 0.7), with rs12030460 as the top associated SNP. All SNPs overlap the intronic region of LINC00970. Common SNPs in LINC00970 have been associated with several complex conditions with the strongest associations (p<10e-55) being venous thromboembolism and different clotting factors levels.
Conclusion: We ran the largest and first multi-ancestry GWAS of sepsis. At our current sample size, the locus with the strongest association overlaps LINC00970, a gene that has been implicated in coagulation disorders. Coagulation disorders are common in sepsis patients which may underlie the association between LINC00970 and sepsis risk. We are currently performing further investigation and replication of results in other biobanks to confirm this hypothesis.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1542 Genome-wide association study for metabolic syndrome reveals *APOA5* SNPs with multilayered effects in Koreans

Authors:

Y. Park[1,2], S. Moon[1], J. Choi[1], J. Kim[1], S-H. Im[3], H-Y. Son[1], J-I. Kim[1,2]; [1]Seoul Natl. Univ., Seoul, Korea, Republic of, [2]Med. Res. Ctr., Genomic Med. Inst., Seoul, Korea, Republic of, [3]Kangwon Natl. Univ., Gangwon, Korea, Republic of

Background and Purpose: Genome wide association studies(GWAS) regarding metabolic syndrome(MetS) have mainly been reported in non-Asian populations, and relatively small populations in Koreans. The interpretation of GWAS usually focuses on the top SNPs, but the specific explanation for other non-leading SNP signals is lacking. This study aimed to identify the interaction between rs2266788 and rs651821, major variants of apolipoprotein A5, and how they collectively influence plasma triglyceride concentrations. **Methods:** Based on the Korean Genome and Epidemiology Study cohort, a total of 58,600 Korean subjects with available biochemical information and demographic variables for MetS were included. These subjects were genotyped with the Korean Biobank Array and imputed with the Northeast Asian Reference Database 2 panel. We performed GWAS on MetS and its five diagnostic components (triglycerides, high-density-lipoprotein cholesterol, blood pressure, fasting blood glucose and waist circumference). **Results:** We conducted GWAS on MetS and its diagnostic components and discovered that APOA5 SNP rs651821 (risk allele C) was the top SNP in both MetS (p-value 1.21E-98, odds ratio (OR) 1.33) and triglyceride (p-value 2.80E-287, OR 1.67) phenotypes. rs2266788 (risk allele G, p-value 6.87E-113, OR 1.45) was another statistically significant SNP in APOA5, regarding the triglyceride phenotype. However, when a conditional analysis was done by rs651821, statistical significance was maintained but the OR direction was reversed. (Conditioned p-value 1.09E-26, Conditioned OR=0.76). Regarding SNP rs2266788, the G allele switched from the risk allele to the protective allele after conditioning. Rs2266788 was previously known to be a causal SNP with enhancer activities via the luciferase assay experiments. Thus, rs651821 and rs2266788 were independent signals but with opposite directions in the extended GWAS analysis. The conspicuous association pattern persisted in subgroup analyses regarding alcohol consumption, exercise frequency, and smoking status. **Conclusions:** Even though an allele might seem to be a risk allele, when the lead SNP's influence is removed, the particular SNP could portray an independent protective signal, indicating a multilayered effect of two SNPs. Both SNPs rs651821 and rs2266788 are independent SNPs with separate functional and enhancer activities. Hence, a new direction for future GWAS research was suggested.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1543 Genome-wide association study identifies candidate genetic variants associated with postoperative nausea and/or vomiting

Authors:

D. Nishizawa[1], R. Morino[2], M. Yokota[3,4], S. Ohka[1], S. Kasai[1], J. Hasegawa[1], Y. Ebata[1], K. Nakayama[1], K. Ikeda[1]; [1]Tokyo Metropolitan Inst. of Med. Sci., Tokyo, Japan, [2]Koujinkai Daiichi Hosp., Tokyo, Japan, [3]Cancer Inst. Hosp., Tokyo, Japan, [4]East Hokkaido Hosp., Hokkaido, Japan

Purpose: General anesthesia is commonly conducted in surgeries for the treatment of cancer or other diseases. Whereas postoperative nausea and/or vomiting (PONV) is known as the most frequent side effect following anesthesia, considerable individual differences have been widely observed in incidence of PONV. We conducted genome-wide association study (GWAS) in patients who underwent general anesthesia to identify potential genetic variants that may significantly contribute to individual differences in PONV. **Methods**: The subjects recruited were a total of 806 patients with written informed consent who underwent general anesthesia by total intravenous anesthesia (TIVA) with propofol and/or desflurane at The Cancer Institute Hospital of Japanese Foundation for Cancer Research (JFCR). The study protocol was approved by the Institutional Review Board at each related institute. Total genomic DNA was extracted from peripheral blood samples and genotyping was conducted using whole-genome genotyping arrays with more than 900,000 markers and iScan System (Illumina, San Diego, CA, USA). GWAS was conducted in the entire 806 patient subjects to identify single nucleotide polymorphisms (SNPs) associated with frequency of nausea and presence/absence of nausea, vomiting, and both of nausea and vomiting (PONV) in trend/additive, dominant, and recessive genetic models. Genome-wide associations were also investigated only in patients with propofol. **Results**: As a result of GWAS in all patients, several two SNPs, rs2776262 SNP and exm2274524 SNP in the *CNTN5* gene region were significantly associated with frequency of nausea in additive or recessive and dominant models, respectively ($P < 7.812 \times 10^{-8}$). In another GWAS only in patients with propofol, rs7212072 and rs12444143 SNPs in the *SHISA6* and *RBFOX1* gene regions, respectively, were significantly associated with frequency of nausea in additive model as well as the rs2776262 SNP in additive and recessive models, and rs1752136 SNP and exm1401859 SNP in the *ATP8B3* gene region were significantly associated with vomiting in trend model. **Conclusions**: The results indicate that the SNPs identified could serve as markers that predict PONV. Our findings will provide valuable information for achieving satisfactory PONV control.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1544 Genome-wide Association Study Identifies Genetic Variants for Blood Sugar Regulation in a Taiwan population.

Authors:

C-L. Tsai, W-C. Chang, M-R. Lin, W-H. Chou; Taipei Med. Univ., Taipei, Taiwan

Type 2 diabetes (T2D) is a metabolic disease characterized by chronically elevated blood sugar. In the past, genome-wide association study (GWAS) studies on diabetes in the Taiwanese population only involved comparisons between cases and controls. However, our study primarily focused on analyzing glycemic indicators as continuous phenotypes in GWAS analysis. We investigated whether potential genes could be identified that would affect the regulation of blood glucose in the Taiwanese population. We combined the single nucleotide polymorphisms (SNPs) from the Taiwan Biobank (TWB) chip with the SNPs from 1000G for imputation, and underwent QC filtering. Ultimately, we obtained a total of 3,636,444 SNPs. We performed GWAS to identify associations of these SNPs with fasting glucose and glycated hemoglobin (HbA1c) in 59,448 samples from TWB. The results showed that 39 SNPs were significantly associated with fasting glucose, and 35 SNPs with HbA1c. After conducting a meta-analysis and comparing our results with previously published GWAS literature on fasting glucose and HbA1c, we identified seven overlapping SNPs associated with fasting glucose and three SNPs associated with HbA1c. Through pathway enrichment analysis, we discovered that the significant SNPs we found collectively affected the pathway related to Maturity Onset Diabetes of the Young (MODY). This further confirmed the reliability of our GWAS results. Among all, we found four novel genome-wide significant fasting glucose-associated SNPs (*rs34874677*, *rs2074489*, *rs12922649* and *rs11650716*), and one novel HbA1c-associated SNP (*rs2074489*). This study presented that in the Taiwanese population, not only identified genes that were commonly involved in regulating blood glucose levels across different populations but also discovered novel genes specific to Taiwan could be associated with the modulation of blood glucose.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1545 Genome-wide association study identifies multiple novel loci associated with circulating metabolite phenotypes involved in energy metabolism and inflammation in type 2 diabetes: results from the Asian Indian Diabetic Heart Study (AIDHS)

Authors:

M. Rout[1], G. Tung[1], D. Sanghera[2]; [1]OUHSC, Oklahoma City, OK, [2]Univ. of Oklahoma Hlth.Sci. Ctr., Oklahoma City, OK

Despite the success of genome-wide association studies (GWAS), the genetic mechanisms that predispose people to type 2 diabetes (T2D) remain poorly understood. The high throughput metabolite profiling techniques have rapidly progressed from a single-gene/single-metabolite association to a genome-wide/metabolome-wide approach. However, limited data are available on non-white populations. Here we have performed global lipidomic profiling and a targeted genetic association study to understand the genetic influences of T2D-associated variants on circulating lipid metabolites using a T2D cohort of Punjabi Sikhs from India as part of the AIDHS. Untargeted lipidomic analysis was performed on blood samples of 3000 individuals (1723 T2D cases and 1277 controls) using liquid chromatography and high-resolution mass spectrophotometry (LC-MS). Of the 1863 identified analytes, including 417 known and 1446 unknown compounds, we identified 354 metabolites associated with T2D using multivariate mixed regression modeling. We further analyzed the association of 243 metabolites (with known structures) by conducting a genome-wide association study (GWAS) using genotyped and imputed at up to 14,959,825 million SNPs as part of discovery and replication cohorts. Our study identified 14 significant associations (p ranging from 3.1×10^{-8} to $3.5x10^{-18}$), including six robust signals in the genes involved in transient receptor potential (TRP) channel, and calcium signaling pathways with the metabolites connected with energy metabolism, fatty acid transport, inflammation, and glucose metabolism. No previous studies have reported the associations between these loci and metabolic traits. In our first metabolome-GWAS in Asian Indians, we have identified novel molecular signatures and pathways associated with T2D. If validated, these results can potentially contribute to developing new therapeutic strategies to improve the prediction and treatment of T2D.**Funding:** The Sikh Diabetes Study/ Asian Indian Diabetic Heart Study was supported by NIH grants R01DK082766 and R01DK118427 (NIDDK) and grants from the Presbyterian Health Foundation of Oklahoma.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1546 † Genome-wide association study of 30000 samples with bone mineral density in UK Biobank

Authors:

Y. Qian, H-f. Zheng; westlake Univ., Hangzhou, China

Osteoporosis, a prevalent aging-related disease, is characterized by decreased bone mineral density (BMD). Here, we conducted genome-wide association studies (GWAS) of BMD phenotypes at eleven skeletal sites, which were measured by dual-energy X-ray absorptiometry (the gold standard technology, referred to as DXA-BMD), in over 30,000 European individuals from the UK Biobank. We identified 91 unique and independent loci across 11 DXA-BMD phenotypes, including six novel loci and two sex-specific loci. By integrating multi-omics datasets, we prioritized six candidate genes for these novel loci (i.e., *ABCA1*, *CHSY1*, *CYP24A1*, *SWAP70*, *ZIC4*, and *PAX1*), providing insights into molecular mechanisms underlying osteoporosis. The annotated candidate genes of two sex-specific loci (i.e., *CYP19A1* and *CYP3A7*) were involved in steroid hormone biosynthesis. Our results demonstrated that the predictive accuracy of fracture risk using DXA-BMD phenotypic values or genetic risk scores (GRSs) outperformed those using BMD measured by quantitative ultrasound, suggesting the use of DXA assessment as a screening tool. We subsequently developed joint models that incorporated multiple GRSs, improving fracture genetic risk stratification and highlighting the utility of combining GWAS of the predicted trait with GWASs of relevant traits for enhancing genetic predictions of complex disorders. In the joint model, we identified eight DXA-BMD specific signals (five of which were novel), which were considered as novel candidate signals for fracture. Additionally, we uncovered shared polygenicity between head BMD and intracranial aneurysms (IA). Mapped genes of shared loci, especially for *SOX6* and *PLCE1* genes (with mixed effect direction on these two phenotypes), provide distinct genetic insights into the interplay between head BMD and IA. Finally, by integrating the druggable genome, gene expression, and GWAS datasets, we identified three genetically-supported drug targets (i.e., *SEBF1*, *CCR1*, and *NCOR1* druggable genes) as priority candidates for evaluation in osteoporosis treatment. This comprehensive investigation provides valuable information for understanding the genetic architecture of BMD and its relationship with aging-related diseases, as well as identifying potential therapeutic targets for osteoporosis.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1547 Genome-wide association study of apolipoprotein A and B and lipoprotein (a) in the China Health and Nutrition Survey

Authors:

C. Zhao, C. Spracklen; Univ. of Massachusetts, Amherst, MA

Lipids are important biomarkers and modifiable risk factors for cardiovascular diseases. The genetic determinants of total cholesterol (TC), high-density lipoprotein cholesterol (HDL), low-density lipoprotein cholesterol (LDL), and triglyceride (TG) levels have been extensively studied in genome-wide association studies (GWAS); however, GWAS for apolipoprotein A and B (ApoA and ApoB) and lipoprotein (a) [Lp(a)] are very limited and primarily investigated populations of European ancestry, limiting the discovery of genetic variants that impact lipid-level variations in other ancestral groups. Using imputed genotypes (1000 Genomes phase 3) from up to 7,632 East Asian-ancestry individuals from the China Health and Nutrition Survey (CHNS), we performed a GWAS of seven lipid traits, including ApoA, ApoB, and Lp(a). Analysis of ApoA identified three genome-wide significant (*P* &lt 5e-8; EPACTS) loci: *LIPC* (rs1077835, $P = 1.66x10^{-14}$), *CETP* (rs72786786, $P = 1.14x10^{-11}$), and *LIPG* (rs9953437, $P = 4.37x10^{-8}$). Sequential conditional analyses identified two additional signals at the *LIPC* locus (rs261291, $P_{cond} = 9.13x10^{-11}$; rs145857216, $P_{cond} = 8.72x10^{-6}$). Analysis of ApoB identified four significant associations near *PCSK9* (rs151193009, $P=1.24x10^{-9}$), *CELSR2* (rs611917, $P = 1.47x10^{-10}$), *LDLR* (rs147540853, $P = 3.18x0^{-10}$), and *APOE* (rs7412, $P = 3.42 \times 10^{-63}$). Sequential conditional analyses identified one additional signal at the *APOE* locus (rs429358, $P_{cond} = 3.55x10^{-10}$). Results were similar for models that included BMI as a covariate. While the ApoA and ApoB loci are well-established for the traditional lipid traits, 6 of the 10 index variants have not previously been associated with ApoA or ApoB. For Lp(a), the initial analysis identified two associated loci: *TOMM40/APOE* (rs61679753, $P = 5.43x10^{-14}$) and *LPA* (rs7770628, $P = 2.78x10^{-204}$). Conditional analyses at *LPA* identified fourteen distinct association signals, of which nine have not been previously reported for association with Lp(a). Within Biobank Japan, we observed a positive correlation between the effect estimates of our lead *LPA* variants and the risk for coronary artery disease ($r^2 = 0.77$, $P = 0.0019$), providing support for the relationship between Lp(a) and coronary artery disease. Our findings provide new insights into the genetic architecture of apolipoproteins and lipoproteins and further highlight the importance of conducting genetic studies across diverse ancestry populations.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1548 Genome-wide association study of blood pressure traits in 140 000 Mexican adults

Authors:

M. Turner[1], J. Alegre-Díaz[2], J. M. Torres[1], F. Rivas[2], D. Aguilar-Ramírez[1], R. Haynes[1], N. Staplin[1], M. Hill[1], R. Collins[1], J. Emberson[1], J. Berumen[2], P. Kuri-Morales[2], R. Tapia-Conyer[2]; [1]Univ. of Oxford, Oxford, United Kingdom, [2]Univ. Natl. Autónoma de México, Mexico City, Mexico

Background: Genome-wide association studies (GWAS) in diverse ancestries have elucidated the genetic architecture of complex traits and diseases, but <0.5% of participants in previous GWAS are of Latin American ancestry. Using data from the Mexico City Prospective Study (MCPS), we aimed to identify potentially-novel blood pressure-influencing genetic variants in an admixed Mexican population.
Methods: We ran GWAS of systolic blood pressure (SBP), diastolic blood pressure (DBP), mean arterial pressure (MAP) and pulse pressure (PP) in 140,559 MCPS participants genotyped with version 2 of the Illumina Global Screening Array. Participants on antihypertensive drugs (16%) had their SBP/DBP increased by 6/3mmHg. Variants with imputation r^2>0.4 and N_{eff}>30 were analysed using SUGEN adjusted for age, age^2, sex, body mass index (BMI), and ancestry principal components. Forwards-selection-backwards-elimination fine-mapping identified conditionally-independent signals. We assessed tissue-specific enrichment with GARFIELD and molecular epigenome annotations from the Common Metabolic Diseases Genome Atlas. Sensitivity analyses included antihypertensive adjustment of 15/10mmHg and no BMI adjustment.
Results: Associations were tested at 42,121,163 variants. Covariate-adjusted LD score regression intercepts were 1.03, 1.03, 1.04 and 0.99, and heritability estimates were 0.10, 0.08, 0.10 and 0.06, for SBP, DBP, MAP and PP respectively. Fine-mapping identified 68 independent signals at 63 loci associated with SBP, of which 15 were considered 'novel'. For DBP, MAP and PP these were, respectively, 53 signals at 51 loci (11 novel), 74 signals at 70 loci (28 novel) and 22 signals at 22 loci (1 novel). After false discovery rate correction, 7 novel SBP variants replicated in a combined meta-analysis of PAGE, pan-UKB and Biobank Japan, including rs1561477 at *ACOXL*, rs7213811 near *YWHAE*, and rs9606885 at *YWHAH-AS1*. Sensitivity analyses gave similar results, but the larger antihypertensive correction yielded additional associations. Novel loci had larger MAFs in Indigenous Mexican-ancestry DNA segments. GARFIELD showed enrichment in blood vessel, kidney and adrenal tissue.
Conclusions: This GWAS of a previously under-studied population offers additional insights into the genetics of blood pressure.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1549 Genome-wide Association Study of Diabetic Retinopathy Identified with an Electronic Health Record Algorithm in the Million Veteran Program

Authors:

J. Breeyear[1,2], A. Giri[3,2], S. L. Mitchell[3], J. N. Hellwege[3,2], B. Charest[4], A. Khakharia[5,4], C. W. Halladay[6], P. D. Reaven[7,8], J. B. Meigs[9,10,11], M. K. Rhee[4,5], Y. Sun[12], M. G. Lynch[4], C. L. Nealon[13], S. K. Iyengar[14,13], M. A. Brantley, Jr.[3], N. S. Peachey[15,13], Y. V. Sun[16,4], P. W. Wilson[4,5], L. S. Phillips[4,5], T. L. Edwards[3,2], VA Million Veteran Program; [1]Vanderbilt Univ., Nashville, TN, [2]VA Tennessee Valley Hlth.care System, Nashville, TN, [3]Vanderbilt Univ. Med. Ctr., Nashville, TN, [4]Atlanta VA Med. Ctr., Atlanta, GA, [5]Emory Univ., Atlanta, GA, [6]Providence VA Med. Ctr., Providence, RI, [7]Phoenix VA Hlth.Care System, Phoenix, AZ, [8]Univ. of Arizona, Phoenix, AZ, [9]Massachusetts Gen Hosp, Boston, MA, [10]Harvard Med. Sch., Boston, MA, [11]Broad Inst., Cambridge, MA, [12]Stanford Univ., Palo Alto, CA, [13]VA Northeast Ohio Hlth.care System, Cleveland, OH, [14]Case Western Reserve Univ., Cleveland, OH, [15]Cole Eye Inst., Cleveland Clinic, Cleveland, OH, [16]Emory Univ., Johns Creek, GA

Diabetic retinopathy (DR) is the leading cause of vision loss and preventable blindness in adults. The course of diabetes and severity of complications vary substantially among individuals, and this variability is not fully explained by known risk factors. While statistically significant loci have been reported from several GWAS of DR, few have been replicated. To address this gap in knowledge, we developed and validated an electronic health record (EHR) algorithm to identify DR cases and controls with diabetes in three independent EHR systems: Vanderbilt University Medical Center, VA Northeast Ohio Healthcare System, and Massachusetts General Brigham, with positive predictive values ranging from 0.840 - 0.953. We applied the DR algorithm in the Million Veteran Program, which identified European (38,944 cases | 69,416 controls) and African (17,221 cases | 26,607 controls) ancestry individuals, identified with the HARE method. GWASs of TOPMed imputed SNPs stratified by ancestry were conducted using logistic regression adjusting for age, sex, duration of diabetes, HbA1c, and first ten principal components. We identified genome-wide significant loci in the European analysis, *TCF7L2* (rs7903146, (OR = 1.12 (95% CI 1.04 - 1.30), $p = 8.57 \times 10^{-24}$)), and *L3MBTL1* (rs6130396, (OR = 0.94 (0.93 - 0.95)), $p = 7.41 \times 10^{-9}$). The African analysis identified one significant locus, *G6PD* (rs1050828 (V68M), (OR = 1.19 (1.14 - 1.24), $p = 7.14 \times 10^{-17}$)). Investigation of genetically predicted gene expression with S-PrediXcan, using GTEx v8 followed by colocalization (>0.85), identified two and one genes significantly associated with DR in the European (*TCF7L2* and *L3MBTL1*) and African (*PSMB4*) analyses, respectively. Based on results from the European analysis, we investigated the association between a published T2D PRS and DR (OR = 1.14 (1.04 - 1.25), $p < 0.05$, per SD increase in PRS) in a European ancestry cohort (905 cases | 1,120 controls), adjusted for age, sex, duration of diabetes, and HbA1c. Our results suggest that those with a higher genetic risk for diabetes are subsequently at a higher risk for DR. The identified *G6PD* missense variant has been previously shown to be under positive evolutionary selection due to a protective effect against *Plasmodium falciparum* infections (i.e., malaria). While a *G6PD* deficiency confers protection from malaria and risk for hemolytic anemia, these results suggest it increases risk for DR as well. With an EHR-based algorithm for DR and the resources of the Million Veteran Program, we were able to identify novel DR-SNP associations that suggest an adaptive variant increases risk for DR and may contribute to ancestral disparities.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1550 Genome-wide association study of Gulf War Illness reveals polygenic architecture and links to multiple health conditions

Authors:

G. Pathak[1], D. Koller[2], B. Cabrera-Mendoza[1], A. NonoDjotsa[3], F. Wendt[4], A. De Lillo[1], E. Friligkou[1], J. He[5], M. Kouakou[1], J. Vahey[6], L. Steele[7], R. Quaden[8], K. Harrington[9], S. Ahmed[10], J. Gaziano[11], J. Concato[12], H. Zhao[1], K. Radhakrishnan[13], J. Gelernter[14], E. Gifford[15], M. Aslan[16], D. Helmer[17], E. Hauser[18], R. Polimanti[19], Department of Veteran Affairs Cooperative Study Program (#2006); [1]Yale Univ., New Haven, CT, [2]Yale Sch. of Med., West Haven, CT, [3]Baylor Coll. of Med., Houston, TX, [4]Yale Sch. of Med., New Haven, CT, [5]Yale Univ. school of medicine, New Haven, CT, [6]Duke Univ., Durham, NC, [7]Dept. of Psychiatry and Behavioral Sci., Baylor Coll. of Med., Houston, TX, [8]Massachusetts Veterans Epidemiology Res. and Information Ctr. (MAVERIC), VA Boston Hlth.care System, Boston, MA, [9]Dept. of Psychiatry, Boston Univ. Sch. of Med., Boston, MA, [10]Dept. of Med., Baylor Coll. of Med., Houston, TX, [11]VA Boston Hlth.care System, Roxbury Crossing, MA, [12]VA Clinical Epidemiology Res. Ctr., VA Connecticut Hlth.care System, West Haven, CT, [13]Natl. Mental Hlth.and Substance Use Policy Lab., Substance Abuse and Mental Hlth.Services Admin., Bethesda, MD, [14]Yale Univ Sch. of Med., West Haven, CT, [15]VA Cooperative Studies Program Epidemiology Ctr.-Durham, Dept. of VA, Durham, NC, [16]VA Clinical Epidemiology Res. Ctr. (CERC), VA Connecticut Hlthcare System, West Haven, CT, [17]Ctr. for Innovations in Quality, Effectiveness, and Safety (IQuESt), Michael E. DeBakey VA Med. Ctr., Houston, TX, [18]Duke Univ Med Ctr, Durham, NC, [19]Yale Univ., West Haven, CT

Among veterans deployed to the Gulf War (1990-1991), Gulf War Illness (GWI) is a chronic multisymptomatic illness characterized by complex and uncertain causes and pathophysiology. To gain new insights into the disease's origin and understand its complex genetic factors, we conducted the first genome-wide association study (GWAS) on GWI in 33,523 veterans of diverse ancestries (African American, Admixed American, and European) from the Million Veteran Program. This study considered GWI definitions based on Kansas and CDC criteria, as recommended by the US Institute of Medicine.

We performed GWAS on four GWI case definitions: i) Kansas GWI symptom and exclusionary condition criteria [KS Sym+/Dx−], ii) Kansas GWI symptom only criteria [KS Sym+], iii) CDC GWI criteria, iv) CDC GWI severe criteria. Transcriptome-wide association study (TWAS) was conducted using PrediXcan and gene expression weights from 49 tissues in GTEx-v8. Polygenic scores (PGS) of 31 complex traits were tested with GWI definitions using PRS-CS. All associations were adjusted for age, sex, and the top ten principal components of within-ancestry population structure and survived the multiple testing threshold.

Overall, eight independent variants were identified ($p<5\times10\text{-}8$), with two of them supported by independent transcriptomic and phenome-wide analyses. Rs4675853 was associated with KS Sym+ and was found to be an expression quantitative trait locus (eQTL) for *AGXT*, *MAB21L4*, and *ATG4B* in several tissues. It has also been reported to be associated with sex hormone-binding globulin levels. Rs138168412, associated with KS Sym+, is a multi-tissue eQTL for *AOPEP* previously linked to respiratory function and physical strength. The TWAS identified five additional GWI-gene associations, including *CEMIP* in the cerebellum and *SNCG* in the adrenal gland. PGS analysis revealed pleiotropy between GWI and several traits, with the strongest evidence observed for type-2 diabetes (T2D), anxiety, and depression. While the PGS of T2D showed a stronger association with GWI in deployed veterans, PGS of anxiety and depression was more strongly associated with GWI in non-deployed veterans. No significant association was found between GWI candidate genes (*ACHE* rs1799805, *BCHE* rs1799807, and *PON1* rs662) and GWI in the single-variant, gene-based, and genetically regulated transcriptome analyses. This study provides the first comprehensive assessment of GWI's polygenic architecture, identifying several risk loci implicated in biological functions potentially relevant to GWI pathogenesis. The pleiotropy analysis suggests that genetic liability to GWI is likely shared with several health outcomes.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1551 Genome-wide association study of hip dysplasia in two Norwegian cohorts.

Authors:

K. Kvarme Jacobsen[1], S. Børthe[2,3,4], L. Bjerke Laborie[5,6], H. Kristiansen[1,6], T. Gundersen[5,6], K. Rosendahl[7,8], B. Winsvold[9,4], T. Zayats[10,11,3]; [1]Førde Central Hosp., Førde, Norway, [2]Oslo Univ. Hosp., Oslo, Norway, [3]Univ. of Oslo, Oslo, Norway, [4]Norwegian Univ. of Sci. and Technology, Trondheim, Norway, [5]Haukeland Univ. Hosp., Bergen, Norway, [6]Univ. of Bergen, Bergen, Norway, [7]Univ. Hosp. of Northern Norway, Tromsø, Norway, [8]Univ. of Tromsø, Tromsø, Norway, [9]Oslo Univ. Hosp. HF, Ullevål, Oslo, Oslo, Norway, [10]Massachusetts Gen. Hosp., Boston, MA, [11]Broad Insitute, Boston, MA

Introduction:Developmental dysplasia of the hip (DDH) is a congenital disorder, resulting in a shallow hip socket where the hip joint may be incongruent, subluxated or completely dislocated. If diagnosed in infancy, non-invasive treatment is available but with an increased risk of early osteoarthritis. Despite a known heritability of about 70 percent in twin studies, little is known about the genetic underpinnings of DDH. Here, we aim to investigate the genetic background of DDH to aid our understanding of the disorder and improve diagnostics and treatment strategies. Methods:We used two established cohort studies: The Norwegian Mother-Father-Child Study (MoBa), and the Trøndelag Health Study (HUNT). In MoBa, the phenotype of DDH diagnosis was derived from a parent-reported question “Does the child have hip dysplasia?”. The HUNT phenotype was based on ICD-9 and ICD-10 diagnoses of DDH and secondary osteoarthritis. The samples have undergone standard quality control and imputation. A total of 32,988 children were available for analyses from MoBa and 69,500 adults from HUNT. For MoBa, we used plink 2.0s glm model for genome wide association analysis, with a 1:4 ratio of cases and controls where excess controls were excluded at random. All analyses were adjusted for sex and the first five principal components. HUNT was analysed using SAIGE, with sex and the first 10 principal components as covariates. Meta-analysis was done in METAL software. Results: From MoBa, a total of 1,362,062 SNPs and 1,340 children were included after quality control and ratio adjustment , of which 268 children had DDH. No SNPs reached genome-wide significance in this initial analysis. The most significant finding was a SNP close to *XPNPEP1* with $P = 1.01 \times 10^{-5}$, a peptidase gene that is putatively involved in collagen assembly. In HUNT, a total of 8,531,386 SNPs were available, and 408 individuals had DDH or secondary osteoarthritis. Rs713162 near *COL11A1* reached genome-wide significance at $P = 8.4 \times 10^{-9}$. *COL11A1* encodes a collagen XI subunit. This finding was not replicated in MoBa, and in the meta-analysis, the P-value of rs713162 was no longer genome-wide significant, as there was a small but opposite direction of effect in MoBa. Discussion:We found a genome-wide significant associations between SNPs and DDH in the HUNT cohort study, that was not replicated in the smaller MoBa study. However, both top hits in the MoBa and HUNT studies point to collagen and its metabolism as important factors in DDH. As the full MoBa sample becomes available, we believe that the increased sample size and power will more likely result in detection of genome-wide significant associations in the MoBa cohort as well.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1552 Genome-wide association study of Long COVID in the Million Veteran Program

Authors:

A. Rodriguez[1], D. Posner[2], M. Murray[3], J. N. Minnier[4], F. Xia[1], N. J. Getty[1], L. Gao[5], T. Cai[6], S. Iyengar[7], S-W. Luoh[8], K. Cho[2], R. Madduri[1]; [1]Argonne Natl. Lab., Lemont, IL, [2]VA Boston Hlth.care System, Boston, MA, [3]VA Boston Hlth.care System, Boston, MA, [4]Dept. of Publ. Hlth.& Preventive Med., Oregon Hlth.& Sci. Univ., Portland, OR, [5]Oregon Hlth.and Sci. Univ., Portland, OR, [6]Dept. of Med., Brigham & Women's Hosp. and Harvard Med. Sch., Boston, MA, [7]Case Western Reserve Univ., Cleveland, OH, [8]VA Portland Hlth.Care System, Portland, OR

Long COVID, also known as post-acute sequelae of SARS-CoV-2 infection (PASC), is a debilitating condition that affects up to 43% of individuals recovering from acute COVID-19. To unravel the genetic basis of PASC, the Department of Energy conducted a genome-wide association study (GWAS) in collaboration with the Million Veteran Program (MVP), a diverse cohort study of US veterans with nearly 30% representation of non-European populations.

All controls and cases had a documented history of SARS-CoV-2 infection. Cases were defined using the International Classification of Diseases (ICD)-10 code (U09.9) for PASC. Controls were individuals who previously had acute COVID-19 but did not have PASC. MVP samples were genotyped on the Thermo Fisher MVP 1.0 Axiom array and imputed with the TOPMed reference panel (version R2). GWAS was conducted for each ancestry group using PLINK2 software for logistic regression, adjusting for age, sex, and the top 10 principal components of ancestry.

The cohort included 3,779 PASC cases and 68,411 controls from MVP participants in European American, Admixed American, African American and Hispanic American ancestral groups. Of 11 million variants, we identified 2 novel loci associated with PASC at genome-wide significance (p-value < 5x10^-8) in MVP and with PASC-related traits. One locus, significant only in participants of African ancestry, contained two intronic variants of *CDH23*, rs10999901 (OR=1.39, p=3.6x10^-8) and rs1868001 (OR=1.40, p=6.7x10^-9). The association between rs10999901 and PASC may be mediated by smoking (per C-allele effect on cigarettes smoked per day in an external European cohort = 0.56, p=7.8x10^-7) or other immune-related traits such as ratio of natural killer cells to lymphocytes (beta=0.23, p=1.3x10^-4). Dysfunction of natural killer cells has been implicated in the pathogenesis of PASC-linked diseases, most notably severe COVID-19 and myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS). The other locus was rs3759084, significant in Admixed American (AMR) ancestry participants (OR=1.34, p=9.7x10^-9). rs3759084 is an eQTL for *LIN7A* in blood, which also harbors an independent risk variant for severe COVID-19, rs10862191 (R^2 with rs3759084 in AMR = 0.06). The involvement of *LIN7A* in neurotransmitter secretion and other neuronal pathways may mediate the association between rs3759084 and PASC. In summary, we identified two novel loci for PASC in the MVP providing the first evidence for genetic involvement. Refinements using a validated PASC phenotypic algorithm using natural language processing and machine learning approaches, as well as additional genetic analyses are underway.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1553 Genome-wide association study of modulators of glucocerebrosidase.

Authors:

E. Somerville[1], L. Krohn[1], K. Senkevich[1], E. Yu[1], J. Ahmad[1], F. Asayesh[1], J. A. Ruskey[1], D. Spiegelman[1], S. Fahn[2], C. Waters[2], S. P. Sardi[3], R. N. Alcalay[2,4], Z. Gan-Or[1]; [1]Montreal Neurological Inst., McGill Univ., Montreal, QC, Canada, [2]Columbia Univ. Med. Ctr., New York, NY, [3]Sanofi, Framingham, MA, [4]Tel Aviv Sourasky Med. Ctr., Tel Aviv, Israel

Parkinson's disease (PD) is a debilitating neurodegenerative disorder characterized by alpha-synuclein deposits and a progressive loss of dopaminergic neurons in the substantia nigra. *GBA1*, responsible for producing the lysosomal hydrolase enzyme glucocerebrosidase (GCase), contains some of the most common pathogenic PD variants. Although deficiencies in GCase are associated with PD risk, not all individuals with low GCase activity carry *GBA1* mutations, suggesting that other genetic variants may be influencing enzyme activity. In the present study, we aimed to uncover additional loci modulating GCase by performing the first GWAS of GCase activity on two independent cohorts: 697 PD cases and 347 controls from Columbia University and 357 PD cases and 163 controls from the Parkinson's Progression Markers Initiative. Meta-analysis revealed the strongest signal to be the *GBA1* locus, as expected, driven by p.N370S (rs76763715, beta = -4.5 se = 0.3, p = 1.6E-41). A novel association in the *GAA* locus was also observed (17:78056851:C:T, beta = -1.0, se = 0.2, p = 1.4e-8). We further investigated this locus by creating an interaction plot of GCase and acid alpha-glucosidase (GAA) enzyme activities in Columbia, which supported the top *GAA* genotype having an effect on the correlation of the enzymes. Additionally, when adding *GAA* genotype and GAA enzyme activity interaction to the linear model for GCase activity, the interaction of homozygous status for the same variant with GAA activity was associated with decreased GCase activity (beta = -0.3, se = 0.1, p = 0.03). We then investigated various PD risk genes and found multiple to be potential modulators of GCase activity. We confirmed a previously reported association of *LRRK2* p.M1646T with GCase activity in our analyses (rs35303786, beta = 1.1, se = 0.4, p = 0.003), but did not find associations with other pathogenic *LRRK2* or *TMEM175* variants. Our results present a novel locus associated with GCase activity, and provide genetic targets to be considered when conducting research and clinical trials involving GCase. Given that GAA is also a lysosomal hydrolase enzyme, an interaction between GAA and GCase may support the idea that overall lysosomal homeostasis can be disrupted by enzyme activity deficiencies. More research will be required to confidently determine the mechanisms behind this relationship.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1554 Genome-wide association study of Myasthenia Gravis uncovers new loci and gives first insights into the performance of polygenic prediction.

Authors:

A. Braun[1], F. Stascheit[1], J. Kraft[1], G. M. Panagiotaropoulou[1], M. G. M. Huijbers[2], G. Kuhlenbäumer[3], F. Leypoldt[3], A. Meisel[4], S. Ripke[1]; [1]Charité – Univ.smedizin Berlin, Dept. of Psychiatry and Psychotherapy, Berlin, Germany, [2]Dept. of Human Genetics, Leiden Univ. Med. Ctr., Leiden, Netherlands, [3]Dept. of Neurology, Univ. Med. Ctr. Schleswig Holstein, Christian-Albrechts-Univ., Kiel, Germany, [4]Charité – Univ.smedizin Berlin, Dept. of Neurology with Experimental Neurology, Berlin, Germany

Background: Myasthenia Gravis (MG) is an autoimmune disease of the neuromuscular junction that affects 14-20 per 100,00 individuals and is characterized by fluctuating weakness of the voluntary muscles. It is frequently categorized by age of onset and autoantibody profiles, as variations in symptom spectrums have been observed between these groups. About 80% of cases test positive for antibodies against the nicotinic acetylcholine receptor. While some cases exhibit other antibodies directed against skeletal muscle receptors, a fraction of patients lack any previously described biomarkers.
Methods: We performed a genome-wide association study (GWAS) of MG cases and unaffected controls of European ancestry from 13 cohorts. Subsequently, we split the sample into early-onset (diagnosed under 50 years of age) and late-onset (diagnosed aged ≥50) and ran separate GWASs. We further conducted association and conditional analyses on imputed human leukocyte antigen (HLA) class I and II alleles for combined, early, and late-onset MG. Finally, we assessed the performance of polygenic risk scores (PRS) by splitting our sample into training and test datasets. PRS were constructed using the clumping and p-value thresholding method.
Results: The GWAS of 5,708 MG cases and 432,028 controls identified 9 genome-wide significant hits ($p<5.0e^{-08}$) and 24 index SNPs below a p-value of $1.0e^{-06}$. Onset-specific analyses revealed 4 additional loci significantly associated with early and 3 with late-onset MG. Imputation of HLA types identified *HLA-B*08:01* as the top risk-conferring allele in MG (OR=2.35, $p=1.15e^{-52}$, SE=0.06) and early-onset MG (OR=4.68, $p=2.18e^{-94}$, SE=0.08) and *HLA-DRB1*03:01* as the top overall allele in a late-onset sub-sample that implied a protective effect (OR=0.48, $p=2.37e^{-9}$, SE=0.12). We observed opposite directional effects for several associated HLA types between early- and late-onset MG subtypes. PRS significantly predicted MG in an independent unstratified target sample, explaining up to 4.21% ($p=5.12e^{-9}$) of variance between cases and controls at multiple p-value thresholds.
Discussion: To our best knowledge, this represents the largest GWAS of MG to date and the first to assess polygenic risk prediction, which has yielded promising first results. Notably, the associated HLA types are part of the 8.1 haplotype which has previously been linked to MG other autoimmune diseases. In this context, the opposite direction of HLA effects between onset subtypes could imply a moderating effect on the age of symptom onset. In the next step, we aim to further expand the sample size, conduct autoantibody-specific analyses, and include non-European samples.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1555 Genome-wide Association Study of Orofacial Clefts in Filipinos Identifies Novel Locus

Authors:

D. Maher[1], N. Mukhopadhyay[2], S. Curtis[3], J. Carlson[1], E. Feingold[4], S. Weinberg[4], C. Padilla[5], J. Murray[6], M. Marazita[7], E. Leslie[8], J. Shaffer[1]; [1]Univ. of Pittsburgh, Pittsburgh, PA, [2]Univ Pittsburgh/Sch Dental Med, Pittsburgh, PA, [3]Emory Univ., Tucker, GA, [4]Univ of Pittsburgh, Pittsburgh, PA, [5]Univ. of the Philippines Manila, Manila, Philippines, [6]U of Iowa, Iowa City, IA, [7]Univ Pittsburgh, Ctr. for Craniofacial and Dental Genetics, Pittsburgh, PA, [8]Emory Univ., Atlanta, GA

Cleft lip with or without cleft palate (CL/P) is among the most common congenital anomalies, with an estimated incidence of approximately 1 in 500 live births in the Philippines. The etiology of CL/P is complex, manifesting in both syndromic and non-syndromic forms, with the majority of cases being non-syndromic, i.e., occurring in the absence of any other signs or symptoms. Despite advances in understanding the genetic basis of CL/P, genes identified to date cumulatively only account for part of the trait heritability. Additionally, as is broadly the case for genetic research, non-European populations have been underrepresented, contributing to health disparities and leading to generalizability issues. We performed a genome-wide association study (GWAS) on 1,399 individuals (cases/controls: 882/517) recruited from the Philippines. The participants were genotyped on the Infinium Global Diversity Array-8v1.0 and imputation was conducted using the TOPMed reference. To mitigate the risk of confounding due to population structure, principal component analysis was conducted and the first five principal components, as well as sex, were included as covariates. SNPs were filtered for imputation quality (R^2 >= 0.8), and minor allele frequency of 1%. Quantile-quantile plots were assessed visually to rule out remaining genomic inflation. One genome-wide significant association was observed near the gene KCNQ5 (rs9446804, p=3.874e-08, minor allele frequency=23.7%). The frequency of the lead variant was similar as has been reported in East Asian populations, though the association of this locus with orofacial clefting has not been reported previously. Independent replication in a cohort of 419 Filipino trios with CL/P from the Gabriella Miller Kids First Research Program did not find that this specific SNP was associated (p = 0.76), but did find two other SNPs within KCNQ5 (>5kb upstream) that were suggestive (chr6:72714789:A:G: p = 0.00029; rs12208063: p = 0.00025). Associations with several known cleft loci (e.g., ARHGAP29, IRF6, SHROOM3) were observed at the suggestive (p<10e-06) threshold and some (e.g., IRF6) were replicated in the Filipino trios. Overall, these results suggest that Filipinos share many of the known common risk variants for CL/P with other populations and additionally have nominated a novel candidate gene for follow-up. This study also showcases the benefit of performing GWAS in diverse populations.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1556 Genome-wide association study reveals novel genetic signals that differentiates neuromyelitis optica spectrum disorder from multiple sclerosis

Authors:

S. Liu[1], E. Misicka[1], D. Crawford[1,2,3], F. Briggs[1,2]; [1]Dept. of Population and Quantitative Hlth.Sci., Case Western Reserve Univ. Sch. of Med., Cleveland, OH, [2]Cleveland Inst. for Computational Biology, Cleveland, OH, [3]Genetics and Genome Sci., Case Western Reserve Univ., Cleveland, OH

Neuromyelitis optica spectrum disorder (NMOSD) is a rare demyelinating autoimmune disease (AD) affecting the optic nerve and spinal cord, with a prevalence of <1/10,000 that varies across populations. In contrast to multiple sclerosis (MS), a relatively common AD of the central nervous disease, NMOSD exhibits unique immunopathological mechanisms suggesting distinct genetic architectures. Here, we conducted genome-wide association studies (GWAS) to identify variants associated exclusively with NMOSD. The study population included 337 NMOSD cases (68% AQP4-Ab+), 718 healthy controls (HCs), and 1,547 MS cases (82% were non-Hispanic White and of European ancestry). The following GWAS for 8.9 million SNPs were conducted adjusting for age at enrollment, sex, and 1st five multidimensional scaling components: 1) NMOSD vs HC+MS, 2) NMOSD vs HC, and 3) NMOSD vs MS. In NMOSD vs HC+MS, the most significant association was rs2858866 (OR=2.0, $p=8x10^{-15}$). Located between *HLA-DRB1* and *-DQA1*, rs2858866 is an eQTL ($p<5x10^{-8}$) for *HLA-DRB1*, *-DQA1*, & *-DRB5* across tissues, including the brain. Similar associations ($p<5x10^{-8}$) were observed across all GWAS comparisons. In the major histocompatibility complex (MHC), there were several other signals with consistent associations ($p<5x10^{-8}$) across GWAS. Outside the MHC, there were also promising suggestive associations on chromosomes 2 (rs10207044-*STAT4*: OR=1.9, $p=2x10^{-7}$) and 5 (rs79272149-*LINC02147*: OR=4.1, $p=2x10^{-7}$) in NMOSD vs HC+MS. Focusing on HLA alleles in those of European ancestry, *HLA-DQA1*05:01* (OR=1.9, $p=2x10^{-9}$) and *HLA-DRB1*03:01* (OR=2.1, $p=2x10^{-8}$) were the most significant associations in NMOSD vs HC+MS and they were genome-wide significant in nearly all other comparisons. Similar results were observed when restricting to AQP4-Ab+ cases. In summary, this is the largest NMOSD sample investigated in a GWAS, including the largest study of AQP4-Ab+ cases. This is also the first GWAS to contrast NMOSD to MS, which enhanced the power to detect novel NMOSD-specific associations for this rare disease. Further investigation of the identified genetic variants targeting their functions and a more sophisticated characterization of the MHC region will offer valuable insights into the mechanisms driving NMOSD pathogenesis, and such analyses are ongoing (i.e., genetic instrument variables and metabolomic/proteomic QTL analyses). Our findings contribute novel insights into the complex genetic architecture underlying NMOSD and shed light on the distinct genetic underpinnings that differentiate NMOSD from MS.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1557 Genome-wide association, colocalization, and Mendelian randomization analyses of the three complement system activation pathways in a general population sample

Authors:

C. Pattaro[1], D. Noce[1,2], L. Foco[1], D. Orth-Höller[2,3], E. König[1], G. Barbieri[1,4], M. Pietzner[5,6], S. Coassin[7], C. Fuchsberger[1], M. Gögele[1], F. Del Greco M[1], A. De Grandi[1], M. Summerer[7], E. Wheeler[6], C. Langenberg[5], C. Lass-Flörl[2], P. P. Pramstaller[1], F. Kronenberg[7], R. Würzner[2]; [1]Inst. for Biomedicine (affiliated to the Univ. of Lübeck), Eurac Res., Bolzano, Italy, [2]Inst. of Hygiene & Med. Microbiol., Dept. of Hygiene, Microbiol. and Publ. Hlth., Med. Univ. of Innsbruck, Innsbruck, Austria, [3]MB-LAB – Clinical Microbiol. Lab., Innsbruck, Austria, [4]Dept. of NeuroSci.s, Biomedicine and Movement Sci., Univ. of Verona, Verona, Italy, [5]Computational Med., Berlin Inst. of Hlth.(BIH) at Charité – Univ.smedizin Berlin, Berlin, Germany, [6]MRC Epidemiology Unit, Univ. of Cambridge, Cambridge, United Kingdom, [7]Inst. of Genetic Epidemiology, Med. Univ. of Innsbruck, Innsbruck, Austria

The complement system (CS) is a fundamental part of the innate immune response. Alterations of CS activation are associated with several human diseases of both syndromic and complex nature. CS activation occurs via three distinct pathways: the classical pathway (CP), the mannose-binding lectin pathway (LP) and the alternative pathway (AP). Given no hypothesis-free genome-wide screen of the three CS pathways has been conducted so far, to identify the CS's genetic basis, we conducted genome-wide association studies of the functional activity of CP, LP, and AP in the Cooperative Health Research in South Tyrol (CHRIS) study (n=4990). We identified 7 loci, including 13 independent and pathway-specific variants (P-value<5e-08) located in or near *CFHR4*, *C7*, *C2*, and *MBL2* (known CS genes) and *PDE3A*, *TNXB*, and *ABO* (novel genes).Variants were associated with inflammatory, autoimmune and coagulation disorders and >400 proteins. We conducted transcriptome- and proteome-wide colocalization analyses based on state-of-the-art datasets, in combination with two-sample Mendelian randomization analysis. We identified three types of results: (1) confirmation of known causal pathways (e.g.: causal role of MBL2 on LP); (2) identification of within-CS feedback loops (e.g.: between AP and complement 7); and (3) identification of novel causal pathways, including: the causal role of ABO protein levels on LP (P-value=1.1e-10; MR-Egger intercept not significant); a causal effect of LP on collectin-11 (P-value=6.3e-44; heterogeneity P-value, P_{het}=0.46) and KAAG1 (P-value=9.0e-25; P_{het}=0.27) levels; a causal effect of LP on mouth ulcers' risk (P-value=9.5e-6; P_{het}=0.71). Overall, these results depict a first, comprehensive and unbiased map of the role of CS on human health.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1558 Genome-wide associations of multi-organ biological age gaps in humans

Authors:

J. WEN[1], Y. Tian[2], I. Skampardoni[1], Z. Yang[1], E. Mamourian[1], A. Toga[3], A. Zaleskey[2], C. Davatzikos[1]; [1]Univ. of Pennsylvania, philadelphia, PA, [2]The Univ. of Melbourne, Melbourne, Australia, [3]Univ. of Southern California, Los Angeles, CA

Introduction: Human aging is evident across multiple organ systems. Machine learning (ML) can be used to determine an individual-level phenotype known as the biological age gap (BAG), which reflects the difference between an individual's ML-derived age and chronological age. To comprehensively depict the genetic heterogeneity of the BAG within nine human organ systems, we performed genome-wide association studies (GWAS) utilizing a large cohort of 30,108 to 111,543 Caucasian participants from the UK Biobank study. **Methods**: We used a linear support vector regressor (SVR) and a 20-fold cross-validation to derive the nine BAGs. We then ran a linear regression using Plink for the brain, cardiovascular, eye, hepatic, immune, metabolic, musculoskeletal, pulmonary, and renal BAG, controlling for confounders of age, dataset status (training/validation/test or independent test dataset), age x squared, sex, age x sex interaction, age-squared x sex interaction, and the first 40 genetic principal components. Additional covariates for total intracranial volume and the brain position in the scanner were included for the brain BAG. We adopted the genome-wide P-value threshold (5x10-8) to determine significant signals. FUMA was used to define the genomic loci considering linkage disequilibrium. **Results**: GWAS identified 11, 44, 17, 41, 61, 76, 24, 67, and 52 genomic loci linked to the brain, cardiovascular, eye, hepatic, immune, metabolic, musculoskeletal, pulmonary, and renal BAGs, respectively. 143 loci are novel - their top lead SNP was never associated with any clinical traits in the EMBL-EBI GWAS Catalog. Specifically, 7, 18, 7, 11, 27, 14, 11, 26, and 22 novel loci were found for the abovementioned nine BAGs. As illustrative examples, we showed the results from two loci. First, we observed that a well-known genetic risk factor (APOE4) for sporadic Alzheimer's disease, located on chromosome 19 (rs429358), was associated with the hepatic BAG. Furthermore, we discovered a novel locus on chromosome 1 (rs76560665) that was associated with the brain BAG. **Conclusions**: In the present GWAS, we identified 394 genomic loci linked to the nine BAGs. These findings provide valuable insights into the genetic architecture underlying the biological age of human organ systems. Further investigations are warranted to elucidate the genetic correlations between the nine BAGs and other clinical traits, particularly chronic diseases.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1559 Genome-wide meta-analyses of cognitive decline across neurocognitive domains in older adults stratified by *APOE* genotype.

Authors:

V. Acharya[1], K-H. Fan[1], B. E. Snitz[2], M. Ganguli[3], S. T. DeKosky[4], O. Lopez[2], E. Feingold[1], M. Kamboh[1]; [1]Dept. of Human Genetics, Univ. of Pittsburgh Sch. of Publ. Hlth., Pittsburgh, PA, [2]Dept. of Neurology, Sch. of Med., Univ. of Pittsburgh, Pittsburgh, PA, [3]Dept. of Psychiatry, Sch. of Med., Univ. of Pittsburgh, Pittsburgh, PA, [4]McKnight Brain Inst. and Dept. of Neurology, Coll. of Med., Univ. of Florida, Gainesville, FL

Cognitive decline has a substantial heritability as shown by twin and population-based studies. *APOE*, one of the robust genetic risk factor for AD, has also been associated with cognitive decline, especially with the decline of memory, executive function, and global cognitive function. *APOE* genotype stratified analysis can help to identify other genetic loci which might be masked due to strong effect of *APOE* 4. Hence, we conducted *APOE* genotype stratified (*APOE*23/22, *APOE*33, *APOE*34/44) genome-wide meta-analyses using slopes across five cognitive domains (attention, language, executive function, memory, and visuospatial function) and global cognitive function . We derived domain slopes for ~3,000 individuals aged 65 and above coming from three longitudinal cohorts: Gingko Evaluation of Memory (GEM), Monongahela-Youghiogheny Heathy Aging Team (MYHAT), and Monongahela Valley Independent Elders Survey (MoVIES). A linear mixed effect model was used to determine the individual cognitive domain change across time by using baseline age, education, and sex as covariates. *APOE* genotype stratified genome-wide association analyses were then conducted using common and low frequency variants (≤1 % minor allele frequency) for each cognitive phenotype adjusting for four genetic principal components in each cohort and later meta-analyzed using inverse-variance method in METAL software. In the *APOE* 33 genotype group, we identified a novel genome-wide significant (GWS) signal for decline of global cognitive function on Chr6q12 (β= -0.507; P=1.44E-09). The same signal was also observed at a nominal significance in the *APOE* 23/22 group, but in the opposite direction (β= 0.57; P= 2.75E-03). A previously described novel signal for decline of attention on Chr9q21.32 in this sample without *APOE* genotype stratification (PMID: 37089073) was found to be confined in the *APOE* 33 group at GWS level (β= -0.288; P=9.95E-09). In the *APOE* 34/44 individuals, we observed a second novel signal for decline of language on Chr5q23.1 (β= 0.693; P=3.79E-08). In addition, we observed three sub-threshold GWS signals, including one for decline of executive function on Chr8p23.2 (β= 0.32; P= 8.49E-08) in the *APOE* 34/44 group, and for the decline of attention and language on Chr2q12.1 (β= 0.86; P=8.64E-08) and Chr5q22.3 (β= 1.41; P=5.66E-08) respectively in the *APOE* 23/22 individuals. In conclusion, the *APOE* genotype stratified GWAS analyses have enabled us to identify additional novel signals associated with cognitive decline in older adults. Our work provides improved understanding of the genetic architecture of cognitive aging.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1560 Genome-wide meta-analysis identifies novel risk loci for uterine fibroids within and across multiple ancestry groups

Authors:

J. Kim[1,2,3], A. Williams[4], H. Noh[5], E. A. Jasper[3], S. H. Jones[6], E. A. Ruiz-Narváez[7], L. A. Wise[8], J. Palmer[8], J. Connolly[9], A. Khan[10], M. Abbass[11], L. Rasmussen-Torvik[11], L. Kottyan[12], W-Q. Wei[13], T. L. Edwards[2,14], D. R. Velez Edwards[2,3,13], J. Hellwege[2,15]; [1]Vanderbilt Univ. Sch. of Med., Nashville, TN, [2]Vanderbilt Genetics Inst., Vanderbilt Univ., Nashville, TN, [3]Div. of Quantitative Sci., Dept. of Obstetrics and Gynecology, Vanderbilt Univ. Med. Ctr., Nashville, TN, [4]NIH, Bethesda, MD, [5]Med., Hlth.and Society, Vanderbilt Univ., Nashville, TN, [6]Inst. for Med. and Publ. Hlth., Vanderbilt Univ. Med. Ctr., Nashville, TN, [7]Dept. of Nutritional Sci., Univ. of Michigan Sch. of Publ. Hlth., Ann Arbor, MI, [8]Dept. of Epidemiology, Boston Univ. Sch. Publ. Hlth., Boston, MA, [9]Ctr. for Applied Genomics, Dept. of Pediatrics, Children's Hosp. of Philadelphia, Philadelphia, PA, [10]Dept. of Med., Div. of Nephrology, Columbia Univ., Coll. of Physicians & Surgeons, Dept. of Surgery, New York, NY, [11]Northwestern Univ. Feinberg Sch. of Med., Chicago, IL, [12]Ctr. for Autoimmune Genomics and Etiology, Cincinnati Children's Hosp. Med. Ctr., Dept. of Pediatrics, Univ. of Cincinnati, Cincinnati, OH, [13]Dept. of BioMed. Informatics, Vanderbilt Univ. Med. Ctr., Nashville, TN, [14]Div. of Epidemiology, Dept. of Med., Vanderbilt Univ. Med. Ctr., Nashville, TN, [15]Div. of Genetic Med., Dept. of Med., Vanderbilt Univ. Med. Ctr., Nashville, TN

Uterine leiomyomata (fibroids) are common benign tumors of the uterus with poorly understood etiology. The prevalence of fibroids ranges between 20 to 80% across reproductive ages. Cost estimates for the United States due to fibroids range from $5.9 to $34.4 billion.

Previous genome-wide association studies (GWAS) have reported 72 associated loci but featured limited sample sizes for non-European populations. Our objective of this study is to identify novel genetic variants associated with fibroids across and within ancestry groups.

We conducted a meta-analysis of fibroid GWAS summary statistics from adult female participants with 74,294 cases and 465,810 controls across European (EUR), African (AFR), East Asian (EAS), and Central South Asian (CSA) ancestry groups from eight datasets. We meta-analyzed the data using METAL: cross-ancestry (all data), EUR, AFR, and EAS/CSA ancestry. Gene and pathway annotation used Functional Mapping and Annotation (FUMA). S-PrediXcan was used to evaluate associations with predicted gene expression in all GTEx v8 tissues, including the uterus.

We identified 371 sentinel SNPs, including 9 novel loci and 15 loci not previously discussed in literature but detected in publicly available biobank summary statistics. Functional analysis identified significant gene-set tissue enrichment in the uterus, cervix, esophagus, fallopian tube, ovary, bladder, and sigmoid colon tissues. These genes were also enriched in DNA damage and cell cycle biological pathways.

The predicted expression of 568 gene-tissue pairs at 180 unique genes were significantly associated with fibroids. Of those, 131 were previously unreported gene associations with fibroids. Within uterine tissue analyses, we observed six significant novel gene associations. There was an association with increased risk for fibroids with increased expression of: *SHMT1* (Odds Ratio (OR) = 1.24, p = 6.1e-7), *RPS26* (OR = 1.04, p = 1.4e-6), and *CD59* (OR = 1.14, p = 1.4e-5). There was an association with decreased risk for fibroids with decreased expression of: *SULT1E1* (OR = 0.72, p = 5.2e-18), *TSGA10* (OR = 0.90, p = 2.9e-10), and *SLC25A17* (OR = 0.84, p = 5.0e-6).

We identified consistent and unique associations at SNPs across populations and tissues in predicted uterine gene expression. There was enrichment in both hormone responsive tissues and tumorigenesis-related pathways. These new genetic loci and uterine expression factors may provide translational opportunities for novel fibroid treatments.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1561 Genome-wide meta-analysis of*orbicularis oris*muscle defects as a subclinical phenotype in orofacial clefting.

Authors:

J. Chernus[1], G. E. Mitrick[1], M. Lee[2], N. Mukhopadhyay[2], J. C. Carlson[3], A. Butali[4], C. J. Buxó-Martinez[5], K. Christensen[6], J. T. Hecht[7], L. M. Moreno Uribe[8], C. D. Padilla[9,10], F. A. Poletta[11,12], G. L. Wehby[13], E. Feingold[1,3], J. R. Shaffer[1], E. J. Leslie[14], J. C. Murray[15], S. M. Weinberg[1,2,16], M. L. Marazita[1,2,17]; [1]Dept. of Human Genetics, Sch. of Publ. Hlth., Univ. of Pittsburgh, Pittsburgh, PA, [2]Ctr. for Craniofacial and Dental Genetics, Dept. of Oral and Craniofacial Sci., Sch. of Dental Med., Univ. of Pittsburgh, Pittsburgh, Pittsburgh, PA, [3]Dept. of Biostatistics, Sch. of Publ. Hlth., Univ. of Pittsburgh, Pittsburgh, PA, [4]Dept. of Oral Pathology, Radiology and Med., Dows Inst. for Dental Res., Coll. of Dentistry, Univ. of Iowa, Coralville, IA, [5]Dental and Craniofacial Genomics Core, Sch. of Dental Med., Univ. of Puerto Rico, San Juan, PR, [6]Unit of Epidemiology, Dept. of Publ. Hlth., Univ. of Southern Denmark, Copenhagen, Denmark, [7]Dept. of Pediatrics, McGovern Med. Sch. and Sch. of Dentistry UT Hlth.at Houston, Houston, TX, [8]Dept. of Orthodontics, Coll. of Dentistry, Univ. of Iowa, Iowa City, IA, [9]Dept. of Pediatrics, Coll. of Med.; and Inst. of Human Genetics, NIH, Univ. of the Philippines Manila, Manila, Philippines, [10]Philippine Genome Ctr., Univ. of the Philippines System, Quezon City, Philippines, [11]ECLAMC (Latin American Collaborative Study of Congenital Malformations) at INAGEMP (Natl. Inst. of Population Med. Genetics), Porto Alegre, Brazil, [12]CEMIC: Ctr. for Med. Ed. and Clinical Res. at CONICET (Natl. Res. Council of Argentina), Buenos Aires, Argentina, [13]Dept. of Hlth.Management and Policy, Coll. of Publ. Hlth., Univ. of Iowa, Iowa City, IA, [14]Dept. of Human Genetics, Emory Univ. Sch. of Med., Emory Univ., Atlanta, GA, [15]Dept. of Pediatrics, Carver Coll. of Med., Univ. of Iowa, Iowa City, IA, [16]Dept. of Anthropology, Univ. of Pittsburgh, Pittsburgh, PA, [17]Clinical and Translational Sci., Sch. of Med., Univ. of Pittsburgh, Pittsburgh, PA

Orofacial clefts, notably clefts of the lip with or without the palate (CL/P), carry a large public health burden, ranking among the most common congenital abnormalities worldwide. Understanding the etiology of CL/P is challenging, because causality is multifactorial - with numerous genetic and environmental influences - and because phenotypes are extremely heterogeneous. Thus while genome-wide association studies (GWAS) have led to functional validation of risk loci, genetic architecture remains poorly understood.

The “subclinical phenotype” paradigm, an epidemiologically supported hypothesis, suggests that cleft lip may exist on a spectrum of severity ranging from subtle, ultrasound-detectable defects in the lip’s orbicularis oris muscle (OOM) to overt clefts. If true, OOM defects and cleft lip may share genetic and non-genetic risk factors. Therefore studying the genetics of OOM defects will improve our understanding of CL/P genetics.

Here we performed the first GWAS and meta-analysis of OOM defects in two cohorts: One GWAS cohort consisted of the individuals without CL/P but with affected close relatives (n≈3800), a cohort expected to be enriched for CL/P risk factors; another consisted of the individuals with no CL/P and no family history (n≈3400). By systematically comparing OOM defect rates, genetic associations, and their overlap with known clefting risk loci, we aimed to clarify the genetic relationship between OOM defects and CL/P.

While our study design does not enable us to formally test whether OOM defects are mild expressions of the processes leading to CL/P, we here interpret our results in the context of the subclinical phenotype framework. There are similar rates of OOM defects observed in families with and without history of CL/P (4.49% and 4.98% respectively, p=0.35, $\chi 2$=0.86). Moreover, there is limited overlap in OOM GWAS signals across the two groups, and also limited overlap between OOM and CL/P GWAS signals. These results appear most consistent with a conclusion OOM defects and CL/P are etiologically distinct phenotypes.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1562 Genome-wide polygenic score for inflammatory bowel disease is associated with diarrheal outcomes in Bangladeshi infants

Authors:

R. Munday[1], R. Haque[2], P. Korpe[3], U. Nayak[4], B. D. Kirkpatrick[5], W. A. Petri, Jr.[4], G. L. Wojcik[3], P. Duggal[3]; [1]Johns Hopkins Sch. of Med., Baltimore, MD, [2]Intl. Ctr. for Diarrhoeal Disease Res., Bangladesh, Dhaka, Bangladesh, [3]Johns Hopkins Bloomberg Sch. of Publ. Hlth., Baltimore, MD, [4]Univ. of Virginia Sch. of Med., Charlottesville, VA, [5]Univ. of Vermont Coll. of Med. and Vaccine Testing Ctr., Burlington, VT

Diarrhea is the second leading cause of death among children under 5 years of age worldwide. In addition to mortality, diarrhea also contributes to malnutrition, cognitive deficits, and linear growth faltering. In previous genome-wide association studies (GWAS) of host genetic susceptibility to enteric pathogens including *Cryptosporidium*, *Campylobacter*, and *Entamoeba histolytica*, we identified shared risk variants with inflammatory bowel disease (IBD) that were different for each pathogen suggesting a relationship between infection and autoimmunity. Children in these cohorts reported an average 3.8 independent episodes of diarrhea (range 0-19) per child and a median of 14 days of diarrhea (range 0-117) per child in the first year of life. The strong environmental risk of diarrheal infection from poor sanitation and overcrowding provides high rates of diarrhea, which represent the extreme cases of diarrhea not present among children in other studies. For this study, we aimed to identify a shared genetic architecture between inflammatory bowel disease and diarrheal outcomes. We included children in the first year of life from three independent birth cohorts in Dhaka, Bangladesh: the Dhaka Birth Cohort, the Performance of Rotavirus and Oral Polio Vaccines in Developing Countries study, and the Cryptosporidiosis Birth Cohort. All three studies prospectively enrolled children from a poor, urban community in Dhaka, Bangladesh and followed them biweekly for diarrhea episodes, infection, and general health. We used 5.37 million SNPs from a publicly available genome-wide polygenic score for inflammatory bowel disease (PGS000017) to derive a score for each participant. We performed logistic regression with 2 severe diarrheal phenotypes, each spanning the first year of life adjusted for sex, genotyping array, and the first principal component: (1) no diarrhea vs 6 or more episodes of diarrhea, indicating frequency, and (2) no diarrhea vs 25 or more days of diarrhea, indicating duration. In the analysis of diarrheal episodes, as the polygenic score increased, individuals were more likely to be cases (per SD OR = 1.31, 95% CI 1.02-1.68, p = 0.03). Similarly, in the analysis of days of diarrhea, the IBD polygenic score was associated with case status (per SD OR = 1.33, 95% CI 1.02-1.72, p = 0.03). These findings provide evidence for a link between the pathogenesis of inflammatory bowel disease and extreme diarrhea in early childhood, both in terms of frequency and duration, and may provide insights into preventative or therapeutic approaches.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1563 Genome-wide study identifies sex-specific risk loci for orofacial cleft disorders

Authors:

K. Kanchan[1], S. Berke[2], J. Ngwa[2], D. Ray[2], C. L. Simpson[3], J. Bailey-Wilson[4], A. F. Scott[5], L. Moreno[6], A. Butali[6], E. J. Leslie[7], S. M. Weinberg[8], T. Beaty[2], M. L. Marazita[8], I. Ruczinski[9]; [1]Johns Hopkins Sch. of Med., Baltimore, MD, [2]Johns Hopkins Bloomberg Sch. of Publ. Hlth., Baltimore, MD, [3]Univ. of Tennessee Hlth.Sci. Ctr., Memphis, TN, [4]Natl. Human Genome Res. Inst., Baltimore, MD, [5]Johns Hopkins Univ. Sch. of Med., Baltimore, MD, [6]Univ. of Iowa, Iowa City, IA, [7]Emory Univ., Atlanta, GA, [8]Univ. of Pittsburgh, Pittsburgh, PA, [9]Johns Hopkins Bloomberg Sch. of Publ. Hlth., BALTIMORE, MD

Orofacial clefts (OFCs) are the most common congenital craniofacial anomalies and the second most common birth defect worldwide. OFCs can be categorized into non-syndromic malformations, which occur in isolation, and syndromic malformations, which are associated with another malformation or as part of a recognized malformation syndrome. Non-syndromic OFCs, including cleft lip, cleft palate, and cleft lip with cleft palate vary between the sexes and birth prevalence rates differ across populations. To elucidate the underlying mechanisms for differential risk to OFCs between the sexes we analyzed whole genome sequences from case-parent trios generated through the Gabriella Miller Kids First (GMKF) Consortium. Our study comprised a total of 890 OFC trios from multiple ethnicities (European (N=371), Latin American (N=265), African (N=130), and Asian (N=124)). We performed genotypic transmission disequilibrium tests (gTDTs) using our in-house Bioconductor package 'trio' to test for the interactions between autosomal SNPs and sex within each ethnic group, followed by a fixed-effects inverse-variance weighted meta-analysis using METAL. We identified one genome-wide significant locus on chromosome 3 where female carriers of variant alleles had a significantly lower OFC risk than male carriers ($p=6.1x10^{-11}$). This chromosome 3 SNP (rs74520055) has an average minor allele frequency (MAF) of 11% and maps to the intronic region of *RFTN1* gene. Its RegulomeDB score of 3a and its chromatin state assignment of 4 show biological evidence as a regulatory element and is associated with a transcription site. In addition, the SNP is also a chromatin signature of three protein-coding genes *BTD, HACL1,* and *ANKRD28,* two long intergenic non-coding RNA genes *LINC00690* and *RP11-194G10.3* and one miRNA gene *MIR563*. We further identified 8 more candidate loci on chromosomes 2, 5, 8, 9, and 10, with average MAF above 9%, where sex-specific effects achieved p-values less than $1x10^{-5}$: rs11693354 (*AGAP1*), rs28799100 (*TMEM161B*), rs1924005650 (*EIF3KP1*), rs2588209 (*SLC7A2*), rs55663336 (*RP11-586K2.1*), rs2456245 (*RP11-122C21.1*), rs78756981 (*CNTFR*), and rs11258513 (*FRMD4A*). Our inference corroborates the finding from an independent case-control study, which previously reported suggestive evidence for *RFTN1* as a sex-specific effect on risk toward non-syndromic OFCs.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1564 Genome-wide survival study identifies three novel assocations between non-HLA donor-recipient mismatches and kidney graft failure

Authors:

V. Mauduit[1], A. Durand[1], M. Morin[1], O. Rousseau[1], P-A. Gourraud[1], N. Vince[2], G. Blancho[3], M. Giral[3], G. Jean[4], M. Ribatet[5], S. Limou[1]; [1]CR2TI, Nantes, France, [2]Inserm, Nantes Université, Nantes Cedex 01, France, [3]CHU de Nantes, Nantes, France, [4]LS2N, Nantes, France, [5]Ecole Centrale de Nantes, Nantes, France

***Purpose**: Kidney transplantation is the best treatment of end stage kidney disease, a pathology affecting a growing proportion of the population. Donor-recipient (D-R) mismatches in HLA genes have been associated with a poorer kidney graft survival. However, HLA mismatches alone do not explain long-term graft function decline. We ran a genome-wide survival study (GWSS) on a large monocentric cohort of kidney transplant pairs in order to characterize D-R mismatches associated with kidney graft loss beyond HLA.
***Methods**: The KiT-GENIE genetic cohort comprises 1482 complete D-R pairs of European ancestry for kidney transplants performed in Nantes since 2000. After imputation with the TopMed reference panel, we performed a genetic association study on >8.2M polymorphism (SNP) mismatches for time-to-kidney graft failure (defined as return to dialysis or preemptive retransplantation). A D-R SNP mismatch was defined by the presence of a donor allele not shared by the recipient. We then tested association between mismatches and time-to-graft failure using multivariate Cox proportional hazards models adjusted for HLA mismatches. P-values below the multiple testing Bonferroni correction threshold (5x10-8) were considered significant.
***Results**: We identified three statistically significant associations: one chromosome 10 intronic mismatch near the promotor of a gene encoding a protein with structural similarities to complement component C9 (p=3.1x10-8, HR=3.5) , a chromosome 17 intergenic mismatch (p=1.0x10-8, HR=4.0) near a gene highly expressed in kidneys and playing a role in signal transduction and protein transport, and a chromosome 21 intronic mismatch of a gene encoding a type I membrane protein expressed in kidney (p=2.0x10-8, HR=5.2).
***Conclusions**: Our genome-wide analysis in a large homogeneous monocentric cohort revealed three novel non-HLA D-R mismatches associated with kidney graft failure. Further validation in external cohorts and non-European pairs is ongoing. In addition, we are constructing risk scores using the GWSS summary statistics for stratifying patients.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1565 Genomic Analysis of Plasmodium falciparum isolates across different altitudinal zones along the slope of Mount Cameroon.

Authors:

M. Ajonina[1,2], T. Apinjoh[1]; [1]Univ. of Buea, Buea, Cameroon, [2]Charisma Univ., Providenciales, Turks and Caicos Islands

Malaria elimination is threatened by the emergence and rapid spread of drug resistance. Understanding the demographic history of *Plasmodium falciparum* and the genetic basis of adaptation to antimalarial treatment and host immunity is critical to elimination efforts. This study sought to characterize the diversity of *P. falciparum* isolates across different altitudes along the slope of Mount Cameroon, genotyped using the Illumina next-generation sequencing platform. A total of 77,253 quality single-nucleotide polymorphisms were identified from 220 *P. falciparum* clinical isolates from high (24,214), intermediate (24,426), and low (28,613) altitude. About 49%, 48.2% and 30% of the parasite isolates from high, intermediate, and low altitudes, respectively had F_{WS} values &gt 0.95 indicative of dominant mixed genotype infections and low population sub-structure with high potential for out-crossing. No significant difference was observed in within-host diversity while population structure analysis did not separate the isolates in the three major altitudinal groups by PCA, F_{ST} and admixtures, suggesting bidirectional gene flow among the populations. A total of 94 antigenic genes under balancing selection were detected in the area including vaccine candidate gene *ama1, eba175, msp1, trap, dblmsp* and *clag2*. Moreover, 17 of these genes were identified to be under both recent positive directional and positive balancing selection including the prominent host immune target genes *surfin 8.2, trap,* and *ama1*. Recent directional selection analysis using integrated standardized haplotype score (iHS) did not detect any selection signatures in the *Pfcrt, Pfdhfr, Pfdhps, Pfmdr1,* and *PfK13* genes. Furthermore, no *PfKelch13* validated mutation associated with artemisinin resistance was identified in this study and no structural divergence was noticed among the *P. falciparum* parasite populations across different altitudes around the Mount Cameroon region.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1566 Genomic and transcriptomic influences on obesity-related traits in a high obesity-risk Hispanic/Latino population

Authors:

D. Kim[1], H. G. Polikowsky[2], H. M. Highland[1], H-H. Chen[2], Y. Wang[1], M. Graff[1], K. L. Young[1], M. Y. Anwar[1], X. Zhang[1], W. Zhu[2], P. Sharma[1], S. Berndt[3], J. D. Arias[3], A. D. Gutierrez[4], M. Lee[5], J. B. McCormick[5], S. P. Fisher-Hoch[5], J. E. Below[2], K. E. North[1]; [1]UNC-Chapel Hill, Chapel Hill, NC, [2]Vanderbilt Univ., Nashville, TN, [3]Natl. Cancer Inst., Rockville, MD, [4]Univ. of Texas Hlth.Sci. Ctr. at Houston, Houston, TX, [5]UTH SPH Brownsville, Brownsville, TX

Approximately 66% of American adults are overweight or obese, fueling a public health crisis with financial costs projected to double every decade. Obesity-related genome-wide association studies (GWAS) have identified >1000 loci to date; however, most studies focus on crude measures like body mass index (BMI), and functional interpretation of GWAS findings remains incomplete. Thus, it is important to investigate precision obesity phenotypes coupled with additional omics data so that we can link genetic variants to causal genes and underlying pathways, in particular for historically marginalized populations. We aimed to discover causal relationships and reveal molecular mechanisms of obesity-related traits in Mexican American participants of the Cameron County Hispanic Cohort (CCHC). We performed GWAS of 13 obesity-related traits, including anthropometric traits, DEXA regional fat measures, and adipokine levels among CCHC study participants (N_{max} for GWAS = 4,271). Using directly measured transcripts in whole blood among the same CCHC participants (N_{max} = 866), we conducted two transcriptomic analyses - i) eQTL and ii) differential expression (DE) analysis. Among suggestive GWAS loci ($p < 5 \times 10^{-6}$), we identified transcripts in the 1 MB region that were significantly associated with the GWAS SNP ($p <$ 0.05 /number of transcripts tested) and for which DE was associated with the given obesity-related trait. We identified five significant SNP-transcript-trait associations from directly measured transcript levels - 1) rs74471501 - *FAXDC2* - BMI; 2) rs2774424 - *ARHGEF7* - visceral adipose tissue (VAT) to subcutaneous adipose tissue (SAT) ratio; 3) rs6810075 - *EIF4A2* - adiponectin-resistin index (ARI); 4) rs3743541 - *RP11-830F9.7* - VAT_{male}; 5) rs543314376 - *MAPK11* - SAT_{female}. We additionally performed colocalization analysis between our five genome-wide significant ($p < 5 \times 10^{-8}$) loci and GTEx gene expression data using fastENLOC. Two GWAS locus-transcript pairs demonstrated the locus-level colocalization posterior probability of ≥ 0.5 - 1) rs7171408 region (SAT and BFP) - *ANKDD1A* in multiple tissues, and 2) rs6810075 region (ARI adjusted for BMI) - *MCF2L2 pseudogene* in Brain). Many of these identified genes are relevant to obesity-related metabolism; for example, the *MAPK* pathway (with respect to SAT - *MAPK11)* plays a crucial role in appetite regulation, adipogenesis, and glucose homeostasis. In conclusion, we provide compelling evidence for novel associations between genetic variants, obesity-related traits, and genes mediating these relationships among Hispanic/Latino populations.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1567 Genomic heterozygosity is associated with a lower risk of osteoarthritis.

Authors:

G. Zhai[1], R. Gill[1], M. Liu[2], G. Sun[2], A. Furey[1], P. Rahman[2]; [1]Fac Med, Mem. Univ Newfoundland, St John's, NL, Canada, [2]Fac Med, Mem. Univ Newfoundland, St. John's, NL, Canada

Objective: Osteoarthritis (OA) is a chronic heterogeneous disease affecting over 500 million people worldwide. Genomic heterozygosity has been shown to confer a health advantage in humans and play a protective role in complex diseases. Given OA is a highly polygenic disease, we set out to determine if an association existed between OA and genomic heterozygosity. **Methods:** A case-control study design was utilized. End-stage knee and hip OA patients and healthy controls were recruited from the Newfoundland and Labrador (NL) population. DNA was extracted from their blood samples and genotyped by the Illumina GWAS microarray genotyping platforms. Directly genotyped GWAS data was imputed by Sanger Imputation Server and individual rates of observed heterozygosity (HetRate) and heterozygosity excess (HetExcess) relative to the expected were mathematically derived, and then standardized to a z-score. Logistic regression modelling was used to examine the association between OA and HetRate or HetExcess with adjustment for age, sex, and BMI. The Arthritis Research UK Osteoarthritis Genetics (arcOGEN) consortium database which contained end-stage OA (knee or hip) and control patients was utilized as a replication cohort to validate our results. **Results:** A total of 559 knee and hip OA patients (mean age 66.5 years, BMI 33.7 kg/m2, and 55% females) and 118 healthy controls (mean age 56.4 years, BMI 29.5 kg/m2, and 59% female) were included in the analysis as a discovery cohort. We found that OA had an inverse relationship with HetRate and HetExcess with an odds ratio of 0.64 (95% CI: 0.45-0.91) and 0.65 (95% CI: 0.45-0.93) per standard deviation (SD), respectively, after adjustment for age, sex, and BMI. This association remained unchanged for knee and hip joint-specific analyses. The findings were validated with the arcOGEN database which showed HetRate and HetExcess were associated with OA with odds ratios of 0.60 (95% CI: 0.56-0.64) and 0.44 (95% CI: 0.40-0.47) per standard deviation (SD), respectively. **Conclusions:** Our results were the first to clearly demonstrate that genomic heterozygosity plays an important role in the risk of OA, which could be used for early identification of individuals with a predisposition for OA.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1568 Genomic rare variant mechanisms for congenital cardiac laterality defect: A digenic model approach.

Authors:

A. Rai[1], B. Yuan[2], P. Lupo[3], C. Lo[4], R. Gibbs[2,5], J. Posey[2], J. Lupski[2,3,5], S. Morris[3], Z. Coban Akdemir[1,2]; [1]Human Genetics Ctr., Dept. of Epidemiology, Human Genetics, and Environmental Sci., Sch. of Publ. Hlth., The Univ. of Texas Hlth.Sci. Ctr. at Houston, Houston, TX, [2]Dept. of Molecular and Human Genetics, Baylor Coll. of Med., Houston, TX, [3]Div. of Pediatric Cardiology, Dept. of Pediatrics, Texas Children's Hosp. and Baylor Coll. of Med., Houston, TX, [4]Dept. of Dev.al Biology, Univ. of Pittsburgh Sch. of Med., Pittsburgh, PA, [5]Human Genome Sequencing Ctr., Baylor Coll. of Med., Houston, TX

Laterality developmental abnormalities are defined by perturbations in the usual left-right (LR) asymmetry of organs in the body. They involve a spectrum of disorders that range from cardiac D-transposition of the great arteries affecting only the heart to complex conditions including situs inversus totalis and heterotaxy which affect the entire thorax and abdomen. The heart is often affected in laterality defects as cardiac formation in early embryo requires proper LR axis signaling cues. Laterality defects show extensive anatomic heterogeneity in their genetic origins (AD, AR, XL, multigenic, etc.) and phenotypic presentation and result from heterogeneous underlying genetic etiology. The majority of their underlying molecular causes remain elusive due to its rarity, genetic and phenotypic heterogeneity, indicating that additional mechanisms remain to be elucidated. Several recent studies suggest that digenic/oligogenic inheritance models may contribute to the complex genetics of these defects. We investigated a digenic model, focusing our analysis on genes with known laterality defect associations. We analyzed a cohort of 271 "unsolved" proband-only exomes from individuals with clinically diagnosed laterality defects using family-based segregation analysis. Exonic and splicing variants with MAF <= 0.01% in gnomAD and CADD score of >=10 were used as a filtering criteria for rare and deleterious variants in 112 proposed laterality defect genes. Thirty probands were found to have >=2 rare and damaging variants, segregation analysis was performed in 21 probands for whom both parental samples were available. Variant allele confirmation and segregation analysis indicated likely damaging trans-heterozygous digenic variants in 7 probands, in genes previously associated with primary ciliary dyskinesia (PCD) such as *DNAH1/DNAH6*, *RSPH4A/DNAH6*, *DNAH6/DNAH9*, *DNAH9/TTC12*, *DNAH1/DNAH11*, *DRC1/GAS8*, and *CCDC40/DNAH1*- consistent with a digenic and or triallelic inheritance ciliopathy model. Parents of all probands were unaffected. In an extended family of 8 members, the female proband was found to have *DNAAF2/DNAH5* digenic variants, with both alleles inherited from the affected father. Of note, both of the variants were not found together in other unaffected 6 family members. These data provide further evidence that digenic epistatic interaction can contribute to the complex genetics of laterality defects. Functional studies for these identified variants and in other PCD genes in patients with laterality defects may provide further insights into the developmental biology and biological perturbations underlying birth defects.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1570 GWAS of Torus Palatinus Reveals Several Suggestive Loci

Authors:

M. Lee[1], A. M. El Sergani[1], C. D. Padilla[2], C. J. Buxó-Martínez[3], R. E. Long, Jr[4], L. M. Moreno Uribe[5], W. L. Adeyemo[6], A. Butali[5], M. L. Marazita[1], J. R. Shaffer[1], S. M. Weinberg[1]; [1]Univ. of Pittsburgh, Pittsburgh, PA, [2]Univ. of the Philippines, Manila, Philippines, [3]Univ. of Puerto Rico, San Juan, Puerto Rico, [4]Lancaster Cleft Palate Ctr., Lancaster, PA, [5]Univ. of Iowa, Iowa city, IA, [6]Univ. of Lagos, Lagos, Nigeria

Torus Palatinus (TP) is a midline bony projection located on the oral surface of the hard palate. It is generally considered benign, although when very large it can interfere with oral function. TP is common with reported incidence varying widely by both sex and ancestry. Early family-based studies have suggested autosomal dominant inheritance with reduced penetrance. Several studies have identified rare mutations associated with TP, but these can only explain a small fraction of cases. Thus, the genetics of TP remains largely unclear. In this study we performed GWAS of TP in a diverse cohort of 829 adults genotyped on the Illumina Infinium Global Diversity Array and imputed to the TOPMed reference. TP was evaluated from 3D digital impressions of the palate. In total, 41.1M SNPs were tested in GENESIS while adjusting for 5 ancestry principal components and sex as covariates. Thirteen SNPs showed suggestive associations. We highlight intronic SNP rs4704136(beta(se)=-0.86(0.17), p-value=7.37 x 10^{-7}), rs17013053(beta(se)=1.62 (0.33), p-value=8.28 x 10^{-7}) and rs1859400(beta(se)=1.25(0.25), p-value=9.51 x 10^{-7}), which are located near facial development genes (e.g., *LRRTM4, HEXB, SUPT4H1*). These results suggest that TP may be an oligogenic or polygenic trait.Results await confirmation via independent replication.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1571 Hemophagocytic lymphohistiocytosis gene variants in severe aplastic anemia and their impact on post hematopoietic cell transplantation outcomes.

Authors:

M. Rafati[1], L. McReynolds[1], Y. Wang[1], K. Jones[1], C. Dagnall[2], B. Hicks[3], S. Spellman[4], M. He[4], Y-T. Bolon[4], S. J. Lee[5], S. Savage[1], S. Gadalla[1]; [1]Natl. Cancer Inst., NIH, Rockville, MD, [2]Natl. Cancer Inst. /Leido, Bethesda, MD, [3]Leidos Biomed/DCEG/NCI/NIH, Bethesda, MD, [4]Ctr. for Intl. Blood and Marrow Transplant Res., Minneapolis, MN, [5]Clinical Res. Div., Fred Hutchinson Cancer Ctr., Seattle, WA

Severe aplastic anemia (SAA) is most frequently immune-mediated, but a subset of patients has a germline genetic predisposition including rare variants in hemophagocytic lymphohistiocytosis-associated (HLH) genes. Germline genetic testing for patients with SAA is important to guide their treatment plan and adjust hematopoietic cell transplantation (HCT) modalities. HLH is predominantly autosomal recessive (AR) or X-linked recessive (XLR). The frequency of monoallelic HLH gene variants and whether they are associated with SAA HCT outcomes are not known. The Transplant Outcomes in Aplastic Anemia project is a collaboration between the NCI and the Center for International Blood and Marrow Transplant Research and consists of genomic and clinical data from 824 patients with SAA of which 140 have an inherited marrow failure syndrome. We used exome sequencing data from the 684 patients with acquired SAA who underwent HCT between 1989-2015 to identify variants in 14 HLH-associated genes (11 AR, 3 XLR) which were curated using ACMG/AMP criteria. Deleterious variants of uncertain significance (dVUS) were defined as those with damaging predictions in => 3/5 meta-predictors (BayesDel, REVEL, CADD, MetaSVM and EIGEN). Patients with potentially causal variants fitting XLR or AR inheritance patterns were deemed unrecognized HLH (UHLH). The Kaplan-Meier estimator was used to calculate the probability of overall survival (OS) after HCT. There were 77 HLH variants in 49 patients; 24 were pathogenic or likely pathogenic variants (PLPVs) identified in 19 patients (2.8%) and 53 were dVUS. Monoallelic PLPVs in *PRF1* were the most frequent, present in 8/19 patients. Three patients fit the criteria for probable UHLH, including a male with a hemizygous PLPV in *SH2D1*, and two patients with dVUSs (one biallelic in *UNC13D*and one homozygous in *AP3B1*). Two UHLH patients survived until last follow-up and the third died one month post-HCT. HCT outcomes in patients with monoallelic PLPVs were statistically comparable to those of patients without variants. The 5-year OS probability in patients with PLPVs was 89% (95% CI=72-99) compared with 66% (95% CI=63-70) in those without variants (p=0.06). No statistically significant associations between HLH variants and other HCT outcomes were noted. We found that 7.2% of patients undergoing HCT for SAA carry potentially deleterious variants in HLH genes, with 2.8% carrying a PLPV. Patients with heterozygous PLPVs did not have different HCT outcomes compared with those without variants. Our data suggest that identification of monoallelic variants in HLH genes does not influence HCT outcomes and no special treatment consideration is warranted.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1572 Hepatic steatosis is reduced in PNPLA3 carriers on the WELL program: a very low carbohydrate diet, mindful eating, and positive affect online intervention

Authors:

E. Speliotes[1], J. Krinock[1], A. O'Brien[1], K. Raymond[1], J. Moskowitz[2], J. Daubenmier[3], H. Bayandorian[4], D. J. Isaman[1], D. H. Griauzde[1], L. R. Saslow[1]; [1]Univ. of Michigan, Ann Arbor, MI, [2]Northwestern Univ., Chicago, IL, [3]San Francisco State Univ., San Francisco, CA, [4]none, San Francisco, CA

Background: Nonalcoholic fatty liver disease (NAFLD) affects about 25% of individuals. No effective medications are currently available to treat NAFLD. Our preliminary data suggests that insulin resistance and the G allele of rs738409 interact to amplify risk of developing NAFLD. One effective way to reduce insulin resistance is using a very low-carbohydrate (VLC) eating pattern which is a moderate protein, higher fat eating pattern. Mindfulness and positive affect teachings may also increase adherence to behavioral changes. **Methods:** We carried out a single-arm, online trial in adults with rs738409-GG or -CG. Participants completed measures on intervention feasibility, acceptability, and physical and patient-reported outcomes. Participants were taught to follow an ad libitum very low-carbohydrate eating pattern (20-35 net or non-fiber grams) of carbohydrates per day for 4 months. They were given supplementary psychological tools on mindfulness and positive affect. **Results:** Eleven participants were enrolled, 9/11 (82%) completed study outcomes, and 8/11 (73%) attended at least half of the sessions. In the intention-to-treat analyses we found: body weight decreased 10.9% ($p < 0.001$), HbA1c reduced 9.4% ($p = 0.001$), insulin reduced 30.7% ($p = 0.020$), insulin resistance reduced 36.8% ($p = 0.007$), triglycerides reduced 14.3% ($p = 0.021$), abdominal symptoms reduced 33.2% ($p = 0.044$), emotional function improved 22.6% ($p = 0.16$) and fatigue reduced 27.1% ($p = 0.024$). Six participants adhered to the eating pattern (defined based on their 4-month 24-hour dietary recalls), 3 participants were not adherent, and 2 had no 4-month 24-hour dietary recall. Adherent participants decreased body weight by 12.0% ($p < 0.001$), decreased liver lobe fat percent by 53.1% ($p = 0.001$), reduced AST by 33.9% ($p = 0.006$), reduced ALT by 47.5% ($p = 0.003$), reduced HbA1c by 7.7% ($p = 0.014$), reduced insulin by 40.5% ($p = 0.023$), and reduced insulin resistance by 44.0% ($p = 0.019$); LDL-cholesterol levels were not significantly changed. Nine participants provided 4-month self-report information. These nine were highly satisfied with the program (mean 6.22, 95% CI 5.58 to 6.85), with 5/9 (56%) giving the intervention the top score. Only 11% reported that they would stop the assigned eating pattern as soon as `the study was over, and 44% stated that they did not plan to ever stop following it. **Conclusion**: These results support the feasibility, acceptability, and preliminary efficacy of the VLC intervention in adults with high genetic risk for NAFLD. The results of this pilot study suggest that a VLC eating pattern is safe and efficacious approach for reducing hepatic steatosis in this population.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1573 Heritability estimates for plasma biomarkers of Alzheimer disease in the Amish population.

Authors:

P. Wang[1], A. Lynn[1,2], Y. E. Song[1], K. Miskimen[1], M. B. Prough[3], D. A. Dorfsman[3,4], A. Fernandez[3], R. Laux[1], S. L. Fuzzell[1], S. D. Hochstetler[1], A. F. Zaman[3], L. D. Adams[3], L. J. Caywood[3], J. E. Clouse[3], S. D. Herington[3], P. Whitehead[3], Y. Liu[5], L. R. Main[2,6], P. Ogrocki[7,8], A. J. Lerner[7,8], A. J. Griswold[3,4], J. M. Vance[3,4], M. L. Cuccaro[3,4], W. K. Scott[3,4], M. A. Pericak-Vance[3,4], J. L. Haines[1,2]; [1]Dept. of Population and Quantitative Hlth.Sci., Case Western Reserve Univ. Sch. of Med., Cleveland, OH, [2]Cleveland Inst. for Computational Biology, Case Western Reserve Univ., Cleveland, OH, [3]John P. Hussman Inst. for Human Genomics, Univ. of Miami Miller Sch. of Med., Miami, FL, [4]The Dr. John T. Macdonald Fndn. Dept. of Human Genetics, Univ. of Miami Miller Sch. of Med., Miami, FL, [5]Dept. of Nutrition, Case Western Reserve Univ. Sch. of Med., Cleveland, OH, [6]Dept. of Genetics and Genome Sci., Case Western Reserve Univ. Sch. of Med., Cleveland, OH, [7]Dept. of Neurology, Univ. Hosp. Cleveland Med. Ctr., Cleveland, OH, [8]Dept. of Neurology, Case Western Reserve Univ. Sch. of Med., Cleveland, OH

Alzheimer Disease (AD) is the most common cause of dementia with a strong genetic component. Emerging studies have found a correlation between various plasma biomarkers and AD pathology, cognitive function, and neurodegeneration. However, limited research has been done to explore the influence of genetic variation on these plasma biomarkers, including the 42- and 40-amino acid forms of amyloid beta (Aβ40, Aβ42), total tau (tTau), phosphorylated tau 181 (pTau181), neurofilament light chain (NfL), and glial fibrillary acidic protein (GFAP).
We recruited 920 individuals from 16 extended, multigenerational Amish families in Ohio and Indiana to estimate the pedigree-based heritability of each biomarker. Genome-wide array data was employed to assess the SNP-based heritability for each biomarker. Plasma Aβ40, Aβ42, tTau, pTau181, NfL, and GFAP concentrations were measured using the Simoa™ Neuro-3Plex, 4Plex, and pTau181 Advantage V2 assays. Genotyping was performed using the Illumina Expanded Multi-Ethnic Genotyping Array and Global Screening Array. All plasma biomarkers were inverse-normalized before analysis. Pedigree-based heritability was estimated using a variance component model, adjusting for study center, age, sex, and diagnosis. *ApoE* e4 dosage was assessed as an additional covariate. SNP-based heritability was estimated using the Genomic Relatedness Restricted Maximum Likelihood model, stratifying SNPs based on linkage disequilibrium and adjusting for the same covariates above.
Pedigree-based heritability estimates were: Aβ40 = 14.8%, Aβ42 = 12.9%, Aβ42/40 = 5.3%, tTau = 38.0%, pTau181 = 25.8%, NfL = 60.3%, GFAP = 79.0%. Including *ApoE* e4 dosage as a covariate significantly affected the heritability of four biomarkers. It increased the heritability of plasma Aβ42 (13.7%, $p = 0.04$) and pTau (27.3%, $p < 0.0001$) but decreased that of Aβ42/40 (2.53%, $p < 0.0001$) and GFAP (29.7%, $p = 0.02$). SNP-based heritability estimates for plasma Aβ40, NfL, and GFAP (4.5%, 38.8%, and 66.2%) were lower than pedigree-based estimates; however, for Aβ42/40, SNP-based approach yielded a higher heritability of 48.4%.
We found that plasma biomarkers, such as Aβ42/40, tTau, pTau181, NfL, and GFAP, are highly heritable in the Amish population. These findings suggest a strong genetic determinant of plasma biomarkers associated with AD in the Amish population. Given plasma biomarkers are more objective than heterogeneous AD diagnosis and easier to measure, they may serve as surrogate markers to study the genetics of AD.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1574 Hidden protein-altering variants influence diverse human phenotypes

Authors:

M. Hujoel[1], R. Handsaker[2], M. Sherman[3], N. Kamitaki[4], A. Barton[5], R. Mukamel[6], C. Terao[7], S. McCarroll[6], P-R. Loh[8]; [1]Brigham & Women s Hosp./Harvard Med. Sch., Boston, MA, [2]Broad Inst., Charlemont, MA, [3]Massachusetts Inst. of Technology, Marblehead, MA, [4]Brigham and Women's Hosp., Boston, MA, [5]Harvard Univ., Cambridge, MA, [6]Harvard Med. Sch., Boston, MA, [7]RIKEN IMS, Yokohama, Japan, [8]Brigham and Women s Hosp. / Harvard Med. Sch, Boston, MA

Structural variants (SVs) comprise the largest genetic variants, altering from 50 base pairs to megabases of DNA. However, SVs have not been effectively ascertained in most genetic association studies, leaving a key gap in our understanding of human complex trait genetics. We identified protein-altering SVs from UK Biobank (UKB) whole-exome sequencing data (*n*=468,570) using haplotype-informed methods capable of detecting sub-exonic SVs and variation within segmental duplications. Subsequent association and fine-mapping analyses identified hundreds of protein-altering SVs that appeared to influence health-related traits in UKB.

Rare CNVs predicted to cause gene loss-of-function (pLoF) enabled new insights into effects of gene inactivation. A partial deletion of *RGL3* exon 6 conferred one of the strongest protective effects on hypertension of all coding variants in the human genome (OR = 0.86 [0.82-0.90], $P = 6.3 \times 10^{-10}$), replicating in *All of Us* (AoU; OR = 0.83 [0.75-0.92], *P*= 0.00026). Knockout of *RGL3* appeared to be well-tolerated based on 37 homozygotes in UKB, suggesting RGL3, or a pathway in which it functions, as a potential drug target. Incorporating pLoF CNVs into gene-level burden analyses of 57 quantitative traits identified 100 pLoF gene-trait associations undetectable from analyses of SNVs and indels alone, representing a 20% increase in power. Several associations implicated new gene-trait relationships, even for well-studied traits such as height for which common-variant GWAS have reached saturation. Replication analyses in BioBank Japan demonstrated consistent effect sizes for five new height genes with adequate replication power.

Common protein-coding variation in rapidly-evolving gene families within segmental duplications—previously invisible to most analysis methods—appeared to generate some of the human genome's largest contributions to variation in type 2 diabetes (T2D) risk, chronotype, and blood cell traits. In a highly polymorphic 99kb segmental duplication at 7q22.1 (1-7 copies per allele), copy number of a common missense variant in *RASA4* appeared to strongly influence T2D risk (1.3-fold [1.2-1.4] range, $P = 1.3 \times 10^{-25}$ in UKB; replication $P = 2.8 \times 10^{-5}$ in AoU). The same paralogous sequence variant in *RASA4* appeared to drive the strongest association genome-wide with chronotype ($P = 2.6 \times 10^{-72}$). Additionally, variation at the immune-related *FCGR2/3* (Fcγ receptor) and *DEFA1/A3* (α-defensin) segmental duplication loci generated two of the top five associations with basophil count. These results illustrate the potential for new genetic insights from genomic variation that has escaped large-scale analysis to date.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1575 Higher genetic risk for type 2 diabetes is associated with progressive decline of beta cell function.

Authors:

H. Lee[1], J. Choi[1], K. Park[1], N. Cho[2], S. Kwak[3]; [1]Seoul Natl. Univ., Seoul, Korea, Republic of, [2]Ajou Univ., Incheon, Korea, Republic of, [3]Seoul Natl. Univ. Hosp., Seoul, Korea, Republic of

Objective
While the association between the polygenic risk score (PRS) for type 2 diabetes (T2D) and static beta cell function is known, the relationship between T2D PRS and the progressive deterioration of beta cell function remains largely unknown. We aimed to evaluate the association of PRS and the trajectory of beta cell function in a community based prospective cohort.
Method
We analyzed 6,311 participants without T2D at baseline from the Ansan-Ansung cohort with 75-g 2-hour oral glucose tolerance tests taken biennially for 14 years. Disposition index (DI), derived from insulinogenic index at 60 minutes and Matsuda index, was used as a marker for beta cell function. PRS was calculated using 1.0 million variants derived from T2D genome-wide association studies from multiple ancestries using Bayesian regression (PRS-CSx). Participants were stratified into low (1^{st} quintile), intermediate (2^{nd} to 4^{th} quintiles) and high genetic risk (5^{th} quintile). Lifestyle was categorized as unfavorable, intermediate and favorable lifestyle according to Life's Essential 8. Linear mixed model was applied.
Result
During a median follow-up of 13 years, 369 (29.4%), 872 (23.2%) and 172 (13.3%) participants developed T2D in high, intermediate and low genetic risk group, respectively. Participants in high genetic risk group, compared to low genetic risk group, had a 27% lower DI at baseline (53 [95% CI 50, 57] *vs* 73 [68, 78]; $P = 7.6x10^{-10}$), and a 2.3-fold faster rate of decline in log_2(DI) per year (-0.037 [-0.043, -0.030] *vs* -0.016 [-0.022, -0.0090]; $P = 1.8x10^{-5}$). Notably, we found significant interaction between the rate of decline in DI and T2D PRS ($P = 1.4x10^{-4}$). Of the 161 known lead variants associated with T2D in East Asians, a variant near SLC30A8 (rs13266634) was significantly associated with the progressive decline in disposition index ($P = 0.000153$, <0.05/161). PRSs of other T2D-related traits (Homeostasis Model Assessment of Beta-cell function (HOMA-B), fasting glucose, fasting insulin, waist hip ratio adjusted by body mass index) did not perform better than T2D PRS. Healthy lifestyle was associated with attenuated rate of decline in DI across all genetic risk group.
Conclusion
Individuals with a high genetic risk for T2D not only exhibit a lower baseline DI, but also experience a more rapid decline in DI over time. Genetic information could be used to identify those at risk for faster decline of beta cell function, enabling a focus on lifestyle intervention.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1576 Higher levels of chronic illness in the Australian Genetics of Bipolar Disorder Study are associated with genetic risk for reduced lifespan, increased BMI, and increased sites of chronic pain: a polygenic risk score analysis

Authors:

P. Lind, B. Mitchell, N. Martin, S. Medland; QIMR Berghofer Med. Res. Inst., Brisbane, Australia

There are extensive research findings that individuals with bipolar disorder (BD) experience high levels of comorbid chronic illness and reduced life expectancy. We examined multimorbidity data in the Australian Genetics of Bipolar Disorder Study using unweighted and weighted Rx-Risk Comorbidity Index scores derived from prescription records. We then investigated, for the first time, whether individual differences in these Rx-Risk Comorbidity Index scores could be predicted by polygenic risk scores (PRS) calculated from genome-wide association studies for lifespan (Timmers et al 2019), chronic pain (Johnston et al 2019) and BMI (leave out QIMR; Yengo et al 2018). Prescription and genotyping data were available for 3,562 unrelated participants (67% female; 64.9% BD-I, 22.5% BD-II and 12.6% sub-threshold BD). Record linkage was available for prescription data from 30/06/2014-30/06/2019 for all dispensed medications listed on the Australian Pharmaceutical Benefits Scheme schedule. The unweighted Rx-Risk score was calculated as the total number of 46 comorbidities mapped to prescriptions using Anatomical Therapeutic Chemical Classification System codes. A weighted score was calculated using mortality-related Rx-Risk weights for 43 comorbidities. PRS were calculated using the PLINK profile score method for clumped SNPs. Linear regression on the profile scores were performed, controlling for ancestry, sex, age and age^2 at survey time, and sex*age. Almost all participants had one or more prescription-based comorbidities, with an average unweighted Rx-Risk score of 6.7. After correcting for multiple testing, genetic predisposition for lifespan, BMI and chronic pain all significantly ($p<2E-06$) predicted the unweighted and weighted Rx-Risk scores (explaining 0.7-2.0% of variance) with effects going in the expected direction. In summary, we confirm individuals with lived experience of BD have high rates of comorbid conditions which are observable from electronic prescription records. We have also shown that the Rx-Risk Comorbidity Index can be predicted by polygenic risk scores.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1577 High-throughput measurement of regulatory SNPs with biased allelic enhancer activity effect for osteoporosis

Authors:

X-F. Chen, Y-Y. Duan, Y-Y. Jia, Y. Guo, T-L. Yang; Xi'an Jiaotong Univ., Xi'an City, China

Osteoporosis is a common age-related skeletal disease characterized by low bone mineral density (BMD) and increased fragility and fracture risk. Recent studies have uncovered crucial roles of chromatin loop in controlling osteogenesis. However, the precise roles of 3D chromatin organization at osteoporosis risk loci remains largely elusive. Here, we firstly integrated GWAS fine-mapping across multiple osteoporosis-relevant traits, including BMD at 6 different sites (forearm, femoral neck, lumbar spine, heel, total body and total-body less head) and fracture from the GEFOS portal, and 3D chromatin interaction analysis in human mesenchymal stem cells (hMSCs) induced osteoblasts to systematically prioritize candidate osteoporosis-associated functional SNPs. We prioritized 5,642 candidate osteoporosis-associated functional SNPs with promoter chromatin interactions. We then utilized an adapted self-transcribing active regulatory region sequencing (STARR-seq) to assess their allelic enhancer regulatory activities in human osteoblast-like cell line U2OS and identified 505 SNPs with biased allelic enhancer activity effect (baaSNPs) from 165 loci (FDR<0.05 by Fisher's exact test). Functional characterization revealed significant enrichment for open chromatin region and active epigenetic markers in osteoblasts for identified baaSNPs. Further functional analysis suggested significant enrichment for biological pathways relevant to skeletal growth and osteoblast differentiation on their chromatin interacted genes on baaSNPs. Collectively, our results highlighted potential important regulatory roles of 3D chromatin organization at osteoporosis risk loci and provided a valuable enhancer-promoter genetic regulatory atlas for osteoporosis, which may help better understand molecular mechanisms underlying osteoporosis-risk loci.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1578 HOMA-B/-IR GWIS using fasting glucose and insulin suggests causal role of inflammation in insulin resistance

Authors:

Z. Balkhiyarova[1,2,3], I. O. Fedko[4,5], L. Zudina[1,3], Meta-Analyses of Glucose and Insulin-related traits Consortium (MAGIC) Investigators, O. Yelizarova[6], R. Mägi[7], M. Kaakinen[1,2,3], D. I. Boomsma[4,5], I. Prokopenko[1,2]; [1]Univ. of Surrey, Guilford, United Kingdom, [2]People-Ctr.d Artificial Intelligence Inst., Univ. of Surrey, Guildford, United Kingdom, [3]Imperial Coll. London, London, United Kingdom, [4]VU Univ. Amsterdam, Amsterdam, Netherlands, [5]Amsterdam Publ. Hlth.Res. Inst., Amsterdam, Netherlands, [6]Inst. for Publ. Hlth.O.M. Marzieiev, Natl. Academy of Med. Sci. of Ukraine, Kyiv, Ukraine, [7]Univ. of Tartu, Tartu, Estonia

Introduction: Studying the development and manifestation of negative effects of obesity, type 2 diabetes (T2D) is a priority for science today. Insulin resistance is one of the predictors of metabolic disorders related to above-mentioned conditions. The aim of this study was to obtain new knowledge about the genetic relationships between insulin resistances, disorders of glucose metabolism, dyslipidemia, fibrinolytic activity (or coagulation markers), and inflammation.**Materials and Methods**: We implemented a Genome-Wide Inferred Statistics (GWIS) approach for homeostasis Model Assessment of β-cell function and insulin resistance (HOMA-B/-IR, N_{GWIS}=75,240). We dissected the effects of glycaemic trait/T2D loci effects on HOMA-B/-IR; genetic relationships between HOMA-B/-IR and 36 inflammatory and cardiometabolic phenotypes. We evaluated causal relationships between PAI-1 and HOMA-IR using two-sample Mendelian randomization approach.**Results**: We identified one novel HOMA-B (*FOXA2*) and three novel HOMA-IR (*LYPLAL1, PER4, PPP1R3B*) loci. We explored and defined a large group of T2D loci influencing insulin resistance. We detected novel strong genetic correlations between HOMA-IR/-B and Plasminogen Activator Inhibitor-1 (PAI-1, r_g=0.92/0.78, P=2.13×10^{-4}/2.54×10^{-3}). HOMA-IR/-B were correlated with C-Reactive Protein (r_g=0.33/0.28, P=4.67×10^{-3}/3.65×10^{-3}) and HOMA-IR was correlated with adiponectin (r_g=-0.30, P=0.012) among others. We detected a causal effect of increased levels of PAI-1 on HOMA-IR (beta = 0.070 ln (ng/mL), 95%CI [0.033, 0.11], P = 2.13×10^{-4}).**Conclusions**: We found support for the role of inflammation in the pathogenesis of metabolic disorders through their genetic relationships improving our understanding of the genetic mechanism underlying development and pathophysiology of complications related to metabolic disorders, including obesity and T2D.**Funding:** LongITools H2020-SC1-2019-874739, Diabetes UK 20/0006307, PreciDIAB ANR-18-IBHU-0001

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1579 Host genetics and gut microbiota in asthma among US Hispanics/Latinos: The Hispanic Community Study / Study of Latinos.

Authors:

M. Stanislawski[1], E. Litkowski[2], C. Arehart[3], N. Gilmore[4], L. Lange[5], E. Lange[4], A. Christy[6], M. A. Katie[6], F. Holguin[4], K. North[7], R. Burk[8], R. Kaplan[9]; [1]Univ. of Colorado Sch. of Med. 1, Aurora, CO, [2]Univ. of Colorado, Aurora, CO, [3]Univ. of Colorado, Boulder, CO, [4]Univ. of Colorado Sch. of Med., Aurora, CO, [5]Univ Colorado Denver, Aurora, CO, [6]Univ. of North Carolina, Chapel Hill, NC, [7]Univ North Carolina, Wake Forest, NC, [8]Albert Einstein Col Med., Bronx, NY, [9]Albert Einstein Coll. of Med., Bronx, NY

Rationale: Asthma is a heterogeneous condition influenced by diverse factors that is often comorbid with obesity. Defining the contribution of genetic and gut microbial factors could improve our understanding of the pathophysiology of these interrelated conditions. **Objectives**: To quantify associations between gut microbiota characteristics and asthma/obesity status among US Hispanic/Latinos, and to integrate genetic and microbiota information to assess their relative contributions to these conditions. **Methods**: We used data from shotgun metagenomic sequencing of stool DNA from N=2404 participants of the Hispanic Community Health Study/Study of Latinos and assessed associations of microbiota characteristics with current asthma and obesity status, defining obesity as body mass index$\geq$30kg/m^2: non-obese asthma (N=86); obese asthma (N=105); and non-asthmatic obesity (N=920) versus non-obese with no history of asthma (N=1293). We used multivariable-adjusted regression-based methods, including permutational ANOVA of overall microbiota composition (beta diversity) and ANCOM-BC of species-level taxa with the Holm method to adjust for multiple comparisons. Based on literature-supported genetic variants, we derived an asthma polygenic risk score (PRS) using host genetic data. We split the data into training (N=1353) and testing (N=1051) sets for taxonomic analyses and to assess the classification accuracy of genetic and microbial risk factors for asthma/obesity status compared to baseline risk factors alone, using a sequential adjustment approach and likelihood ratio tests to compare models. **Measurements and Main Results**: We observed significant associations of overall gut microbiota composition with non-obese asthma, asthmatic obesity, and non-asthmatic obesity versus non-obese non-asthmatic (all $p<0.002$). We observed distinct taxonomic associations with non-obese and obese asthma, such as *Acidaminococcus intestini* with increased risk of non-obese asthma and *Fournierella massiliensis* with decreased risk of obese asthma. The asthma PRS improved models of non-obese ($p=0.02$; AUC=0.79) and obese asthma ($p=0.0008$; AUC=0.83) above baseline risk factors (AUC=0.78 and AUC=0.81, respectively); the addition of microbial factors further improved models of obese asthma (non-obese asthma: $p=0.65$; AUC=0.79; obese asthma $p=0.002$; AUC=0.86). **Conclusions**: Our results support that genetic and microbiota characteristics are independently associated with obese asthma in adults. While we observed some microbiota characteristics associated with non-obese asthma, the associations were strongest in asthma comorbid with obesity.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1580 Human body fat distribution genes affect adipocyte function via Wnt signaling and mitochondrial activity: a systems genetics analysis.

Authors:

J. Reed[1], J. Huang[2], Y. Li[2], L. Ma[3], D. Banka[1], T. Wang[2], W. Ding[2], J. L. Bjorkegren[3,4], M. Civelek[1]; [1]Univ. of Virginia, Charlottesville, VA, [2]Novo Nordisk Res. Ctr. China, Beijing, China, [3]Icahn Sch. of Med. at Mount Sinai, New York, NY, [4]Karolinska Inst.t, Stockholm, Sweden

BACKGROUND: Excess fat in the abdomen is a sexually dimorphic risk factor for cardio-metabolic disease. The relative storage between abdominal and lower-body subcutaneous adipose tissue depots is approximated by the waist-to-hip ratio adjusted for body mass index (WHRadj). Genome-wide association studies (GWAS) have identified ~495 genes (346 loci) associated with WHRadj. Most of these genes have unknown roles, but many are expressed and putatively act in adipose tissue. We aimed to identify novel sex- and depot-specific drivers of WHRadj using a systems genetics approach.
METHODS: We used two independent cohorts of adipose tissue gene expression data (STARNET, GTEx) with 362 - 444 males and 147 - 219 females, primarily of European ancestry. Sex- and depot-specific Bayesian networks were used to model the gene-gene interactions from 8,492 adipose tissue genes. Key driver analysis identified genes that, *in silico* and putatively *in vitro*, regulate many others. Key driver gene function was determined by perturbing their expression in human subcutaneous pre-adipocytes.
RESULTS: A total of 51 - 119 key drivers in each network were replicated in both cohorts. We used single-cell expression data to select replicated key drivers expressed in adipocytes, prioritized genes unstudied in adipose tissue, and integrated public datasets to nominate 53 novel key driver genes (10 - 21 from each network) that may regulate fat distribution by altering adipocyte function. In other cell types, 23 of these genes are found in crucial adipocyte pathways - Wnt signaling or mitochondrial function. We selected eight genes whose expression is highly correlated with WHRadj to study their effects on adipogenesis/Wnt signaling (*ANAPC2, ANTXR1, PSME3, RSPO1, TYRO3*) or mitochondrial function (*C1QTNF3, MIGA1, PSME3, UBR1*).
Adipogenesis was inhibited in cells overexpressing *ANAPC2* and *RSPO1* compared to controls, consistent with a third cohort (METSIM) that shows higher expression is associated with higher WHRadj, and thus lower relative storage in the subcutaneous depot. *RSPO1* inhibited adipogenesis by increasing β-catenin activation and Wnt-related transcription, repressing *PPARG* and *CEBPA*. *PSME3* overexpression led to more adipogenesis than controls, inconsistent with association data. In mature adipocytes, *MIGA1* and *UBR1* downregulation led to mitochondrial dysfunction, with lower oxygen consumption than controls; additional research (e.g., deeper phenotyping of cells from both depots) is needed to support that lower expression is associated with higher WHRadj.
SUMMARY: *ANAPC2, MIGA1, PSME3, RSPO1,* and *UBR1* affect adipocyte function and may drive body fat distribution.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1581 Identification and functional characterization of rare mutations of the SYNGR family genes associated with schizophrenia.

Authors:

M-C. Cheng, S-H. Hsu, H-Y. Tsai; Yuli Branch, Taipei Veterans Gen. Hosp., Hualient Country, Taiwan

Schizophrenia is a severe psychiatric illness experienced by approximately 1% of individuals worldwide and has a debilitating impact on perception, cognition, and social function. The heritability of schizophrenia is estimated at 80% on average. Synaptic vesicle-related genes significantly regulate synaptic transmission and play a critical role in various psychiatric diseases. Synaptogyrin (SYNGR) gene family are the transmembrane protein genes of presynaptic vesicles that warrant genetic and functional analysis for the pathogenesis of schizophrenia. For pathogenic mutation identification, we sequenced the exonic regions of the *SYNGR* family genes in 516 unrelated patients with schizophrenia from Taiwan. We analyzed the Synaptogyrin protein functions of the identified protein-altering mutants via bioinformatic analysis, immunoblotting, exocytosis assay, and mass spectrometry analysis. After the resequencing, 17 protein-altering variants were identified, including 5 in the *SYNGR1* gene, 4 in the *SYNGR2* gene, 3 in the *SYNGR3* gene, and 5 in the *SYNGR4* gene. Among these mutations, 15 had minor allele frequencies (MAFs) less than 0.5% in the gnomAD and Taiwan BioBank database. Two mutations (*SYNGR1*$^{p.Tyr196Ter}$ and *SYNGR3*$^{p.Pro197Thr}$) were not observed among 1,517 healthy controls from Taiwan BioBank and the gnomAD database. The immunoblot assay demonstrated that the *SYNGR1*$^{p.Tyr196Ter}$ mutant as a loss-of-function mutant in the cultured cells. In the exocytosis assay, we found that overexpressing *SYGRN1*^{WT} and *SYNGR3*^{WT} inhibited Human growth hormone 1 (hGH1) secretion in PC12 cells, whereas two mutations (*SYNGR1*$^{p.Thr144Met}$ and *SYNGR1*$^{p.Tyr196Ter}$) did not inhibit hGH1 secretion. Mass spectrometry assay showed that the phosphorylated serine^{210} site of SYNGR3 protein was identified in wild-type (*SYNGR3*^{WT}) but not in three mutant-types (*SYNGR3*$^{p.Ser190Arg}$, *SYNGR3*$^{p.Pro197Thr}$, and *SYNGR3*$^{p.Ser210Asn}$). The results suggest that the SYNGR family genes, especially two neuronal genes (*SYNGR*1 and *SYNGR3*), harbor rare and functional disrupting mutations in some patients with schizophrenia, supporting contributing rare coding variants to the genetic architecture of schizophrenia. Of particular interest is that *SYNGR1* gene mutations might be associated with exocytosis, while *SYNGR3* gene mutations might be involved in regulating protein kinase C function.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1582 Identification of 39 Gene-Lifestyle interaction Loci for cardio-metabolic traits in Sub-Saharan African populations: An AWI-Gen study

Authors:

P. Boua[1,2], J-T. Brandenburg[2], S. Singh[2], H. Sorgho[1], A. Choudhury[2], H. Tinto[1], M. Ramsay[2], as members of AWI-Gen, and the H3Africa Consortium; [1]Clinical Res. Unit of Nanoro (CRUN), Inst. de Recherche en Sci. de la Santé (IRSS), Nanoro, Burkina Faso, [2]Sydney Brenner Inst. for Molecular BioSci., Faculty of Hlth.Sci., Univ. of the Witwatersrand, Johannesburg, South Africa

Gene-environment (GxE) interaction refers both to the modification of genetic risk factors by environmental risk and protective factors and to the role of specific genetic risk factors in determining individual differences in vulnerability to environmental risk factors. Gene-environment interactions affect cardiometabolic traits and can help identify novel pathways of diseases susceptibility. Several studies have identified environmental or behavioural factors that influence the genetic risk associated to cardiometabolic traits, with few conducted in Africa. In a sub-Saharan African population cohort (AWI-Gen study), we investigated the gene-lifestyle interaction of smoking and alcohol consumption on lipid and blood pressure traits. Participants were genotyped using the H3Africa genotyping array (2.3M SNPs), followed by imputation using the Sanger Imputation Server (African panel). GxE analyses were performed, using LinGxEScanR v1.0 (https://github.com/USCbiostats/LinGxEScanR/tree/CHARGE), in this resident African population-based cohort of over 9000 middle-aged adults. Exposure to smoking tobacco was defined by current smoking status and intensity (cigarette pack years and cigarettes per day), and alcohol consumption was defined by current drinking and intensity (heavy, light, never and the number of standard drinks). We found 39 significant loci for gene-alcohol interaction for blood pressure traits (3 Loci), gene-alcohol interactions for lipid (6 loci) and gene-smoking interaction for blood pressure traits (21 Loci). The identified loci were associated with exposure to alcohol and smoking, but also to the intensity of the exposure (standard drinks, heavy drinker, and cigarettes per year). There were also shared loci for blood pressure traits and lipid traits. The post-GWAS analysis revealed that some of the genes involved in the interaction were previously reported as markers of exposure, showing differential methylation in smoking. Most variants were population-specific and had very low allele frequencies in non-African populations. Inclusion of under-represented populations in medical genomics research is essential for advancing an understanding of complex trait genetics.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1583 Identification of an East Asian-specific variant associated with Lewy bodies dementia by genome-wide association study in Japanese subjects.

Authors:

R. Mitsumori[1], Y. Asanomi[2], D. Shigemizu[3,4], S. Akiyama[1], T. Morizono[1], S. Niida[1], K. Ozaki[5,4]; [1]Natl. Ctr. for Geriatrics and Gerontology, Obu, Japan, [2]Natl. Ctr. for Geriatrics and Gerontology, Obu City, Aichi Prefecture, Japan, [3]Natl. Ctr. for Geriatrics and Gerontology, Obu, Aichi, Japan, [4]RIKEN Ctr. for Integrative Med. Sci., Yokohama, Japan, [5]The Natl. Ctr. for Geriatrics and Gerontology, Obu, Aichi, Japan

Dementia with Lewy bodies (DLB) is the second most common clinically neurocognitive disorder among the elderly, following Alzheimer's disease (AD). The current therapy is limited to symptomatic and supportive care. The hallmarks of DLB are α-synuclein neuronal inclusions (Lewy bodies and Lewy neurites) that induce neuronal loss and deposit in the brain cortex and limbic system. DLB results from interactions among multiple genetic and environmental risk factor, none of which can cause disease solely by each of themselves. Recent large-scale genetic analysis such as genome-wide association study (GWAS) and genome sequencing analysis with Caucasian descent have identified several susceptibility loci for DLB. Whereas reports of large-scale genetic analysis with Asian are relatively modest. To clarify genetic architecture in the pathogenesis of DLB, we conducted GWAS in the Japanese population comprising of 221 DLB and 6,113 controls recruited at the National Center for Geriatrics and Gerontology (NCGG) biobank using ethnicity-specific genotyping array, Asia screening array. We identified a new East Asian-specific genetic risk variant for DLB with GWAS significance ($P < 5.0\times10^{-8}$) in the Japanese GWAS. We also replicated *APOE* locus, a known risk locus for DLB and AD in Caucasians. We found that the East Asian specific DLB associated variant affected the expression of the closest gene by assessing the expression quantitative trait loci analysis in whole blood. We performed trans-ethnic meta-analysis with the GWAS data from the NCGG biobank, UK biobank and another Caucasian cohort described elsewhere (Chia et al. Nat. Genet. 2021). In the meta-analysis, *APOE* and *SNCA* loci were successfully replicated with GWAS significance. We are further conducting the transcriptome-wide association analysis (TWAS), mendelian randomization analysis, construction of polygenic risk score (PRS) and genetic correlation analysis, which could contribute to better understanding of the polygenic mechanism for DLB.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1584 Identification of cells, genes, and regulatory elements affecting human kidney function.

Authors:

G. Loeb[1], P. Kathail[2], R. Shuai[2], R. Chung[2], R. Grona[1], S. Peddada[1], S. Volkan[3], S. Federman[1], K. Mader[1], A. Chu[4], J. Davitte[5], S. Shafer[3], L. Przybyla[1], R. Rapiteanu[6], J. Reiter[1]; [1]UCSF, San Francisco, CA, [2]UC Berkeley, Berkeley, CA, [3]GSK, San Francisco, CA, [4]GSK, Jamaica Plain, MA, [5]GSK, Exton, PA, [6]GSK, Stevenage, United Kingdom

Background The molecular mechanisms underlying human kidney disease remain largely unknown. Genome-wide association studies (GWAS) have identified hundreds of genomic loci associated with kidney function and disease. However, the genetic variants and genes mediating the effect of most of these loci remain unclear. **Methods** We performed a GWAS of kidney function estimated using multiple biomarkers in the UK Biobank and used functionally informed fine-mapping to identify putative causal kidney function variants. To determine the effect of genetic variants implicated in kidney function, we used scATAC-seq in human kidneys, genome-wide measurements of H3K27 acetylation (CUT&RUN), and measured the effect of CRISPR-mediated perturbation of regulatory elements on gene expression.
Results We found that 58% of kidney function SNP-heritability localized to candidate regulatory elements of kidney tubule epithelial cell types, 10% localized to podocyte-specific regulatory elements, and <1% localized within endothelial, stromal, or immune cell-specific regulatory elements. We identified putative causal kidney function variants using functionally-informed fine-mapping and used these variants to identify regulatory elements and genes involved in kidney function. In human kidneys and primary tubule epithelial cells, we assessed how kidney function variants affect chromatin accessibility, enhancer activation, and enhancer function. We found that kidney function variants alter chromatin accessibility and regulatory element function within tubule epithelial regulatory elements. A pooled screen targeting kidney function noncoding regulatory elements with CRISPR interference (CRISPRi) identified novel genes involved in human kidney function. **Conclusion** By integrating human genetics and studies of enhancer function, this work provides a framework for identifying variants, regulatory elements, and genes involved in human kidney disease. The combination of fine-mapping of GWAS-nominated variants, regulatory element identification, and mapping enhancers to regulated genes provides a framework for moving from GWAS to molecular mechanisms of disease.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1585 Identification of circulating protein biomarkers for childhood obesity using Mendelian randomization

Authors:

R. AVOCEGAMOU[1], N. Yazdanpanah[2], M. Yazdanpanah[2], D. Manousaki[3]; [1]Univ. of Montreal, Montreal, QC, Canada, [2]CHU Sainte-Justine, Montréal, QC, Canada, [3]Sainte Justine Hosp. Univ. of Montreal, Montreal, QC, Canada

BACKGROUND: Childhood obesity is a major public health problem affecting one in 10 youth and is associated with several short and long term cardiometabolic complications. Characterisation of early biomarkers for pediatric obesity is necessary to develop screening tools, to prioritize children at risk for preventive interventions, while these biomarkers represent potential therapeutic targets. Here, our goal was to identify circulating proteins with causal effects on childhood obesity, leveraging proteomic and pediatric body mass index (BMI) GWAS data, in a two-sample Mendelian randomization (MR) study.**METHODS**: We identified genome-wide significant (P < 5x10-8) polymorphisms (SNPs) associated with circulating levels of 1,297 proteins in 5 large European GWAS (Sun et al, Emilsson et al, Folkersen et al, Yao et al , Suhre et al). We obtained the effects of SNPs on the level of each protein from these GWAS and their effects on pediatric BMI from a European GWAS by Vogelezang et al. (N=61,111 children). Subsequently, we estimated the causal effect of circulating proteins on BMI by calculating the Wald ratio for proteins with a single SNP-instrument or by meta-analyzing the effects of multiple SNP-instruments per protein on pediatric BMI using the Inverse Variance Weighted (IVW) approach. We next applied genetic colocalization for the proteins prioritized by our MR study.**RESULTS**: Among 1068 proteins tested (with SNPs available in Suhre et al, Sun et al, Folkersen et al, Yao et al et Emilsson et al GWAS), we identified two proteins whose MR effect (Wald ratio) had a P-value below the Bonferroni-corrected threshold (<2.4 x10-4). These are the ENG (endoglin; MR beta: -0.07, 95% CI= -0.10, -0.05, P= 4.4 x10-5) and the FABP4 (fatty acid binding protein 4, MR beta: - 0.33 95% CI -0.5, -0.16, P=1.3 x10-4), both having a negative effect on BMI per standard deviation increase in their serum level. Both proteins colocalized with childhood BMI (posterior probabilities for shared causal variant 95.9% for ENG and 75.9% for FABP4). ENG is expressed in the vascular wall, while FABP4 has a known role in fatty acid transport and metabolism. Neither ENG nor FABP4 have been previously associated with adult or pediatric BMI in GWAS.**CONCLUSION**: We identified two circulating proteins as candidate causal biomarkers of pediatric obesity. MR studies are ongoing to test specific proteins prioritized by previous MR studies as biomarkers of adult BMI for causal effects on pediatric BMI. Upon the number of identified proteins, we will use the candidate proteins to identify protein-protein interaction networks that will point to predominant pathophysiological pathways in pediatric obesity.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1586 † Identification of insulin trafficking and secretion genes by a targeted prime editing screen

Authors:

V. Parsons, S. Vadlamudi, A. S. Moxley, A. Etheridge, J. D. Rosen, K. A. Broadaway, K. Mohlke; UNC Chapel Hill, Chapel Hill, NC

Altered proinsulin secretion due to incomplete insulin processing has been linked to pancreatic islet beta-cell stress, but the genes involved are incompletely understood. A recent genome-wide association study (GWAS) meta-analysis of plasma proinsulin levels identified 36 independent signals at 30 loci. To elucidate the underlying genes and variants driving association signals, we are performing a targeted CRISPR prime editing screen. We selected 11 genes at eight loci to knock out in INS1 832/13 rat insulinoma cells; genes were selected based on pancreatic islet expression, colocalized islet eQTL, and putative function. We also chose to edit the lead GWAS variants at two of these loci (*TBC1D30, SGSM2*) because these variants encode amino acid substitutions. For each target, we designed six to eight prime editing and secondary nicking guide combinations. We observed editing efficiencies of ~50-80% for four genes (K*ank1, Madd, Sgsm2 knockout, and Lrrc49*), ~20% for one gene (*Dlc1*), and <2% for one gene (*Larp6*) and the *Sgsm2* substitution; editing of four additional targets and repeated targeting of *Larp6* and the *Sgsm2* variant are underway. For targets with ≥30% editing efficiency, we will isolate clonal lines and assay proinsulin secretion. The first target we investigated was a proinsulin-associated locus on chr12 for which the lead variant rs150781447 encodes an Arg279Cys substitution in *TBC1D30* that is predicted to damage protein function. We generated 14 clonal lines with 47% *Tbc1d30* expression and compared them to 14 mock-edited control lines. When stimulated with glucose to induce proinsulin secretion, *Tbc1d30* knockdown cells secreted 1.6-fold more proinsulin ($p=2.1x10^{-3}$) compared to mock-edited controls. These results validate effects on proinsulin secretion of two sets of mutant lines created by CRISPR editing with double strand breaks. For the *Tbc1d30* Arg279Cys variant, we generated seven Arg/Cys and four Cys/Cys lines and compared them to mock-edited Arg/Arg lines. The Arg/Cys lines secreted 2.3-fold more (p=0.034) and Cys/Cys lines secreted 2.2-fold more (p=0.037) proinsulin, respectively, than mock-edited lines. These data are concordant with the GWAS direction of effect, support the hypothesis that rs150781447 may be the causal variant at this locus, and the similar effect size of for Arg/Cys and Cys/Cys lines on proinsulin secretion highlights the challenges of modeling variants with modest effects. In summary, we have established a prime editing strategy to functionally identify genes and variants at GWAS loci and demonstrated a role for *TBC1D30* in proinsulin secretion, suggesting it is responsible for the chr12 GWAS locus.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1587 Identification of microbiota features correlated with host lifestyle and molecular, biochemical, and immunophenotypic measurements in a deeply phenotyped Sardinian cohort.

Authors:

M. Diana[1,2], M. Pitzalis[2], M. Masala[2], A. Maschio[2], F. Busonero[2], M. Lobina[2], M. Devoto[2,3], M. Gorospe[4], D. Schlessinger[4], V. Orrù[2], E. Fiorillo[5], F. Santoni[6,2], F. Cucca[2,1], M. Pala[2]; [1]Dipartimento di Scienze Biomediche, Università di Sassari, Sassari, Italy, [2]Istituto di ricerca genetica e biomedica (IRGB), CNR, Monserrato, Italy, [3]Dipartimento di medicina traslazionale e di precisione, Università Sapienza, Roma, Italy, [4]Lab. of genetics and genomics, NIH, Baltimore, MD, [5]Istituto di ricerca genetica e biomedica (IRGB), CNR, Lanusei, Italy, [6]Ctr. Hosp.ier Univ.ire Vaudois (CHUV), lausanne, Switzerland

Extensive literature exists on the association between the microbiota and various host phenotypic traits, lifestyles, and diseases. These studies have limitations including small sample sizes, reliance on non-whole-genome technologies (e.g., 16S rRNA), limited capacity to correlate metagenomic data with a large number of concurrently measured characteristics in the same individuals, and predominant use of observational or case-control experimental designs, which are susceptible to experimental confounders and reverse causality effects. To address these limitations, we propose leveraging metagenomic data obtained through whole-genome shotgun sequencing of fecal microbiota from 2,700 volunteers in the ProgeNIA cohort. Our objective is to establish correlations between these data and a diverse range of quantitative, qualitative, and genetic measurements available within the same cohort. These measurements encompass biochemical blood markers, anthropometric traits, disease status, lifestyle factors (such as alcohol consumption, coffee intake, cigarette smoking, and physical activity), immunophenotypic measurements (including cellular composition and protein expression of surface markers), leukocyte transcriptional profiling, and high-resolution genetic characterization (involving approximately 25 million variants, including SNPs and INDELs). The ProgeNIA cohort uniquely facilitates the study of the microbiota's impact on a broad spectrum of host characteristics. Thus far, we have quantified 5,566 taxonomic levels, encompassing 22 phyla, 139 classes, 173 orders, 227 forms, 895 genera, 1,724 species, and 2,317 taxa. Preliminary findings have identified more than 100 taxonomic levels displaying significant correlations with factors such as alcohol consumption (padj<1.7E-14), smoking (padj=4.2E-08), Proton Pump Inhibitors (PPI) (padj=1.7E-09), BMI (padj=2.9E-06), anti-diabetics intake (padj=2.2E-07), cancer (padj=7E-04), diabetes (padj=8.7E-07), glycemia (padj=1.4E-03), sex (padj=3E-09) and age (padj=5.8E-09). Our study demonstrates both agreement and disagreement with existing literature, providing a better resolution of the taxa involved.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1588 Identification of rare functional genetic variants and cell-type gene regulatory networks in Progressive Supranuclear Palsy

Authors:

J. Grundman[1], T. Chang[1], R. Kawaguchi[1], L. Qi[1], L. Chen[1], K. Gao[1], J. Ou[1], J. Nguyen[1], M. DeTure[2], B. Rawlinson[2], C. L. Dalgard[3], H. Morris[4], D. Dickson[2], T. Li[5], A. Battle[5], W-P. Lee[6], H. Wang[6], Y. Leung[6], L-S. Wang[6], G. Schellenberg[6], D. Geschwind[1]; [1]Univ. of California, Los Angeles, Los Angeles, CA, [2]Mayo Clinic, Jacksonville, FL, [3]Uniformed Services Univ. of the Hlth.Sci., Bethesda, MD, [4]UCL Inst. of Neurology, London, United Kingdom, [5]Johns Hopkins Univ., Baltimore, MD, [6]Univ. of Pennsylvania, Philadelphia, PA

Progressive supranuclear palsy (PSP) is a rare, relatively homogeneous neurodegenerative tauopathy with some known common and rare genetic risk factors and proposed cell type and regional vulnerability patterns. Here, we aimed to use paired RNA-sequencing to inform interpretation of whole genome sequencing (WGS) based on outlier calls across both gene expression and alternative splicing. Specifically, we employ a recently developed framework, Watershed (Ferraro et al., 2020), that leverages transcriptomic outliers and genomic annotations to construct a Bayesian model that scores the likelihood of rare functional genetic variants within +/- 10 kb of outlier genes. Our cohort is comprised of over 300 European-ancestry PSP patients and age- and sex-matched controls who have undergone both WGS and bulk RNA-sequencing across prefrontal cortex and caudate nucleus. All samples have identity by descent (IBD) < 0.25. We additionally examine the association of common and rare genetic risk variants in the context of cell-type gene regulatory networks through the use of single-nucleus RNA-sequencing and ATAC-sequencing across both prefrontal cortex and caudate nucleus in a subset of our samples (N=23 per region). Identifying additional genetic risk variants and specific cell type-specific mechanisms related to PSP will further our understanding of causative disease mechanisms and provide new avenues for therapeutic developments. More broadly, such integration of transcriptome data with rare variant data is a potentially powerful means identify new risk variants and genes.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1589 Identifying causal genes modulated by rare regulatory variants in Alzheimer's disease.

Authors:

T. Patel; Icahn Sch. of Med. at Mount Sinai, New York, NY

Genetic studies have identified more than 75 loci associated with Alzheimer's disease (AD) risk to date, however most are localized to non-coding regions of the genome, making it difficult to pinpoint the causal genes and pathways. We have previously shown that common AD risk variants are enriched in active enhancers of myeloid cells, which likely modify disease risk by regulating gene expression and, in turn, function of these cells. However, the impact of rare non-coding variation on myeloid gene expression and function remains to be elucidated. We intersected whole-genome sequencing (WGS) data from 6,161 AD cases and 6,502 controls with microglial epigenomic annotations (H3k27Ac ChIP-seq) and chromatin interactions (PLAC-seq) (from Nott et al., 2019) to identify rare variants (MAF < 1%) that fell within these cell-type-specific regulatory elements. Using GENESIS, we performed single variant and set-based association tests by aggregating variants to the gene regulatory elements (GRE) they were located in, and GRE-target genes based on the chromatin interactions. Variants were assigned weights using functional impact scores (e.g. disruption to transcription factor binding sites, DNase hypersensitivity sites) from various databases including RegulomeDB. Preliminary analyses identified several rare variants significantly associated with AD, including variants that were in GREs interacting with known myeloid expressed gene promoters CX3CR1 ($P=3.2x10^{-10}$) and IL4R ($P=7.5x10^{-10}$), and potentially novel genes of interest such as RIPK2 ($P=3.8x10^{-8}$). Aggregating variants to GRE-target gene level identified clusters of significant genes in the VEGFA-TMEM63B locus ($P=1.6x10^{-7}$) that appear to be regulated by the same GREs. We plan to replicate our analyses using the larger AD case-control 36k WGS data to increase power to detect additional disease-relevant regulatory variants and their target genes in microglia. Integrative approaches have already nominated candidate AD risk genes whose expression may be modulated by common AD risk variants in myeloid cells. By investigating the contribution of rare regulatory variation in AD, we can better understand the mechanisms through which these rare variants regulate gene expression and disrupt gene regulatory networks in disease-relevant cell types.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1590 Identifying individuals at risk for surgical supravalvar aortic stenosis in Williams-Beuren syndrome by polygenic score

Authors:

D. Liu[1], E. Ruiz Escobar[1], N. Raja[2], M. Levin[2], E. Biamino[3], M. Bedeschi[4], M. Digilio[5], G. Squeo[6], R. Villa[4], S. Osgood[2], J. Freeman[1], G. Merla[7], B. Pober[8], C. Mervis[9], A. Roberts[10], C. Morris[11], L. Osborne[12], B. Kozel[13]; [1]Natl. Heart, Lung, and Blood Inst., Bethesda, MD, [2]NIH, Bethesda, MD, [3]Univ. of Turin, Turin, Italy, [4]Fondazione IRCCS Ca Granda Ospedale Maggiore Policlinico Med. Genetic Unit, Milan, Italy, [5]Bambino Gesù Children's Hosp., Rome, Italy, [6]Fondazione IRCCS Casa Sollievo della Sofferenza, San Giovanni Rotondo, Italy, [7]IRCCS Casa Sollievo Della Soff, San Giovanni Ro, Italy, [8]Massachusetts Gen. Hosp., Boston, MA, [9]Univ. of Louisville, Louisville, KY, [10]Boston Children's Hosp., Boston, MA, [11]Univ Nevada Sch Med, Las Vegas, NV, [12]Univ Toronto, Toronto, ON, Canada, [13]NIH, Bethesda, MD

Introduction: Supravalvar aortic stenosis (SVAS), a narrowing of the aorta above the aortic valve, is a characteristic feature of Williams-Beuren syndrome (WBS). SVAS is present in about two-thirds of those with WBS, but severity varies; ~20% of people with WBS have SVAS requiring surgical intervention in their early childhood, while ~35% have no appreciable SVAS and ~45% have mild or intermediate narrowing. Recently, we reported a new method for identifying genetic modifiers of SVAS outcomes. However, until now a clinically usable scoring system for early identification of children with WBS at risk of surgical SVAS has not been available. **Methods:** In our previous modifier study evaluating 217 individuals with WBS (87 surgical SVAS and 130 no SVAS), we identified 360 genes (containing a total of 427 common non-synonymous variants; [hereafter variants]) in 13 pathways that were associated with extreme SVAS outcomes: surgical SVAS *vs.* no SVAS. These pathways include innate/adaptive immune and matrisome. We first assessed whether the polygenic score (PGS) calculated using 361 of the variants (after clumping) in the extreme phenotype subgroups (87 surgical SVAS and 130 no SVAS) could be useful for classification of SVAS severity. We further assessed whether the same set of variants was also able to distinguish surgical SVAS (n=87) from mild/intermediate stenosis (n=184) by the PGS. **Results:** When 90% of the samples are used for training, and 10% of the samples for testing, the classification accuracy of the PGS measured by the area under the curve (AUC) is 0.97 for surgical SVAS vs no SVAS and is 0.89 for surgical SVAS vs mild/intermediate stenosis. **Conclusion:** The variants from 13 pathways described here not only provide insight into the molecular underpinnings of SVAS but can also be used to calculate a highly predictive PGS. Together, the combination of a strongly predictive PGS and the identification of key modifier pathways offers the previously inconceivable possibility of molecularly targeting the most severe form of SVAS before it becomes life threatening.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1591 Identifying the best performing Alzheimer's disease PRS with the aim of risk prediction and the role of age in the genetic architecture.

Authors:

Y. Abakkouy[1], E. Luckett[1], S. Becelaere[1], R. Vandenberghe[1,2], I. Cleynen[1]; [1]KU Leuven, Leuven, Belgium, [2]UZ Leuven, Leuven, Belgium

Polygenic risk scores (PRS) combining the effect of disease-associated variants offer possibilities for risk prediction, also in Alzheimer's disease (AD). However, there are many PRS tools available and there is no consensus on how to model a PRS for AD in terms of f.e. p-value threshold (pT) settings for SNP inclusion and how to best deal with the strong *APOE* effect.

We designed a PRS-pipeline comparing PRSice and Lassosum for different parameter settings and PRS models: PRS_{AD} (incl all SNPs), PRS_{noAPOE} and PRS_{APOE} (excl and only incl *APOE*), $APOE_{\varepsilon2+\varepsilon4}$ (incl the two major *APOE* variants), and the weighted sum of PRS_{noAPOE}+$APOE_{\varepsilon2+\varepsilon4}$. Scores were built from summary statistics of Kunkle et al (Nat Genet 2019) and Bellenguez et al (Nat Genet 2022) to test the effect of the inclusion of proxy cases in the latter. PRS were derived from and standardized against the 1000G European dataset, and built on a publicly available test dataset (ADNI, N_{case}=241, $N_{control}$=256) from which we selected the best performing PRS model for each summary statistic based on variance explained (R2). This best model was then validated on an in-house AD case-control dataset (N_{case}=179, $N_{control}$=98).

We found that PRSice and the PRS_{noAPOE}+$APOE_{\varepsilon2+\varepsilon4}$ model performed best for both the Kunkle- and Bellenguez-based scores (R2=25.1%, AUC=0.64, pval=2.7e-19; R2=21.9%, AUC=0.65, pval=1.7e-17 respectively). The pT for SNP inclusion for the best Kunkle-based score was 1; while the pT for the best Bellenguez-based score was 1e-05. The best Kunkle score had an accuracy of 70.4% (sensitivity=67.0%, specificity=76.5%). The best Bellenguez score had an accuracy of 73.7% (sensitivity=69.3%, specificity=81.6%). The two scores showed a correlation coefficient of 77% (p<2.2e-16) but performed poorly in identifying the same individuals in the top/bottom 5% PRS (28.6% overlap). The best Kunkle score was correlated with age at diagnosis (R=0.17, p=0.02), while the best Bellenguez score was not (R=0.11, p=0.2). Considering the known association between APOE and age at onset, we also tested the PRS_{noAPOE} for correlation with age, and found no significant correlations. *APOE4* carriers however had a higher Kunkle-based PRS_{noAPOE} with pT=1 than non-carriers (p=0.03, b=0.56, se=0.25), which was not the case for a Bellenguez-based PRS_{noAPOE} with pT=1e-05 (p=0.9).

Our analyses confirm that a PRS in which the two major APOE variants are modeled separately outperforms any other PRS model tested, irrespective of GWAS base set used. Other analyses however show inconsistent results depending on GWAS base set used, with fe Kunkle scores suggesting a polygenic architecture, and Bellenguez scores an oligogenic architecture.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1592 Identifying the Genes Responsible for Natural Genetic Resistance to Type 1 Diabetes in a Middle Eastern Cohort

Authors:

A. Akil[1], I. Ahmed[1], B. Ajaz[1], K. Fakhro[2], W. Hagopian[3], R. Oram[4]; [1]Sidra Med., Doha, Qatar, [2]Sidra Med., Doha, Qatar, Qatar, [3]Pacific Northwest Res. Inst., Seattle, WA, [4]Univ. of Exeter, Exeter, United Kingdom

Background: Specific HLA alleles have been linked to autoimmune diseases including susceptibility to type 1 diabetes (T1D). The T1D risk has been shown high for certain HLA class II genes while others provide protection from disease development. Defining the underlying mechanisms of DQB1*06:02-mediated protection will help identify the molecular, metabolic, and immunologic pathways responsible for T1D pathogenicity.
Aim: Fine mapping of the HLA alleles and haplotypes from 15k whole genomes of Qatar Genome subjects. Assessment of the risk and protective HLA alleles that exist in Qatari population associating with an increased or a lower risk of T1D development.
Methods: HLA type inference was performed using multiple independent typing methods with high accuracy on the WGS data and association with clinical traits of T1D was performed using linear regression models.
Results: We found a high diversity of rare alleles among class II HLA genes in our population. We further tested the association of observed HLA alleles with five clinically relevant traits of T1D and identified several potentially protective and risk alleles for the disease. We particularly analyzed the associations of the HLA-DQB1*06:02 allele and found that the protective effects of DQB1*06:02 could be gender specific in our population. Compared to males, homozygous females for DQB1*06:02 allele are likely to have greater levels of hemoglobin A1c (HbA1c), glucose and thyroid stimulating hormone (TSH), some of the clinical traits that are associated with an increased risk for T1D. Conversely, homozygous males have significantly higher levels of Insulin and C-peptide, the clinical traits that are associated with a lower risk for T1D.
Conclusion: Multiple alleles from genes DRB1, DQA1 and DQB1 which are known to segregate with T1D predisposition showed a significant association with clinical phenotypes of T1D suggesting a greater genetic susceptibility for T1D in the general population.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1593 Impact of genome-wide rare copy number variants on autism spectrum disorders.

Authors:

Z. He[1,2], K. S. Benke[2,3], J. Daniels[2,3], V. Morrill[1,2], M. Fallin[4], H. Volk[2,3], C. Ladd-Acosta[1,2,3]; [1]Dept. of Epidemiology, Bloomberg Sch. of Publ. Hlth., Johns Hopkins Univ., Baltimore, MD, [2]Wendy Klag Ctr. for Autism and Dev.al Disabilities, Bloomberg Sch. of Publ. Hlth., Johns Hopkins Univ., Baltimore, MD, [3]Dept. of Mental Hlth., Bloomberg Sch. of Publ. Hlth., Johns Hopkins Univ., Baltimore, MD, [4]Rollins Sch. of Publ. Hlth., Emory Univ., Atlanta, GA

Background: Rare single nucleotide variants have been identified and common single nucleotide polymorphisms (SNPs) from genome-wide association studies continue to emerge as autism spectrum disorder (ASD) risk factors. Rare copy number variants (CNVs) have also been implicated in the genetic etiology of neurodevelopmental disorders but further discovery and characterization in US ASD populations is needed. Thus, we leveraged 1005 participants in Study to Explore Early Development (SEED) 1 and 3 to examine the association of rare CNVs with ASD. **Methods:** Genotyping was performed on Illumina Infinium HumanOmni1-Quad BeadChip, with over 1.1 million probes among a subset of SEED participants. We generated raw autosomal CNV calls using PennCNV (v1.0.5) and QuantiSNP (v2.3) and retained consensus CNV calls detected by both algorithms. Extensive sample and CNV level quality control was performed. We filtered CNV calls with minor allele frequency <1%, length ≥10 kilobases and <10 megabases, spanning ≥10 consecutive probes as confidence rare CNVs. We performed functional annotation using ANNOVAR. We examined CNV burden measures of total number/length of CNVs at four levels: genome-wide, genes, exons of 470 autistic genes, and functional categories. We used logistic regression to examine associations of CNV burden with ASD, adjusting for maternal and child characteristics, study site, wave, and principal components 1-10. **Results:** We identified 5804 rare CNVs: 2192 in 337 ASD cases (CASE), 3068 in 513 general population controls (POP), and 544 in 155 non-ASD developmental delay controls (DD). Genome-wide CNV burden combining copy number (CN) gain and loss was greater among ASD cases (CASE vs POP, odds ratio (OR) = 1.18, 95% confidence interval (CI): 1.01-1.39, P = 0.043; CASE vs POP & DD, OR = 1.16, 95% CI: 1.003-1.35, P = 0.045), measured as total length of CNVs. ASD cases had a higher burden of CNVs spanning exons of autistic genes for CN gain and loss combined (CASE vs POP, OR = 1.32, 95% CI: 1.01-1.73, P = 0.042; CASE vs POP & DD, OR = 1.34, 95% CI: 1.05-1.72, P = 0.019), and for CN loss (CASE vs POP & DD, OR = 1.93, 95% CI: 1.14-3.30, P = 0.015), measured as total length of CNVs. Functional annotation suggested a higher burden of exonic CNVs among ASD cases (CASE vs POP, OR = 1.20, 95% CI: 1.01-1.42, P = 0.041), measured as total length of CNVs. SNPs associated with ASD were enriched in CNVs in ASD cases compared to POP/DD. **Conclusion:** These results contribute to the growing evidence that rare CNVs, both genome-wide and at specific loci, contribute to risk of ASD. Rare CNVs are essential for understanding the genetic architecture of ASD and may provide novel genetic testing or therapeutic targets.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1594 Implementation of a cardiometabolic genetic testing panel facilitates diagnosis, intervention, and preventative care in a predominantly Hispanic population

Authors:

B. Yuan[1], L. Abushamat[2], E. Venner[1], D. Muzny[1], G. Metcalf[1], M. Murugan[1], S. Kadirvel[1], L. Li[1], J. Stephen[1], S. Xu[1], V. Yi[1], M. Sulit[1], C. L. Kovar[1], M. Gingras[1], J. Hu[1], V. Korchina[1], M. Varughese[2], C. Fordis[3], S. Pereira[4], R. Gaba[5], M. Changlani[6], S. Mulukutla[7], H. Silberman[6], J. E. Posey[8], L. R. Treviño[9], C. M. Ballantyne[2], R. A. Gibbs[1], A. Balasubramanyam[5]; [1]Human Genome Sequencing Ctr., Baylor Coll. of Med., Houston, TX, [2]Section of Cardiovascular Res., Dept. of Med., Baylor Coll. of Med., Houston, TX, [3]Inst. for Continuing Professional Dev. in Quality Improvement and Patient Safety, Baylor Coll. of Med., Houston, TX, [4]Ctr. for Med. Ethics and Hlth.Policy, Baylor Coll. of Med., Houston, TX, [5]Section of Diabetes, Endocrinology and Metabolism, Baylor Coll. of Med., Houston, TX, [6]DHR Hlth.Heart Inst., McAllen and Edinburg, TX, [7]Diabetes and Endocrinology Inst., DHR Hlth., Edinburg, TX, [8]Dept. of Molecular and Human Genetics, Baylor Coll. of Med., Houston, TX, [9]DHR Hlth.Inst. for Res. and Dev., Edinburg, TX

Background: Cardiometabolic diseases are the leading cause of mortality worldwide. The genetic etiology has been extensively studied, ranging from high-penetrance monogenic conditions to polygenic traits. However, genetic testing uptake in community settings has been low and notably lower in Hispanic populations, an underrepresented minority in genomic research.
Methods: We developed a genome sequencing (GS) based testing panel, targeting a majority Hispanic population in community-based cardiology and endocrinology clinics in the Rio Grande Valley, Texas, that interrogated 177 genes related to monogenic cardiometabolic conditions, two *LPA* polymorphisms (rs3798220 and rs10455872), and specific pharmacogenomic (PGx) alleles of *SLCO1B1* and *G6PD* related to simvastatin and rasburicase metabolism. Results were returned to referring physicians and reviewed during a virtual multi-disciplinary sign-out conference.
Results: As of April 2023, testing was completed for 694 individuals (291 females, 403 males, 18-92 years old, 92% Hispanic). 24 individuals (3.5%) were positive for a pathogenic or likely pathogenic variant in a monogenic disease gene, including 8 also positive for an LPA risk allele, conferring increased risk of comorbidities. 11 findings were considered diagnostic, whereas6 were nondiagnostic but provided valuable disease risk information. The remaining 7 findings were pending review with referring physicians. Most diagnostic findings were related to cardiomyopathies (*TTN* n=5, *MYH7* n=1, *MYL2* n=1, *TNNI3* n=1), followed by *LDLR*, *HNF1A* and *FBN1* (n=1, respectively). The nondiagnostic findings included *ABCC8* (n=2), *SDHA* (n=2), *MYH7* (n=1), and *SCN5A* (n=1), suggesting subclinical traits or incomplete penetrance. Additionally, 252 individuals (36.3%) were positive for at least one *LPA* risk allele, and 159 (22.9%) individuals were positive for at least one PGx allele.
Discussion: Implementation of a diagnostic and pharmacogenomic cardiometabolic gene panel signifies clinical utility in disease diagnosis, intervention, and preventative care for historically underserved populations in community settings. The diagnostic rate of this study (3.5%) is lower than the previous HeartCare project (9%) (PMID:34363016), potentially due to different clinical spectrums and settings (community-based vs academic), and underrepresentation of Hispanic populations in present genomic knowledgebases. Expanded GS analysis may increase diagnostic yield. Developing and implementing population-specific *LPA* risk alleles and polygenic risk scores may improve individual-level risk assessment and management.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1595 † Improving Polygenic Scores accuracy in non-European populations for lifetime disease risk estimation

Authors:

S. Kanoni[1], B. Jermy[2], B. Wolford[3], M. Kanai[4,5], K. Läll[6], Y. Wang[4,5], K. Zguro[7], Y. Cheng[8], Z. Yang[2], T. Hartonen[2], R. Monti[9], J. Wanner[2,9], O. Youssef[10,11], Genes & Health Research Team, Estonian Biobank Research Team, FinnGen, G. Matullo[12], C. Lippert[9,13], D. McCartney[8], C. Hayward[14], R. Marioni[8], S. Furini[7,15], A. Renieri[7,16,17], A. Martin[4,5], B. Neale[4,5], K. Hveem[3,18], R. Mägi[6], A. Palotie[2,4,5], Y. Okada[19,20,21,22], D. van heel[23], H. Heyne[2,9,13], N. Mars[2,4,5], A. Ganna[2,24], S. Ripatti[2,24,25]; [1]William Harvey Res. Inst., Barts and the London Sch. of Med. and Dentistry, Queen Mary Univ. of London, Cambridge, United Kingdom, [2]Inst. for Molecular Med. Finland, FIMM, HiLIFE, Univ. of Helsinki, Helsinki, Finland, [3]K. G. Jebsen Ctr. for Genetic Epidemiology, Dept. of Publ. Hlth.and Nursing, Faculty of Med. and Hlth., Norwegian Univ. of Sci. and Technology, Trondheim, Norway, [4]Analytic and Translational Genetics Unit, Massachusetts Gen. Hosp., Boston, MA, [5]Stanley Ctr. for Psychiatric Res. and Program in Med. and Population Genetics, Broad Inst. of MIT and Harvard, Cambridge, MA, [6]Estonian Genome Ctr., Inst. of Genomics, Univ. of Tartu, Tartu, Estonia, [7]Med Biotech Hub and Competence Ctr., Dept. of Med. Biotechnologies, Univ. of Siena, Siena, Italy, [8]Ctr. for Genomic and Experimental Med., Inst. of Genetics and Cancer, Univ. of Edinburgh, Edinburgh, United Kingdom, [9]Hasso Plattner Inst., Univ. of Potsdam, Digital Engineering Faculty, Potsdam, Germany, [10]Helsinki Biobank, Hosp. District of Helsinki and Uusimaa (HUS), Helsinki, Finland, [11]Pathology Dept., Univ. of Helsinki, Helsinki, Finland, [12]Dept. of Med. Sci., Univ. of Turin, Turin, Italy, [13]Hasso Plattner Inst. for Digital Hlth., Icahn Sch. of Med. at Mount Sinai, New York, NY, [14]Med. Res. Council Human Genetics Unit, Inst. of Genetics and Cancer, Univ. of Edinburgh, Edinburgh, United Kingdom, [15]Dept. of Electrical, Electronic and Information Engineering "Guglielmo Marconi", Univ. of Bologna, Bologna, Italy, [16]Med. Genetics, Univ. of Siena, Siena, Italy, [17]Genetica Medica, Azienda Ospedaliera Univ.ria Senese, Siena, Italy, [18]Levanger Hosp., Nord-Trøndelag Hosp. Trust, Levanger, Norway, [19]Dept. of Genome Informatics, Graduate Sch. of Med., the Univ. of Tokyo, Tokyo, Japan, [20]Dept. of Statistical Genetics, Osaka Univ. Graduate Sch. of Med., Osaka, Japan, [21]Lab. for Systems Genetics, RIKEN Ctr. for Integrative Med. Sci., Kanagawa, Japan, [22]Lab. of Statistical Immunology, Immunology Frontier Res. Ctr. (WPI-IFReC), Osaka Univ., Osaka, Japan, [23]Blizard Inst., Barts and the London Sch. of Med. and Dentistry, Queen Mary Univ. of London, London, United Kingdom, [24]Massachusetts Gen. Hosp. and Broad Inst. of MIT and Harvard, Cambridge, MA, [25]Dept. of Publ. Hlth., Univ. of Helsinki, Helsinki, Finland

Polygenic scores (PGS) estimate the cumulative genetic risk for common diseases but are mostly developed in Europeans leading to limited global clinical utility. Within the INTERnational consortium of integratiVE geNomics prEdiction (INTERVENE), we developed a novel framework to estimate country-, age- and sex-specific absolute lifetime risk across ancestries. We leveraged 9 biobanks, including 2 Asian biobanks (East and South Asian) and estimated absolute lifetime risks across PGS strata for 18 high global burden diseases. We integrated population-specific metrics from the Global Burden of Disease (GBD) and harmonized disease outcomes across biobanks. PGSs were computed with MegaPRS, using available GWAS summary statistics. Overall, we had a combined sample size of 1.38M (7.6% non-European), with extensive (up to 50 years) follow-up within health registries. We employed Cox proportional hazards models and GBD baseline risks to estimate overall, age- and sex-stratified PGS effects by ancestry per disease. We observed high variability in several diseases' prevalence. Overall cancer prevalence was >20% among Europeans and East-Asians but lower in South Asians (3.15%). Prevalence of type 2 diabetes (T2D) was 11.8% in Europeans, but twice as high in South-Asians and three times higher in East Asians. All PGSs were associated with the diseases in the European ancestry but only a fraction remained significant in the other ancestries. Similar to Europeans, PGS had a larger effect for coronary heart disease (CHD in South-Asian men compared to women (HR_{male} = 1.34 (95% CI: 1.29 - 1.39) vs HR_{female} = 1.20 (95% CI: 1.14 - 1.27), $PGS*Sex_{Interaction}$ P-value = $7.92x10^{-4}$), whereas for T2D the effect was larger in East Asian women compared to men (HR_{male} = 1.53 (95% CI: 1.50 - 1.55) vs HR_{female} = 1.62 (95% CI: 1.59 - 1.66), $PGS*Sex_{Interaction}$ P-value = $1.76x10^{-5}$). Varying effect sizes by age were detected for CHD, type 1 diabetes and T2D in South Asians, whereas T2D and prostate cancer exhibited age-specific effects in East Asians. We also observed a reduced cumulative incidence of T2D in East-Asians relative to Europeans; at age 80, cumulative incidence for men and women in the top 5% PGS was 37.46% (95% CI: 34.55% - 40.27%) and 28.37% (95% CI: 25.89% - 31.11%), respectively. The variability by ancestry in cumulative incidence was driven by baseline hazard variation. Our results demonstrate variable sex- and age-specific PGS associations across ancestries and emphasize the importance of calibrating lifetime risk estimates.

We aim to further improve PGS accuracy in non-Europeans by including primary care data and additional global biobanks for better cumulative risk assessment.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1596 Incorporating genetic and environmental variables to understand disease risk in The Biobank at the Colorado Center for Personalized Medicine.

Authors:

J. Shortt, M. Lin, M. Fisher, H. Smith, S. Chavan, N. Pozdeyev, N. Rafaels, K. Crooks, C. Gignoux; Univ. of Colorado Anschutz Med. Campus, Aurora, CO

Prediction of complex disease risk promises to be another important milestone in personalized medicine and population health but accurate prediction depends on sufficient knowledge of both genetic and non-genetic factors. Although environmental information is typically difficult to gather in contrast to genetics, as we build more complex models and try to capture more trait variance environmental information, and any interactions between, population-scale biobanks with linked electronic health records (EHR) provide opportunities to test how this can be accomplished. Here, we build a multi-modal dataset comprised of over 1500 phecodes derived from EHR data, over 2500 polygenic risk scores (PRS) calculated from dense genotyping, and both direct environmental variables (e.g., altitude) and indirect environmental variables (neighborhood social vulnerability and urban-rural commuting areas) in 73,346 participants in the Colorado Center for Personalized Medicine (CCPM) biobank. We identify hundreds of associations between phecodes and environmental variables at an FDR of 0.05, including known associations between elevation and carcinoma of skin (OR per elevation quantile [95% CI] = 1.19 [1.08-1.31], p ~ 0.0003). Additionally, we find associations between altitude and cardiovascular symptoms (OR = 1.35 [1.30-1.40], p ~ 1.46e-58), socioeconomic status and alcohol-related disorders (OR=6.3 [3.28-12.01], p ~ 2.65e-8), and urban-rural commuting areas and upper respiratory infections (OR = 0.86 [0.84-0.89], p ~ 6.71e-27). We then evaluate how including environmental variables and polygenic risk scores affects prediction of complex disease across diverse genetic similarity groups and test for gene-environment interactions. Future efforts will focus on identifying additional environmental variables and other factors to better characterize their contributions to health.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1597 Infectious disease as a trigger: The relationship between infection and immunological disease.

Authors:

L. Howell[1], M. Davis[2]; [1]Brigham Young Univ., Provo, UT, [2]Brigham Young Univ, Provo, UT

Infectious disease and genetic or immunological diseases have long been considered separate entities with completely distinct etiologies and no overlap in causality. However, some evidence suggests that there is a link between infection and diseases like multiple sclerosis (MS) and Alzheimer's disease (AD), or even that infection could act as a trigger for such conditions as well as causing acute infection. Preliminary analysis in the All of Us database compared the rates of encephalitis in individuals with dementia against the greater population, using multivariate logistic regression, and found significantly higher rates of infection among individuals with dementia even after correcting for age ($p < 2 \times 10^{-16}$). The same procedure was done for rates of meningitis, again showing significantly higher rates of infection among individuals with dementia ($p < 2 \times 10^{-16}$). In order to continue to unravel the multifaceted association between infection and other complex diseases, we will perform a PheWAS on two pathogens, focusing on herpesvirus infections (HSV1 and HHV6) and *Mycobacterium tuberculosis* infections, with the goal of identifying associated diseases. We will also perform a GWAS on infection with these two pathogens to screen for genetic variants that could contribute to infection. Our research will be performed in the All of Us database, using the substantial genomic data and EHR data available there, including approximately 24,000 cases of herpesvirus infection and approximately 2,000 cases of tuberculosis. We anticipate that the results will suggest that the relationship between infectious disease and chronic immune conditions is complex and deserving of further study.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1598 Influence of autozygosity on complex disease risk across the phenotypic spectrum

Authors:

H. Martin[1], D. Malawsky[1], E. van Walree[2], B. Jacobs[3], T. Heng[4], Q. Huang[1], A. H. Sabir[5], S. Rahman[6], S. Malik Sharif[7], A. Khan[8], M. Umićević Mirkov[9], 23andMe Research Team, Genes & Health Research Team, D. Posthuma[10], W. Newman[11], C. Griffiths[12], R. Mathur[12], D. van heel[3], S. Finer[3], J. O'Connell[13]; [1]Wellcome Sanger Inst., Hinxton, United Kingdom, [2]Dept. of Clinical Genetics, Amsterdam UMC, Amsterdam, Netherlands, [3]Blizard Inst., Barts and the London Sch. of Med. and Dentistry, Queen Mary Univ. of London, London, United Kingdom, [4]Wellcome Sanger Inst., Hinxton, Cambridgeshire, United Kingdom, [5]West Midlands Regional Clinical Genetics Unit, Birmingham Women's and Children's NHS FT, Birmingham, United Kingdom, [6]Queen Square Inst. of Neurology, Univ. Coll. London, London, United Kingdom, [7]Yorkshire Regional Genetics Service, Leeds Teaching Hosp. NHS Trust, Leeds, United Kingdom, [8]Waltham Forest Council, London, United Kingdom, [9]Congenica Limited, Hinxton, United Kingdom, [10]Dept. of Complex Trait Genetics, Ctr. for Neurogenomics and Cognitive Res., Amsterdam NeuroSci., VU Amsterdam, Amsterdam, Netherlands, [11]Univ. of Manchester, Manchester, United Kingdom, [12]Wolfson Inst. of Population Hlth., Queen Mary Univ. of London, London, United Kingdom, [13]23andMe Inc, Sunnyvale, CA

Consanguinity increases autozygosity, which is associated with rare Mendelian disorders and various quantitative traits. Previous studies investigating its impact on complex disease have been plagued by confounding by socioeconomic and cultural factors. We investigated patterns of consanguinity and associations between autozygosity and complex disease in unrelated individuals from the Genes & Health cohort (G&H; N=23,978 with genetically-inferred Pakistani or Bangladeshi ancestries) and UK Biobank (UKB; N=397,184 with genetically-inferred European or South Asian ancestries), using a new strategy to reduce confounding. Based on their distribution of runs of homozygosity (ROHs), we inferred that 33% of G&H individuals are offspring of second cousins or closer, versus 2% of European-ancestry UKB individuals. We found significant changes in rates of consanguinity with age that differ between ancestry groups.

We examined associations between F_{ROH} (fraction of the genome in ROHs) and traits using only individuals inferred to be offspring of first cousins or closer, reasoning that, within this group, variation in F_{ROH} is primarily driven by stochastic recombination events and Mendelian segregation. We showed that subsetting to this highly consanguineous group greatly attenuates confounding that hampered previous studies. We tested associations between 61 common diseases and F_{ROH} amongst highly consanguineous individuals from G&H and UKB, and found significant associations (FDR<5%) with twelve of them, including type 2 diabetes (T2D) and stress-/anxiety-related disorders. We attempted to replicate seven of these in a within-sibling analysis in 23andMe (median N=478,590). All seven showed concordant directions of effect and we replicated two, T2D and post-traumatic stress disorder (p<0.05/7), despite having only 43% power to do so. We estimated that autozygosity due to consanguinity accounts for 5-18% of T2D cases in British Pakistanis, in which T2D is particularly common. We show through simulations that although in theory, autozygosity can influence risk of binary traits with a purely additive genetic architecture by increasing the variance in polygenic risk, in practice non-additive genetic effects are more likely to be driving our observed signals.

Our work quantifies, for the first time, the impact of autozygosity on dozens of common diseases using a novel approach to control for confounding (i.e. restricting to offspring of first cousins). It has important implications for global populations with high rates of consanguinity, and highlights the possibility of widespread non-additive effects on several diseases.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1599 Influence of germline variants of *IGHV3-53* and *IGHV3-66* genes on antibody responses to BNT162b2 mRNA COVID-19 vaccine

Authors:

Y. Mashimo[1], K. Yamazaki[1], T. Kageyama[2], S. Tanaka[2], T. Taniguchi[3,4], K. Matsushita[5], H. Igari[3,4], H. Hanaoka[6], K. Yokote[7], H. Nakajima[2,4], Y. Onouchi[1]; [1]Dept. of Publ. Hlth., Chiba Univ. Graduate Sch. of Med., Chiba, Japan, [2]Dept. of Allergy and Clinical Immunology, Chiba Univ. Graduate Sch. of Med., Chiba, Japan, [3]Dept. of Infectious Diseases, Chiba Univ. Hosp., Chiba, Japan, [4]Chiba Univ. Hosp. COVID-19 Vaccine Ctr., Chiba, Japan, [5]Div. of Lab. Med., Chiba Univ. Hosp., Chiba, Japan, [6]Clinical Res. Ctr., Chiba Univ. Hosp., Chiba, Japan, [7]Dept. of Endocrinology, Hematology and Gerontology, Chiba Univ. Graduate Sch. of Med., Chiba, Japan

Understanding the factors affecting acquisition of neutralizing antibodies (nAbs) after COVID-19 vaccination may be useful to lead better prevention of COVID-19. Multiple reports have suggested the unique and central role of two paralogous immunoglobulin heavy variable (*IGHV*) genes (*IGHV3-53* and *IGHV3-66*) in protection against SARS-CoV-2 as the components of nAbs. We hypothesized that germline variants affecting the function or expression of these genes may influence antibody acquisition in recipients of COVID-19 vaccines designed to induce nAbs. To validate this hypothesis, we investigated the relationship between post-vaccination antibody titers of 1,773 Japanese healthcare workers who received two doses of BNT162b2 mRNA COVID-19 vaccine and the germline variants of *IGHV3-53* and *IGHV3-66*. Based on the existing information on expression quantitative trait loci (eQTL) and linkage disequilibrium of the surrounding variants and data in our previous study, two single nucleotide variants rs11623191 and rs6423677 within *IGHV3-53* and *IGHV3-66*, respectively were selected as the promising candidates for the predictors of the post-vaccination antibody titer. Significant impacts of rs11623191-T and rs6423677-C alleles on the increasing usage of *IGHV3-53* and *IGHV3-66*, respectively in *IGH* transcripts were confirmed by analyses of genotypes and *IGHV* repertoire in peripheral blood mononuclear cells of 96 participants. Linear regression showed significant effect of the number of rs6423677-C allele on log2-transformed antibody titers (coefficient [B] = 0.08, 95% confidence interval [CI], 0.0085-0.15). Although this relationship was not observed for rs11623191 (B = 0.0021, 95% CI, -0.071-0.075), the sum of the numbers of rs6423677-C and rs11623191-T alleles appeared to predict a larger titer increase than rs6423677 alone (B = 0.095, 95% CI, 0.018-0.17). In addition, an independent effect of these two alleles on the antibody titer was also found in a multivariate linear regression analysis additionally using previously reported predictors as explanatory variables, including age, sex, and several factors such as medication and drinking behavior (n = 1,511, standardized coefficient [β] = 0.056, 95% CI, 0.0088-0.10). Interestingly, the effect of the two SNVs was not uniform among the participant subpopulations stratified by sex and/or age. Our study is the first to reveal the influence of the germline variants of two *IGHV* genes on the antibody response against SARS-CoV-2 after BNT126b2 vaccination. This result might provide an important clue to future optimization of COVID-19 prevention strategies.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1600 Insights into genetics affecting brain morphology and the endocrine system using image-derived phenotypes from large-scale magnetic resonance imaging.

Authors:

H. Currant[1], C. Arthofer[2], G. Douaud[2], T. Ferreira[3], S. S. Venkatesh[3,4], C. Nellaker[3], S. Smith[2], B. Hill[3], N. A. Baya[3,4], D. S. Palmer[3], A. Bartsch[5], J. Andersson[2], M. F. Lippincott[6,7], Y-M. Chan[8], S. B. Seminara[7,6], T. E. Nichols[2,3], S. Brunak[1], F. J. Lange[2], C. M. Lindgren[3,4]; [1]Novo Nordisk Fndn. Ctr. for Protein Res., Univ. of Copenhagen, Copenhagen, Denmark, [2]Oxford Ctr. for Functional MRI of the Brain, Oxford Univ., Oxford, United Kingdom, [3]Big Data Inst., Oxford Univ., Oxford, United Kingdom, [4]Wellcome Ctr. for Human Genetics, Oxford Univ., Oxford, United Kingdom, [5]Univ. of Heidelberg, Heidelberg, Germany, [6]Reproductive Endocrine Unit, Massachusetts Gen. Hosp., Boston, MA, [7]Harvard Ctr. for Reproductive Med., Massachusetts Gen. Hosp., Boston, MA, [8]Boston Children's Hosp., Boston, MA

Background: Several brain structures are involved in regulating the endocrine system, and their dysmorphology predicts function: pituitary gland volume correlates with sex-steroid concentrations and hypoplasia of olfactory bulbs is a phenotype of Kallmann syndrome which causes infertility. Magnetic resonance imaging (MRI) provides a tool for acquiring rich phenotypes that reflect brain morphology. Here we conduct genetic association analyses at different phenotypic scales to elucidate the biology underlying brain morphology and explore its link to endocrinology.

Methods: We performed genome-wide association studies utilising image-derived phenotypes from UK Biobank brain MRI (n=34,834, 53% female) representing the volume of endocrine-related brain structures: the hypothalamus, pituitary gland and olfactory bulbs. We identified sexually dimorphic loci and performed phenome-wide association studies (PheWAS). To elucidate links between brain morphology and reproductive hormone levels, we calculated genetic correlation, performed Mendelian Randomisation (MR) and constructed genotype-stratified hormone trajectories. We further expanded phenotypic resolution by supplementing the original phenotypes with detailed spatial maps of the genetic association between loci of interest and voxel-level volume changes across the entire brain.

Results: We discovered 48 loci ($p<5\times10^{-8}$) associated with the volume of the hypothalamus, pituitary and olfactory bulbs, two of which were sexually dimorphic ($pdiff<5.5\times10^{-4}$). PheWAS identified associations with endocrine conditions including diabetes mellitus ($p<1.69\times10^{-5}$). There was evidence of genetic correlation of pituitary gland volume with follicle stimulating hormone (FSH) and testosterone levels ($p<0.05$). MR further identified a possible bidirectional link between brain structure volumes and FSH and testosterone levels. Analysis of primary care data identified age-dependent differences in sex hormone levels when stratified by genotype at identified loci. Voxel-wise modelling found additional associations between loci associated with the three brain structures and finer-grain brain morphology that reinforce the relationship between these brain structures in regulating the endocrine system. These include a genetic variant associated with pituitary gland volume that also showed associations with the volume of parts of the olfactory cortex.

Conclusions: Empowered by its unparalleled size and high-resolution of phenotype, our project furthers our understanding of the genetics underlying brain structure, and the interaction of these effects with the endocrine system.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1601 Integrate RNAseq gene expression and genome-wide genetic data to identify genes modifying age-at-onset of Alzheimer's disease

Authors:

Y-J. Li[1,2], M-F. Hsueh[1,2], J. La[1,2], E. Martin[3,4]; [1]Duke Univ., Durham, NC, [2]Duke Molecular Physiology Inst., Durham, NC, [3]Univ. of Miami Miller Sch. of Med., Miami, FL, [4]John P. Hussman Inst. for Human Genomics, Univ. of Miami, Miami, FL

Objective: Genetic effects on age-at-onset (AAO) of Alzheimer's disease (AD) are supported by high heritability (e.g., 42%). In recent years, several large genome-wide association studies (GWAS) have identified genes for AAO of AD, including known AD-risk genes and novel ones. However, GWAS may not capture all AAO genes. Here, we sought to use RNAseq data to identify additional genes for AAO of AD and genetic variants with pleiotropic effects on both gene expression and AAO of AD.
Methods: RNAseq data from dorsolateral prefrontal cortex tissues were available for 636 participants from ROSMAP cohort, of which 249 Non-Hispanic White (NHW) AD cases with AAO data were analyzed. Fragments per kilobase exon per million (FPKM) were used as expression measures. Excluding genes with > 10% missing expression data, 19,173 genes were tested. Multivariable linear regression models regressed log10(FPKM) on AAO, disease duration, sex, postmortem interval, RNA integrity number, and the first principal component derived from expression data. This model allows us to identify genes associated with AAO and disease duration, respectively. Qvalues were computed to correct for multiple testing, and genes with qvalue < 0.05 were considered genome-wide significant. Suggestive genes were defined by $p < 5x10^{-4}$. eQTLs for the top AAO genes were queried from GTEx and crossed checked for their association with AAO of AD from our AAO GWAS results. To investigate genetic variants with pleiotropic effects on both gene expression and AAO of AD, we performed fastQTL analysis for each gene expression using variants within 5KB of the gene and will follow with colocalization and SMR analyses using results of fastQTL and AAO GWAS.
Results: Among 249 NHW AD cases, the median (Q1, Q3) was 87.29 (83.4, 91.3) years for AAO and 2.8 (1.2, 5.0) years for AD duration with moderate correlation between them (r=-0.37). *MSLNL* reached genome-wide significance (beta (95% CI) =0.02 (0.01, 0.03), $p=1.1x10^{-6}$, qvalue=0.019) for AAO and 14 suggestive genes were identified ($p< 5x10^{-4}$) with *ITM2A* gene ranked at the second (p=9.83x10-5). On the other hand, *WDR55* was significantly associated with disease duration ($p=2.84x10^{-6}$, qvalue=0.026). The best eQTLs for *MSLNL* were in *MIR662* (in the overlapping region). These eQTLs also showed association with AAO of AD (e.g., rs9927150, p=0.003).
Conclusion: Gene-based expression association analysis for AAO demonstrated the utility to discover novel AAO gene such as *MSLNL*. We also identified *WDR55* associated with AD disease duration. More analysis is underway to understanding these findings further.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1602 Integrated analysis of rare exonic variants provides additional insights into alcohol use disorder risk.

Authors:

T-H. Nguyen[1], A. Gentry[1], H. Vy[2], M. Ahangari[1], R. Peterson[3], K. Kendler[4], B. Riley[4], S-A. Bacanu[4], B. Webb[5]; [1]Virginia Commonwealth Univ., Richmond, VA, [2]Icahn Sch. of Med. at Mount Sinai, New York, NY, [3]SUNY Downstate Hlth.Sci. Univ., Brooklyn, NY, [4]Virginia Commonwealth Univ, Richmond, VA, [5]RTI Intl., Research Triangle Park, NC

Alcohol use disorder (AUD) is common and affects millions of people in the United States. Twin and family studies show that this disorder is heritable, and genome wide association studies (GWASs) report multiple loci associated with AUD. Post-GWAS analyses of GWAS signals yield enrichment in several gene-sets that show nominal enrichment in brain tissues. Unfortunately, only a handful of genes have been reported from rare-variant studies. However, recent studies strongly imply that for many complex disorders, common and rare variant findings, while not converging at gene level, converge at gene-set level. In addition, recent studies have showed that using phenotypic information predicted from biobanks can increase genetic discoveries. These suggest that using different approaches for rare variant analyses can improve detection power for genes associated with AUD.
Here, we explore integrative approaches to prioritize genes associated with a proxy phenotype of AUD, the Alcohol Use Disorder Identification Test-Problems (AUDIT-P) phenotype. We analyzed the whole-exome-sequencing (WES) datasets of 500K people from the UK Biobank. First, we conducted WES analyses for individuals whose AUDIT scores are available. Second, we developed a pipeline to jointly model rare variants and gene-sets to improve statistical power. Finally, we used a machine learning approach developed by our group to predict AUDIT scores for all the 500K people, and re-analyzed their WES datasets.
We first analyzed loss-of-function and missense variants from WES sequencing datasets of 133,914 people with AUDIT scores available. Three significant genes (*ADH1C, FPR1, VPS29*) were observed. The most significant signal was for *ADH1C* (adjusted p-value = 0.4×10^{-5}). Next, to prioritize additional genes, we jointly analyzed gene-level statistics and 181 gene sets curated from previous studies. We prioritized several genes (max posterior probability > 0.8) including *ADH1C*. Finally, to see if the prediction of phenotypic information can help prioritize genes, we conducted WES analysis for all 414,508 samples of European descent with full predicted AUDIT scores. Statistical power for *ADH1C* was substantially improved (adjusted p-value = 3.3×10^{-21}). Our results present top significant genes for AUD obtained by analyzing rare variants from a large-scale WES dataset. The results also include 1) the additional biological information into AUD, and 2) integrative approaches for incorporating functional genomics and health care record datasets to improve genetic discoveries. This research has been conducted using the UK Biobank Resource application number 30782.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1603 Integrated Illumina and PacBio genome sequencing of Middle Eastern families identifies novel pathogenic variants underlying neuro-developmental disorders

Authors:

Z. Siddig[1], M. Ghorbani[1], R. Abouelhassan[1], S. Hassan[1], H. Naeem[1], R. Farhad[1], I. Diboun[1], S. Moosa[1], M. Tauro[1], A. El Khouly[1], S. Poolat[1], S. Tomei[1], S. Lorenz[1], D. Love[1], K. Fakhro[1,2,3], A. Satti[1], R. Benini[1], Y. Mokrab[1,2,3]; [1]Sidra Med., Doha, Qatar, [2]Weill Cornell Med., New York, U.S.A, NY, [3]Hamad Bin Khalifa Univ., Doha, Qatar

Establishing genotype-phenotype correlations for neurodevelopmental disorders (NDDs) is a substantial challenge, especially amongst Middle Eastern populations enriched for these conditions due to high level of consanguinity. We implemented a platform integrating data from Illumina WGS, PacBio long-read sequencing, and EHRs to diagnose and characterize unresolved pediatric cases with NDDs in Qatar. In a pilot project, we recruited a cohort with intractable epilepsy or GDD, consisting of 168 families of patients. Family segregation genomic analysis was done using Congenica©, and variants were annotated with allele frequency from Qatar Genome Project and other public datasets. Candidate variants were prioritized using ACMG guidelines. Using this approach, we analyzed a set of 15 families leading to a positive genetic diagnosis yield of 40%. We identified 4 novel variants within previously reported disease genes using Illumina HiSeq, in addition to a likely pathogenic homozygous deletion identified using PacBio HiFi in two patients. Through deploying family segregation analysis on Congenica©, we identified a likely pathogenic missense mutation (NM_004113.5:c.155G>A, chr3:192335434 based on GRCh38, p. R52H) in *FGF12* in a 2-year-old boy who exhibits early-onset epileptic encephalopathy and global developmental delay. This gene encodes a cytosolic neuronal sodium channel binding protein and increases their voltage-dependent fast inactivation. The reported protein-altering variant has a change of amino acid from arginine to histidine, and the affected arginine residue is part of the highly conserved surface region that binds voltage-gated sodium channels which is necessary for *FGF12* modulation of sodium channels' fast inactivation gating. The mutation was predicted to be damaging by in silico prediction scores, including CADD (>25) and SIFT (<0.05). This gene is reported to have a definitive association with neonatal onset epileptic encephalopathy and intellectual disability in the Decipher Gene2Phenotype database, matching the reported patient's phenotype. Functional studies reported in the literature have shown that mutant *FGF12* proteins have a strong gain-of-function phenotype, increasing neuronal excitability in zebrafish. Currently results are being analyzed for the rest of the cohort, the results of which will have significance impact on precision diagnosis and understanding of NDDs in the Middle East and worldwide.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1604 † Integrating COVID-19 severity genetics with single-cell omics implicates alveolar type II cells and CD14 monocytes

Authors:

D. Chang, H. Shi, D. Garfield, C. M. Rosenberger, M. Reyes, S. Kapadia, M. I. McCarthy, B. L. Yaspan; Genentech, South San Francisco, CA

Large-scale efforts to characterize host genetic factors underlying COVID-19 severity have associated over 40 loci (The COVID-19 Host Genetics Initiative, 2022), highlighting genes involved in innate immunity and type 1 interferon response. Single-cell studies have also associated monocyte transcriptional states with severity in COVID-19 patients (Schulte-Schrepping et al, 2020), providing further support for the role of innate immunity in the pathogenesis of COVID-19. Beyond peripheral blood, alveolar type II (ATII) cells in the lung have been implicated from observations of ATII hyperplasia and loss in COVID-19 patients, and the genetic association with hospitalized risk at the SFTPD locus. To investigate whether additional cell types play a role in COVID-19 severity and identify pathogenic cell functions, we integrated a genome-wide association study (GWAS) of hospitalized COVID-19 risk (release 7 COVID-19 HGI) with studies measuring the transcriptional and chromatin accessibility landscape across single cells in the blood and the lung. First, we used the LD-score regression framework to investigate whether hospitalized COVID-19 genetic risk was enriched in regions of open chromatin in >200 cell types captured across 30 tissues (Zhang et al 2021, Wilk et al 2022). We found nominal evidence for enriched heritability in open chromatin of ATII cells (P=0.016) and CD14 monocytes (P=0.030). We next applied the scDRS method to integrate polygenic signals from hospitalized COVID19 risk GWAS with single cell transcriptomic data from the Human Lung Cell Atlas (Sikkema et al, 2022) and a study of PBMCs from COVID19 participants and healthy controls (Wilk et al, 2020). Similar to the partitioned heritability results above, we found that in the lung, ATII cells were enriched for higher scDRS scores ($P=1.00x10^{-3}$, FDR=0.058), while in the blood, activated granulocytes, CD14 monocytes and dendritic cells from COVID-19 patients were enriched (P<0.010, FDR<0.1). In ATII cells, gene expression correlating with the scDRS score mapped to several pathways including fatty acid metabolism and response to IL6 (FDR<0.029). In contrast, gene expression correlating with the scDRS score in COVID-19 patient derived CD14 monocytes mapped to inflammation associated pathways such as type 1 interferon response, and granulocyte activation ($FDR<2.20x10^{-16}$). Together these results provide support for the causal roles of ATII and CD14 monocytes in the pathogenesis of severe COVID-19, and highlight key pathogenic pathways in these cells.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1605 Integrating metabolomics with proteomics identifies novel drug targets for heart failure and atrial fibrillation.

Authors:

M. van Vugt[1,2,3], C. Finan[3,1,4], S. Chopade[3,4], F. W. Asselbergs[2,3], J. van Setten[1], F. Schmidt[2,3]; [1]UMC Utrecht, Utrecht, Netherlands, [2]Amsterdam UMC, Amsterdam, Netherlands, [3]Univ. Coll. London, London, United Kingdom, [4]UCL BHF Res. Accelerator Ctr., London, United Kingdom

Background. Altered metabolism plays a role in the pathophysiology of heart diseases, such as atrial fibrillation (AF) and heart failure (HF). Aside from NT-pro-BNP, which is used for HF diagnosis and prognosis, there is a paucity of plasma metabolites for patient monitoring and potential therapeutic targets. We aimed to identify plasma metabolites and biological pathways associated with the onset of cardiac disease, and subsequently prioritized plasma proteins that affect these metabolites and are potentially relevant for therapeutic intervention. **Methods.** Mendelian randomization (MR) was used to assess the causal association of 174 metabolites measured in up to 86,507 participants, spanning seven distinct metabolic classes, on the following cardiac outcomes: AF, HF, dilated cardiomyopathy (DCM), and non-ischemic cardiomyopathy. Subsequently, we sourced data on 1,567 plasma proteins and performed *cis* MR to identify proteins affecting the identified metabolites as well as the cardiac diseases. Results were subsequently mapped to biological pathways and prioritized for cardiac tissue expression and druggability. **Results.** Of the 174 metabolites spanning seven metabolic classes, 34 metabolites belonging to six metabolite classes were prioritized for involvement with cardiac disease. While most metabolite classes affected multiple cardiac outcomes, phosphatidylcholines were enriched for HF (p=0.029) and acylcarnitines for DCM (p=0.001). A total of 87 proteins affected these metabolites and cardiac diseases, including APOC3 and PLA2R for HF, and RET and DPEP1 for AF. Furthermore, 38 proteins were drugged or druggable, of which eight (ACES, AT1B2, DPEP1, FA10, IL6RA, KPCA, PCSK9, and RET) were targeted by drugs with cardiovascular indications, side-effects, or both. **Conclusion.** This study identified plasma metabolites involved with cardiac diseases, providing information on potential diagnostic markers. By linking these metabolites to druggable proteins with a concordant effect on cardiac disease, we provide actionable leads to facilitate the development of novel treatments for AF and HF.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1606 Integrating phenome and gene level approaches to identify genetic associations with glaucoma

Authors:

A. Scalici[1,2,3], T. Miller-Fleming[4,2,5], J. Hirbo[6,2,5], N. Cox[4,2,5]; [1]Vanderbilt Univ., Nashville, TN, [2]Vanderbilt Genetics Inst., Vanderbilt Univ. Med. Ctr., Nashville, TN, [3]Div. of Genetic Med., Vanderbilt Univ. Med. Ctr., Nashville, TN, [4]Vanderbilt Univ Med Ctr., Nashville, TN, [5]Div. of Genetic Med., Dept. of Med., Vanderbilt Univ. Med. Ctr., Nashville, TN, [6]Vanderbilt Univ. Sch. of Med., Nashville, TN

Glaucoma is a highly heritable progressive eye disorder defined by optic nerve degeneration. While age and biological sex are major risk factors for developing glaucoma, recent studies have shown that there are significant disparities in glaucoma prevalence, severity of disease, and age of onset between individuals of African and European ancestry. Despite this disparity, the majority of genome-wide association studies (GWAS) of glaucoma are conducted among individuals of European ancestry. Leveraging electronic health record (EHR) based studies that integrate genetically predicted gene expression (GPGE) and glaucoma comorbidities have the potential to shed greater insight into disease mechanisms and expand the known genes and pathways associated with glaucoma pathogenesis. We identified glaucoma cases (N=11,888) and age matched controls (N=59,440) within Vanderbilt's EHR that had at least three visits in five years and had an EHR-reported race as white. We conducted a phenome-wide association study (PheWAS) to identify glaucoma comorbidities. Known glaucoma comorbidities such as radiotherapy (p=5.30x10-241), and diabetes (p=5.05x10-72) were significantly enriched among glaucoma cases. Using the beta estimates from the significantly associated comorbidities as weights, we constructed a phenotypic risk score (PheRS) in an independent population (BioVU, N=70,493 participants with European genetic ancestry). This PheRS represents a weighted sum of an individual's glaucoma comorbidities. To test whether the PheRS was associated with changes in GPGE, we conducted a transcriptome-wide association study (TWAS) in BioVU. We identified several genes (NEU2, TDRKH, TPP1) (p<8x10-5) that have previously been associated with neurological phenotypes, but not glaucoma. In future studies, we will apply this phenome-based approach to individuals of African ancestry to identify ancestry-specific genetic factors that may be involved in glaucoma pathogenesis.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1607 Integration of 3D human stem cell models derived from post-mortem tissue and statistical genomics to guide schizophrenia therapeutic development

Authors:

J. Erwin[1,2], T. Sawada[1], K. Benjamin[1], A. Barbosa[1], B. Araujo[1], A. McCord[1], L. D'Ignazio[1], B. Sheehan[3], M. Zabolocki[4], J. Kleinman[1], T. Hyde[1], C. Bardy[5], D. Weinberger[3], A. Paquola[6]; [1]Lieber Inst., Baltimore, MD, [2]Johns Hopkins Sch. of Med., Baltimore, MD, [3]Lieber Inst. for Brain Dev., Baltimore, MD, [4]South Australian Hlth.and Med. Res. Inst. (SAHMRI), Adelaide, Australia, [5]South Australian Hlth.and Med. Res. Inst., Adelaide, Australia, [6]Lieber Inst. for Brain Dev. / Johns Hopkins Med., Baltimore, MD

Schizophrenia is a neuropsychiatric disorder characterized by positive symptoms (such as hallucinations and delusions), negative symptoms. Schizophrenia is highly heritable, and genetic studies are playing a pivotal role in identifying potential biomarkers and causal disease mechanisms with the hope of informing new treatments. Genome-wide association studies (GWAS) identified nearly 270 loci with a high statistical association with schizophrenia risk; however each locus confers only a small increase in risk therefore it is difficult to translate these findings into understanding disease biology that can lead to treatments. Induced pluripotent stem cell (iPSC) models are a tractable system to translate genetic findings and interrogate mechanisms of pathogenesis. Mounting research with patient-derived iPSCs has proposed several neurodevelopmental pathways altered in SCZ. However, it is unclear what exactly these iPS models recapitulate, how potential perturbations of early brain development translates into illness in adults and how iPS models that represent fetal stages can be utilized to further drug development efforts to treat adult illness. I will present the largest transcriptome analysis of post-mortem caudate nucleus in schizophrenia where we discovered that decreased presynaptic DRD2 autoregulation is the causal dopamine risk factor for schizophrenia. We developed stem cell models from a subset of the postmortem cohort to better understand the molecular underpinnings of human psychiatric disorders. We established a method for the differentiation of iPS cells into ventral forebrain organoids and performed single cell RNAseq and cellular phenotyping. To our knowledge, this is the first study to evaluate iPSC models of SZ from the same individuals with postmortem tissue. Our study establishes that striatal neurons in the patients with SCZ carry abnormalities that originated during early brain development. Differentiation of inhibitory neurons is accelerated whereas excitatory neuronal development is delayed, implicating an excitation and inhibition (E-I) imbalance during early brain development in SCZ. We found a significant overlap of genes upregulated in the inhibitory neurons in SCZ organoids with upregulated genes in postmortem caudate tissues from the same patients with SCZ compared with control individuals. Altogether, we demonstrate that ventral forebrain organoids derived from postmortem tissue of individuals with schizophrenia recapitulate perturbed striatal gene expression dynamics of the donors' brains.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1608 † Integration of protein-disease associations with Mendelian randomization reveals potential drug targets and disease biomarkers in UK Biobank Pharma Proteomics Project (UKB-PPP)

Authors:

L. Hou[1], A. Li[1], Y. Zhang[1], E. Baugh[1], A. Doostparast Torshizi[1], S. Guo[2], A. Hart[1], J. Molineros[1], D. Waterworth[1], T. Mansi[3], C. Whelan[4], S. Li[1]; [1]Janssen R&D, Spring House, PA, [2]Janssen R&D, San Francisco, CA, [3]Janssen R&D, Titusville, NJ, [4]Janssen R&D, Boston, MA

The UK Biobank Pharma Proteomics Project (UKB-PPP) collected plasma proteomics data of 2,923 proteins across 54,219 participants. The proteomics data along with extensive phenotypic and genetic data offers unprecedented opportunity for human omics-driven drug discovery. In this study, we developed a framework for biomarker and target identification by integrating: 1) cross-sectional association analysis between plasma protein levels with prevalent and incident cases of ~1,500 diseases defined by combined ICD codes (PheCODEs); 2) protein-change trajectory estimation for patients before and after disease onset; 3) systematic Mendelian randomization (MR) analysis for ~800 traits using UKB-PPP cis-pQTL as instrumental variables. Our analysis revealed 241,501 significant ($p < 6.0\times10^{-8}$) associations between 2,389 proteins with 860 diseases. Protein-disease pairs were further classified into different categories based on 1) whether the protein is casually related with the diseases from MR analysis, 2) whether the protein is significantly up/downregulated in prevalent, incident cases or both, 3) trajectory of protein-change before/after disease onset. Our results confirmed some well-known causal relationships (e.g., ANGPTL3 for hypercholesterolemia and coronary artery disease) and revealed novel causal risk factors (e.g., IL4R for type 1 diabetes, UMOD for hypertension). Among non-causal protein-disease pairs, protein association and change trajectory analysis replicated diagnostic biomarkers established previously (e.g., prostate-specific antigen for prostate cancer) and revealed novel insights into candidate biomarkers: glial fibrillary acidic protein (GFAP), for example, is a more neurodegeneration-specific biomarker elevated in Alzheimer's disease (AD) and dementia cases at least 5 years prior to diagnosis, compared to neurofilament light-chain (NfL), which is a more non-specific marker associated with >200 diseases, including AD and other neurological, metabolic, and respiratory diseases. Our work demonstrates that integrating large-scale human omics analyses has great potential to identify therapeutic targets and potential biomarkers for disease prediction, diagnosis, or subtyping, especially with the public release of UKB-PPP data and the growth of diverse biobanks.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1609 Integrative Analysis of Cell-Specific Human Endogenous Retrovirus Expression and Host Gene Expression Identifies Associations with Clinical Phenotypes in Systemic Lupus Erythematosus.

Authors:

M. Sirota[1], C. Lanata[2], Z. Cutts[1], S. Patterson[1], L. Maliskova[1], K. Taylor[1], C. Ye[1], M. Dall'Era[1], J. Yazdany[1], L. Criswell[3], G. Fragiadakis[1], C. Langelier[1], T. Capra[1]; [1]UCSF, San Francisco, CA, [2]NHGRI, Bethesda, MD, [3]Natl. Human Genome Res. Inst., Bethesda, MD

Introduction: Transposable elements (TEs) consist of human endogenous retroviruses (HERVs) and long interspersed nuclear elements (LINEs). Their expression may contribute to production of type I IFN and the generation of autoantibodies. The objective of this study was to detect TEs in 4 cell-types in systemic lupus erythematosus (SLE) patients and characterize their relationship to SLE phenotypes and host gene expression.
Methods: We profiled cell-sorted RNA-seq data (CD4+ T cells, CD14+ monocytes, CD19+ B cells, and NK cells) from PBMCs of 119 SLE patients from the California Lupus Epidemiology Study. We performed cell specific differential expression analysis of TEs across 11 SLE clinical phenotypes adjusting for sex, age, genetic principal components 1-10 and medication use. TE family and viral gene enrichment was performed. We explored which genes and pathways were differentially expressed with SLE phenotypes associated TEs.
Results: After standard quality control, we quantified TEs in 4 cell-types using Telescope resulting in a total of 27,135 TEs. TE expression was cell specific as observed by dimensionality reduction clustering. Differential expression analysis on SLE subphenotypes identified a moderate number of significant differentially expressed TEs in different cell types and SLE subphenotypes (adj $p < 0.05$). For example, 53 TEs were differentially expressed in CD19 cells by disease severity. Some commonly positively associated TEs across phenotypes include MER4B, MER101, ERVLB4, ERVLE and HERVH. Viral protein enrichment analysis found protease (Pro), integrase, reverse transcriptase (RT) enriched in the CD4 cells, dutpase enriched in CD14, Pro and RT enriched in CD19, and Ribonuclease H, Pro, and RT enriched in NK cells. We then performed differential expression analysis of host gene expression with SLE phenotype associated TEs across cell types. There was an upregulation of the IFN signaling pathway in CD4, CD19 and CD14 cells (FDR <0.05). CD19 cells had upregulation of genes in pathways related to BCR signaling, NF-kB activation, and DNA repair. CD4 cells revealed a downregulation of metabolic pathways and viral influenza pathways. CD14 showed upregulation of OAS antiviral response, and NK cells had downregulation of transcription pathways.
Conclusion: We have described differentially expressed TEs in 4 cell types according to several SLE subphenotypes. Furthermore, these TEs correlate with differential gene expression in SLE relevant pathways, in particular in CD19 cells. These findings suggest that expression of TEs might be contributing to activation of SLE-related mechanisms in a cell-specific manner.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1610 Integrative analysis of the colonic transcriptome and colonoscopy-based diverticula enumeration unveils diverticulosis-specific causal genes

Authors:

J. Seo, H. Liu, K. Young, X. Zhang, T. O. Keku, C. D. Jones, K. E. North, S. S. Robert, A. F. Peery; Univ. of North Carolina, Chapel Hill, NC

Colonic diverticulosis is a prevalent condition among older adults, marked by the presence of thin-walled pockets in the colon wall that can become inflamed, infected, hemorrhage, and rupture. The genetic factors that contribute to the pathogenesis of the disease, including the transcriptome, remain understudied. Here, we report the first large colonoscopy-based transcriptomic study that investigates variations in the colonic transcriptome from well-characterized patients with (N=172) and without (N=232) diverticulosis. Between patient groups, 38 genes were differentially expressed, and 17 genes showed evidence of differential transcript usage. These genes suggested tissue remodeling as a potential mechanism of diverticula formation. In addition, expression quantitative trait loci (eQTL) mapping identified a total of 6,143 genes that exhibited at least one *cis*-eQTLs at a false discovery rate (FDR) of less than 1%. Mendelian randomization combining GWAS and colonic *cis*-eQTLs identified potential causal genes at five loci, three of which had been previously implicated in diverticulosis GWAS (*ENTPD7*, *CRISPLD2*, and *PHGR1*). Notably, the *ENTPD7* gene was upregulated in diverticulosis patients. Furthermore, polygenic risk score analysis revealed that a higher prevalence of diverticula is associated with elevated genetic risk for developing diverticulitis. These discoveries shed light on the underlying gene regulation contributing to diverticulosis and provide insights into the potential molecular pathways involved in its pathogenesis.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1611 Interplay between Polygenic Risk Score and Solar Radiation Exposure: Implication for Systemic Lupus Erythematosus Onset and Pathogenesis

Authors:

Y-M. Chen[1], T-C. Chen[2], H-W. Yang[1], Y-J. Chen[1], I-C. Chen[1]; [1]Taichung Veterans Gen. Hosp., Taiwan, Taichung, Taiwan, [2]Academia Sinica, Taiwan, Taipei, Taiwan

Objectives: This research elucidates the correlation between solar radiation exposure, polygenic risk score (PRS), and systemic lupus erythematosus (SLE) onset, utilizing genomic, environmental, and clinical data.
Methods: We included 1,800 SLE participants and 1,800 controls from the Taiwan Precision Medicine Initiative, genotyped via the Affymetrix Genome-Wide TWB 2.0 SNP Array. The SLE PRS, a sum of weighted genetic variants encompassing 27 SNPs, was calculated. QGIS computed solar radiation exposure from participants' residences. We utilized logistic regression and mediation effect analyses to explore connections between SLE PRS, solar insolation susceptibility, and SLE.
Results: SLE patients showed decreased solar insolation ($p<0.001$). The highest decile of SLE-PRS had lower solar insolation before 1, 3, 6, and 12 months compared to the lowest decile, with statistical significance for 1 and 12 months ($p=0.025$, $p=0.004$, respectively). An inverse association between SLE prevalence and solar insolation was found. Solar insolation didn't mediate the SLE-PRS effect on SLE in the high insolation group. A substantial direct effect of SLE-PRS on SLE (0.44, 95% CI: 0.23-0.68, $p<0.001$) was observed in the high solar insolation group. Compared to the lowest decile, the highest SLE-PRS decile showed a 10.98-fold SLE risk increase (95% CI, 3.773-31.952, $p<0.001$). High SLE-PRS and high solar insolation exposure significantly raised SLE risk.
Conclusions: Our study unveils the intertwined nature of UV exposure and polygenic risks in SLE. Future studies should explore the preventative potential of robust solar radiation protection for high-risk individuals before the disease onset.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1612 Interrogating aging-associated transcriptional changes in the murine enteric nervous system

Authors:

J. Slosberg[1], S. Kulkarni[2,3], L. Goff[1]; [1]Johns Hopkins Univ., Baltimore, MD, [2]Beth Israel Deaconess Med. Ctr., Boston, MA, [3]Harvard Med. Sch., Boston, MA

Diverse gut functions, including intestinal motility, absorption, and immunity, are regulated by the cells of the enteric nervous system (ENS) which resides entirely within the gut wall. These processes change significantly during maturation and are increasingly dysregulated with aging - suggesting that the ENS changes structurally and functionally during these phases. However, the molecular factors that contribute to ENS maturation and aging are unknown. Identifying these will provide a more complete understanding of the etiology of age-associated ENS disorders and will yield putative targets for therapies against diseases, for which incidence is increasing in aging populations.

To this goal, we investigated how the motility-regulating enteric neurons and their local environment, the longitudinal muscle-contained myenteric plexus (LM-MP) of the gut wall, changes at a molecular level over the course of aging. We performed RNA-sequencing to profile the transcriptome of the ileal LM-MP tissue of male and female C57BL/6J mice of 3 different ages: juvenile (1 month-old), mature adult (6 month-old), and aging (17 month-old). Using a DESeq2 analysis framework, 809 genes were identified as varying with age (388 genes upregulated, 411 genes downregulated, and 10 genes with inconsistent direction of effect). Between tissues from juvenile and aging mice, a gene set enrichment analysis highlighted the downregulation of the TGF-beta signaling pathway and of genes involved in extracellular matrix (ECM) organization. Specifically, we observe decreased expression in diverse collagen subunits, elastin (Eln), and matrix metalloproteinases (Mmp2, Mmp11, Mmp15), suggesting aging-associated decreased rates of ECM turnover and remodeling. We also found significant changes in genes involved in synapse organization, notably a decrease in the neuronally-expressed proteoglycan Agrin (Agrn). Aging was further associated with increased expression of Ric3 (resistance to inhibition of cholinesterase 3) and Vsnl1 (visinin-like 1), which are known to modulate neuronal signaling.

Our observations suggest important aging-associated changes in gene regulatory networks that are important for normal ENS and gut function. These data provide us with important putative targets for therapeutic interventions to provide treatment for aging-associated gut dysfunction.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1613 Investigating alternative splice variants as potential biomarkers in Parkinson's disease: Insights from targeted sequencing of plasma extracellular vesicles

Authors:

T. Cole[1], D. L. Miller[2], S. K. Chakrabortty[2], K. Babcock[2], S. Xing[2], J. J. Ray[2], B. C. Haynes[1], S. Yu[2], J. Skog[2]; [1]Bio-techne, AUSTIN, TX, [2]Bio-techne, Waltham, MA

Introduction: Parkinson's disease (PD) is a neurodegenerative disorder with rapidly expanding global prevalence, ranking as the second most common neurodegenerative disorder worldwide. Dysregulation of RNA splicing has emerged as a hallmark of neurodegenerative diseases, including PD, underscoring the importance of investigating alternative splice variants. The use of minimally invasive plasma liquid biopsies provides an opportunity to monitor neuronal-origin cell-free RNA (cfRNA) and holds promise for surveillance purposes in the context of PD. While the low abundance of neuronal extracellular vesicles in hematopoietic-rich plasma poses challenges, leveraging targeted profiling of brain-derived transcripts can unveil rare splice variants and enhance their detection.
Methods: We conducted whole transcriptome sequencing (WTS), whole exome sequencing (WES), and brain-specific exosomal RNA profiling on a cohort of 20 plasma samples from PD patients and 20 healthy controls. We performed differential splice variant analysis to examine the influence of each RNA-Seq enrichment platform on the resolution of gene expression and the detection of isoforms.
Results: The brain-specific exosomal RNA profiling exhibited remarkable performance in terms of coverage and detection of brain-specific transcripts compared to the WTS and WES panels. We observed orders of magnitude deeper coverage of transcripts specific to the brain tissue using brain-specific exosomal RNA profiling and identified a substantial number of transcripts that were not detected by the other panels. Notably, our analysis unveiled a set of 37 genes that demonstrated robust evidence of differential splice variant usage across all three panels. Furthermore, the brain-specific panel revealed an additional 30 genes that were exclusively identified.
Conclusions: Our findings underscore the significance of leveraging brain-specific exosomal RNA profiling for a more comprehensive understanding of the transcriptomic landscape in the context of PD. The identification of genes exhibiting significant alternative splice usage across multiple panels emphasizes their potential as robust biomarker candidates for PD. Moreover, the discovery of additional genes exclusively captured by the brain-specific panel adds to the richness of the transcriptomic landscape and further expands the repertoire of potential PD biomarkers. These findings underscore the value of integrating multiple profiling approaches and provide valuable insights for future investigations in larger clinical cohorts.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1614 Investigating ancestry-specific genetic variation in apolipoprotein L genes associated with electronic health record phenotypes in diverse patient biobanks

Authors:

D. Zhang[1], M. Levin[1], S. Damrauer[2], M. Ritchie[1], D. Rader[1]; [1]Perelman Sch. of Med., Philadelphia, PA, [2]Corporal Michael J. Crescenz VA Med. Ctr., Philadelphia, PA

Health care disparities between people of different ancestries and ethnicities are well-documented in every field of medicine. As a poignant example, African Americans are more than 3 times as likely to have kidney failure compared to White Americans per the National Kidney Foundation. A major contributor to this disparity is the sample bias underlying existing genomic studies. Of the ~6,401 studies compiled in the genome-wide association studies (GWAS) catalog, ~95% of all GWAS participants are of European (EUR) ancestry with less than 1% of participants being of African American (AFR) ancestry. Using the Penn Medicine Biobank (PMBB) and adopting a genome-first approach, we investigated 572 protein-altering variants in the apolipoprotein L gene family, including 69 predicted loss-of-function (pLOF) and 503 missense variants, with a specific interest in those more common in non-European populations. We performed phenome-wide association studies (PheWAS) on 48 variants with a MAF > 0.1% in the PMBB AFR population (n = 11,198) against 1,236 binary phenotypes derived from electronic health records data with at least 20 cases. In addition to confirming the known AFR-specific rs73885319 and rs60910145 *APOL1* variants as strong positive controls for their association with end-stage renal disease (ESRD), our results identified a stop-gain variant rs11089781 (p.Gln58*) in the *APOL3* gene also found to be significantly associated with increased risk for ESRD (OR = 1.38, p = 3.64e-08). This variant has a gnomAD minor allele frequency of 0.22 in AFR compared to 3.97e-04 in EUR. It is also in linkage equilibrium (r^2 < 0.05) with the *APOL1* G1 and G2 known risk alleles for renal disease. Furthermore, the association between rs11089781 under a recessive model and ESRD is strongest in individuals who have low *APOL1* risk and carry either 0 or 1 copy of the risk alleles (OR = 1.79, p = 2.48e-03) compared to those who carry 2 copies of the risk alleles (OR = 0.98, p = 0.914). Using laboratory values for creatinine and estimated glomerular filtration rate (eGFR), we found that rs11089781 is also significantly associated with increased levels of max creatinine (beta = 0.07, p = 2.96e-05) and decreased levels of max eGFR (beta = -0.08, p = 2.36e-06). Replication of this association in up to 121,790 AFR individuals from the Million Veterans Program also yielded significant associations, both with ESRD (OR = 1.16, p = 1.01e-08) and mean eGFR (beta = -0.04, p = 2.42e-13). Initial hypotheses suggest that *APOL3* may play a protective role against *APOL1* and loss-of-function in *APOL3* increases susceptibility to *APOL1*-induced kidney dysfunction, though further functional characterization is still required.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1615 Investigating genetic architecture of cognitive functioning in correlation to metabolic traits in the Indian population

Authors:

S. CHAKRABORTY[1,2], SANSCOG Study Team, J. S. Sundarakumar[2], B. Kahali[3]; [1]Indian Inst. of Sci., Bangalore, India, [2]Ctr. for Brain Res., Bangalore, India, [3]Ctr. for Brain Res., Indian Inst. of Sci., Bangalore, India

Several large-scale genome-wide studies on general intelligence and common executive function phenotypes have established association of hundreds of independent genomic loci with cognition. Such large-scale association studies on either general cognition or domain-specific cognitive measures based entirely on the Indian population has not yet been conducted. Through our ongoing study on healthy individuals (aged 45 years and above) residing in south India, we aim to study the genetic landscape of normal cognitive functioning in a population-specific manner. In our study of ~3000 individuals, we have imputed ~400,00 (post quality check) genetic variants to three reference imputation panels - the haplotype reference consortium (HRC), TOPMed and our in-house panel based on whole-genome sequencing of Indian samples. The imputation accuracy of our in-house panel is higher than that of TOPMed and HRC panel for rare, low-frequency and common variants. With these imputed genotypes from our in-house panel scanning ~13 million markers, we carry out genome-wide association analysis (GWAS) for Hindi Mini Mental State Examination (HMSE) scores which assess general cognitive functioning and COGNITO (Computerized assessment of adult information processing) scores which report scores for diverse cognition domains adapted to rural settings. We also carry out GWAS for anthropometric and metabolic traits, namely body mass index, waist-hip ratio adjusted for BMI, waist-height ratio adjusted for BMI, fasting blood sugar, fasting insulin, HbA1c, HDL, LDL, triglycerides, and total cholesterol, since evidence from a multitude of studies highlight that age-associated metabolic risks affect cognitive functioning and could contribute to cognitive decline in ageing. We report association signals for each of the phenotypes and assess novelty by comparing with variants in the proximity (±500kbp and LD R^2>0.8) listed in EBI-GWAS catalogue associated with cognition and related neurological traits. Our initial analysis for HMSE has yielded one novel genome-wide UTR variant in chr 7(p11.2) and several suggestive novel associations in intronic and UTR regions of genes, a few of which are brain-specific eQTLs, and have been implicated with neuronal mechanisms and neurotransmission. With one or multiple combinations of the above metabolic markers, we find suggestive causal association with general cognition. The results from this study uncover genetic architecture of complex traits (metabolic and cognitive) in Indians, and forms basis for further studies aiming towards understanding cognitive decline in ageing Indian population.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1616 Investigating Immune Functions of neuronal genes burdening Autism Spectrum Disorder: An integrated gene expression and mutation analysis

Authors:

A. Siddappa Niranjana Murthy[1], S. V.P.[2], M. K.R.[3], R. Nallur B[3]; [1]Natl. Inst. of Mental Hlth.and NeuroSci.s, Bengaluru, India, [2]St Johns Res. Inst., Bengaluru, India, [3]Univ. of Mysore, Mysore, India

Autism Spectrum Disorder (ASD) is a neurodevelopmental disorder characterized by social-communication deficits, restricted-repetitive behaviours and interests. Neuroinflammation is often implicated in ASD pathogenesis. Recent studies in rats indicate immune activation deregulates the expression of genes associated with synaptogenesis, axonal guidance, synaptic contact and neurogenesis. Thereby, we explored the overlapping immune functions of neuronal genes through systematic in-silico functional analysis of ASD-associated genes from databases (db) like SFARI-db, Denovo-db and Genes4denovo-db. Enrichment analysis was performed with EnrichR to identify genes with prominent neuro-immune functions that were further subjected to protein-protein interaction (PPI) and tissue-specific functional network analysis on STRING and HumanBase, respectively. Prioritised neuro-immune genes were investigated for ASD brain-tissue specific gene expression, biological functional similarity with high-risk ASD genes using tool GOSemSim and were checked for mutations within 15 in-house WES data from Indian ASD children obtained on Illumina HiSeq 2500. From three databases ten thousand ASD-associated genes were collated; 1518 genes had neuroimmune functions, including 950 genes with statistically significant PPIs. Of these, 618 and 590 genes displayed neuronal and immune cell specific functional networks, respectively with 193 common genes holding prominent neuro-immune functions. Their involvement in ASD pathogenesis were validated using ASD brain post-mortem tissues and 15 in-house ASD WES data. Genes BDNF, ANK3, PLCB1, SET, SYT1, LAMB1, PRKAR1, BRCA2, NTRK1 were found differentially expressed in brain tissues and carried deleterious variation in our WES data. Thus, our study provides the first genetic evidence for neuroinflammation commonly observed in ASD subjects.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1617 Investigating the additive and interactive interactions between ADHD polygenic risk and environmental features on ADHD

Authors:

M. Shoaib[1], C. Burton[1], N. Rosalin[1], P. Arnold[1], P. Peters[2], R. Schachar[1], J. Crosbie[1]; [1]Hosp. for Sick Children, Toronto, ON, Canada, [2]Carleton Univ., Ottawa, ON, Canada

Attention-deficit/hyperactivity disorder (ADHD) is a neurodevelopmental disorder with a polygenic component. According to recent evidence, air pollutants, NO2 and fine particulate matter (PM2.5) may increase risk for ADHD while residential greenspace may decrease risk for ADHD among children. The aim of this study is to test the associations between ADHD polygenic risk (PRS) and environmental exposures (NO2, PM2.5 and greenspace) on ADHD trait scores using additive and interactive models.

The study utilizes participants from the Spit for Science population-based study (age 6 to 17 years) where ADHD traits were measured using Strengths and Weaknesses of ADHD Symptoms and Normal Behavior Rating Scale (SWAN). We estimated PRS using a principal component extension to PRSice-2 in genotyped European ancestry individuals (n= 3197) using SNP effect sizes from the largest available ADHD genome wide association study. Environmental exposures (NO2, PM2.5, greenness) were obtained from linking geospatial coordinates from Canadian Postal Codes to Canadian Urban Environmental Health Research Consortium (CANUE) databases. We implemented regression models to test both additive (PRS + E) and interactive (PRS x E) effects on SWAN scores. Each analysis included age, sex and four population principal components as covariates by considering three environmental variables in separate models.

Preliminary analysis revealed a significant association between ADHD PRS and ADHD traits in all additive models (p-values < 0.001). We found a negative association between greenness and SWAN scores trending towards significance (p-value = 0.09) implying exposure to more green space might be associated with lower ADHD traits. Other environmental features NO2 and PM2.5 were not significantly associated with ADHD traits in additive regression models. There were no significant interactions between ADHD PRS and any of the environmental measures on ADHD traits (PRS x NO2 or PM2.5 or greenness).

In this study, we have tested the additive and interactive interactions between ADHD genetics and environment based on the polygenic risk of ADHD individuals. In additive models, we found PRS as a significant predictor of ADHD traits in population-based samples. We observed higher green spaces potentially be associated with low ADHD traits with a marginal significance. No significant interactions emerged between ADHD genetic risk and environmental features. The initial findings motivate us to extend our analysis in a larger sample size to improve power, the next phase of this study will be conducted using a larger cohort (N= 15000) and we will apply PRS methods suitable for diverse ancestry groups

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1618 Investigating the genetic and phenotypic landscape of Ectodermal Dysplasia.

Authors:

C. Munderloh, R. Belanger Deloge, E. A. Rivera-Munoz, H. Du, R. A. Lewis, J. E. Posey; Baylor Coll. of Med., Houston, TX

Ectodermal Dysplasias (EDs) encompass a diverse group of inherited disorders of the embryonic ectoderm that demonstrate substantial genetic and phenotypic heterogeneity. Currently, 49 ED subtypes have a known molecular etiology involving genes encoding critical components of the *TP63*, *WNT*, and *EDA* developmental pathways (~36%), genes in other varied pathways (~8%), and genes encoding structural framework proteins (~11%). The molecular etiologies of the remaining 40 clinically defined EDs are unknown presently. We recognize that many ED subtypes and syndromes meet the criteria for an ED clinical diagnosis via involvement of two or more affected ectodermal derivatives. Comprehensive characterization of ED's genetic and phenotypic spectrum is required to better understand ED and enhance diagnostic evaluation and molecular testing. In this study, we recruited a cohort of 196 individuals from 48 ED families with no firm molecular diagnosis but with observed and suspected modes of inheritance consistent with previously described autosomal dominant, autosomal recessive, and X-linked forms of ED. We collected blood samples from 196 individuals, consisting of 68 affected individuals and 128 unaffected individuals. Generation and analysis of exome and genome sequencing data to identify potentially deleterious variants in both established and/or novel candidate ED genes are ongoing. To date, an analysis of eight cases has revealed three notable findings. These include the identification of a pathogenic variant in the known disease gene *TP63* (NM_003722.5:c.953G>A (p.R318H)) and variants of unknown significance in two novel candidate genes involved in epidermis formation, *KDF1* and *KRT5* (NM_152365.3:c.795C>A (p.D265E); NM_000424.4:c.1682T>G (p.L561R)). We are presently performing further investigations to determine whether *KDF1* and *KRT5* indeed represent novel ED disease genes, and also performing phenotypic clustering analysis of studied cases. This work will inform the phenotypic and molecular contributors that drive ED subtypes, yield potential phenotypic expansions, and reveal genotype-phenotype correlations.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1619 Investigating the Interplay of Sex and the Genetic Risk Factor, APOE, Across Ages in Alzheimer's Disease at a Single Cell Resolution

Authors:

Y. Li[1], Y. Hao[2], C. Pereda Serras[1], S. Yoon[2], B. Grone[2], L. Ding[2], J. Blumenfeld[2], T. Oskotsky[1], Y. Huang[2], M. Sirota[1]; [1]Univ. of California, San Francisco, San Francisco, CA, [2]Gladstone Inst.s, San Francisco, CA

Background: Old age, apolipoprotein E4 (APOE4) genotype, and the female sex are well-established risk factors for Alzheimer's Disease (AD). Almost two-thirds of AD patients are women and most are APOE4 carriers. We aim to study the interplay between these risk factors by analyzing single-nucleus RNA-sequencing (snRNA-seq) data from the hippocampus of human APOE4 and APOE3 knock-in (KI) female and male mice across ages.
Methods: A sex-balanced cohort of APOE3-KI and APOE4-KI homozygous mice were sacrificed at 6 (young), 12 (adult), and 18 (old) months of age. Single nuclei were prepared from the hippocampus of each mouse for snRNA-seq. Demultiplexed fastq files were aligned to a custom reference genome mm10-1.2.0, using the Cellranger v2.0.1. Quality assurance, count normalization, clustering, cell type identification, and differentially expressed gene (DEG) analysis were performed using the Seurat.
Results: Our dataset consists of a total of 527,395 cells covering 27,153 genes. A total of 36 cell clusters were identified, with six major brain cell types (astrocytes, microglia, oligodendrocytes, oligodendrocyte precursor cells, and excitatory and inhibitory neurons). Analysis of the abundance of each cell type showed no differences in cellular composition across ages, APOE genotypes, and sexes in all cell types except astrocytes. Astrocyte abundance was significantly higher in old females versus old males (padj=0.025, Tukey's multiple comparison test) and in young APOE4 versus young APOE3 males (padj=0.042).
Sex and age-stratified DEG analysis comparing APOE4 versus APOE3 astrocytes revealed many shared DEGs between young and adult males but not with old males, while females had few consistent DEGs across the three age groups. We also identified several DEGs associated with plasma membrane assembling and post-synaptic transmission/signaling that are upregulated in old APOE4 versus old APOE3 females while downregulated in young APOE4 females and young and adult APOE4 males versus their APOE3 counterparts. Analysis of interactions among cell types using CellChat revealed sex-biased communication from glial cells to inhibitory neurons.
Conclusion: We identified a clear interplay between age, APOE genotype, and sex in mouse models of AD and revealed the interaction and convergence of these risk factors at a single-cell resolution.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1620 Investigating the pleiotropic genetic architecture of impulsivity across biological scales

Authors:

T. Mallard[1,2], M. Jennings[3], Y. Zhang[4], J. Tubbs[1], D. Gustavson[5], A. Grotzinger[5], M. Westwater[6], 23andMe Research Team, S. Elson[7], P. Fontanillas[7], T. Ge[1], K. Choi[1], J. Smoller[1], A. Palmer[3], S. Sanchez-Roige[3]; [1]Massachusetts Gen. Hosp., Boston, MA, [2]Harvard Med. Sch., Boston, MA, [3]Univ. of California San Diego, La Jolla, CA, [4]Harvard T.H. Chan Sch. of Publ. Hlth., Boston, MA, [5]Univ. of Colorado Boulder, Boulder, CO, [6]Univ. of Oxford, Oxfordshire, United Kingdom, [7]23andMe, Inc., Sunnyvale, CA

Background: Impulsivity is an important mental construct implicated in various psychiatric conditions, but consensus on its measurement and corresponding biology is lacking. Neuroscientific insights support 'splitting' impulsivity into distinct facets; however, psychological research shows that many mental constructs can be 'lumped' together at multiple levels of analysis. Here, we use multivariate genetic methods to advance genomic discovery for impulsivity by (i) testing the splitting vs lumping hypotheses as they relate to common variant architecture and (ii) characterizing pleiotropy across biological scales.
Methods: We used the Genomic SEM framework to conduct a multivariate analysis of eight impulsivity facets measured in 23andMe (*N*s ~130k), employing two statistical models: (1) an "omnibus" model that provides an unstructured meta-analytic test of association, and (2) a common factor of impulsivity, where genetic associations operate via a latent genetic factor. To gain a comprehensive understanding of the etiology of impulsivity, we examined these models across molecular genetic, transcriptomic, and neurogenomic levels. We then conducted polygenic score (PGS) analyses in two independent cohorts to characterize links with human health and wellbeing.
Results: Results revealed widespread pleiotropy among impulsivity facets, with most trait pairs having moderate-to-large genetic correlations (mean r_g = 0.43, range = -0.09-0.79). Genomic factor analysis indicated that the observed patterns of covariance could be approximated with a common factor model (CFI = 0.94, SRMR = 0.08). However, GWAS analyses revealed that SNP-level effects were largely inconsistent with this model of general liability. This pattern was also seen in multivariate TWAS and neurogenomic analyses, where impulsivity facets had many trait-specific patterns of association, including heterogeneous signals in the inferior frontal gyrus, a region strongly implicated in self-control (all Q_{Trait} P < 2.43e-7 for the pars orbitalis, pars triangularis, and pars opercularis). Finally, PGS results underscored the consequences of lumping impulsivity facets, as 67 links to health outcomes were lost when a common factor PGS was used, including novel links between urgency and neurodevelopmental disorders.
Discussion: By formally comparing splitting versus lumping models, these findings generate new insights into the genetic architecture of impulsivity, implicating novel risk genes and links to human neurobiology and health. Our results support multidimensional approaches as a more informative avenue for studying the heterogeneous biology of impulsivity.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1621 Investigating the underlying genetic architecture of personality traits and overlap with psychopathology in 240,000 US military veterans.

Authors:

P. Gupta[1,2], M. Stein[3,4], S. Beck[1,2], K. Adhikari[1,2], J. Gelernter[1,2], D. Levey[1,2]; [1]Yale Sch. of Med., New Haven, CT, [2]VA Connecticut Hlth.care System, West Haven, CT, [3]Univ. of California San Diego, San Diego, CA, [4]VA San Diego Hlth.care System, San Diego, CA

Background: Personality is influenced by both genetic and environmental factors and is associated with to other complex traits such as anxiety and depression. Of the five commonly classified personality traits - neuroticism, agreeableness, extraversion, conscientiousness and openness, neuroticism has most often been the focus of genetic studies.
Methods: The Million Veteran Program (MVP) cohort has around 900,000 enrollees, with more than 650,000 so far with available genetic data and also has survey response data from the Big Five Inventory-Short. The Big Five assesses five personality traits: neuroticism, agreeableness, extraversion, conscientiousness and openness. We conducted GWAS in the MVP of each trait in approx. 240,000 individuals with available survey personality data. We conducted linkage disequilibrium score regression to calculate heritability and genetic correlation among personality traits and each with psychopathologies.
Results: We identified genome-wide significant loci associated with the 5 personality traits (7 with neuroticism and openness, 11 with extraversion, 2 with conscientiousness and 3 with agreeableness). Gene expression associations were found in TWAS analysis for each of the trait with highest number of associations observed for neuroticism and extraversion. Further, neuroticism showed highest positive correlation with anxiety (rG=0.69) while agreeableness showed lowest negative correlation (rG= -0.31) with anxiety suggesting a differing overlap in their respective genetic architecture.
Conclusions: We provide evidence for a substantial differential genomic architecture underlying the personality traits. Our work yields novel insights and opens new questions about the influence of genetic factors on personality and how this intersects with psychopathology.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1622 Investigation of familial hypercholesterolemia prevalence and association with cardiovascular disease in an African ancestry cohort.

Authors:

A. Winters, M. Kelly, N. Mohammed, A. Berry, L. Jones, S. Gidding, M. Oetjens; Geisinger, Danville, PA

Familial hypercholesterolemia (FH) is an inherited condition consisting of lifelong elevated LDL cholesterol levels and increased risk of cardiovascular disease. Variants in three genes involved in LDL receptor function, *LDLR*, *APOB*, and *PCSK9*, account for most cases. Among Western populations, the prevalence of FH has been reported to be between 1:212 to 1:331. African ancestry individuals are underrepresented in submissions to genetic resources used to determine variant pathogenicity and there are no known reports on the prevalence of FH in those populations. To create an African ancestry cohort, we examined genetic data from Geisinger's MyCode (N = 175,500) and the NIH's All of Us (N = 98,590) cohorts with principal component analyses to identify individuals who cluster with 1000 Genomes African superpopulation (1KG AFR). We assessed sequence variants in FH genes from our African ancestry cohort for pathogenicity using American College of Medical Genetics criteria. Lipid levels and use of lipid lowering medications were compared among those with and without a variant. We identified 29,931 individuals who clustered with the 1KG AFR group. Among these, we identified 31 unique pathogenic or likely pathogenic FH variants among 108 individuals. Most of the pathogenic variants identified (94%) were missense variants in the *LDLR* gene. This gave a prevalence of FH of 1:277 among African ancestry individuals. Mean age for individuals with a pathogenic FH variant was 40.5 years in MyCode and 50.4 years in All of Us. Of those with a pathogenic FH variant and at least one LDL-C measurement, we found that 52.5% had an LDL-C &gt190mg/dl and 37.5% had an LDL-C &gt250 mg/dl, compared to 5.7% and 0.6% of African ancestry individuals without an FH variant. The effect size of an FH variant on maximum LDL-C levels adjusted for age, sex, and the first four principal components of ancestry was 90.4 mg/dL (95% CI [69.8, 111.0], p=1.1x10-17) in MyCode and 126.5 mg/dL (95% CI [110.5, 142.6], p=4.0x10-53) in All of Us. In the combined cohort, 29.6% of individuals with a pathogenic FH variant had been prescribed an LDL-lowering medication, compared to 20.9% of individuals without a variant. We identified a prevalence in an African Ancestry cohort of FH consistent with that observed in studies of European ancestry cohorts. The effect size of an FH variant in this cohort was large. These data suggest that African ancestry individuals with pathogenic FH variants have a high likelihood of increased LDL and concomitant cardiovascular disease risk and that there is an opportunity for additional identification and treatment of FH in this population.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1623 Investigation of the genetic basis of pediatric psoriasis reveals contributions from both common risk alleles and rare *de novo* mutations.

Authors:

J. Frost[1], S. Kumar[2], M. Fang[3], T. Erickson[3], A. Hajek[3], A. Kwon[3], J. Liu[2], A. Zhu[1], R. Pandya[1], S. Rangel[3], M. Lebwohl[1], J. Oksenberg[4], M. Seyger[5], W. Liao[2], A. Paller[3], A. Bowcock[6]; [1]Icahn Sch. of Med. at Mount Sinai, New York City, NY, [2]Univ. of California, San Francisco, San Francisco, CA, [3]Northwestern Univ., Feinberg Sch. of Med., Chicago, IL, [4]Univ California San Francisco, San Francisco, CA, [5]Radboud Univ. Nijmegen Med. Ctr., Nijmegen, Netherlands, [6]Icahn Sch. of Med. at Mount Sinai, New York, NY

An understanding of the genetic basis of psoriasis can help in its management, including treatment choice. The need is particularly high for children with psoriasis, which is often a life-altering, chronic disorder associated with extracutaneous comorbidities. Our previous studies revealed that some pediatric psoriasis patients harbor highly penetrant mutations in *CARD14*, but these are rare or *de novo*. The goal of the current study was to examine the genetic underpinnings of pediatric psoriasis via the identification of *de novo* mutations, the presence of HLA class I risk alleles, and weighted genetic risk score (wGRS). Following exome-sequencing of 41 trio families (pediatric patients with no first-degree family history of psoriasis) we first imputed HLA class I alleles with "Optitype". We observed that 60% of pediatric psoriasis cases, including the six guttate psoriasis cases, harbored at least one copy of the major known psoriasis risk allele HLA-C*06:02. None of the probands harbored *CARD14* mutations so we then searched for *de novo* mutations to identify novel highly penetrant coding variants predisposing to psoriasis. In 23 probands we identified 37 *de novo* mutations in multiple genes. A number of these genes encoded proteins with a role in pathogen responsiveness such as *LCAT, SPNS1, CORO1A, SNX6*, and *AMPD3*. Pathogens have long been considered a trigger for psoriasis. One case harbored a p.Arg90Cys alteration in *SNAP29,* a gene mutated in the rare autosomal recessive disorder CEDNIK (Cerebral dysgenesis, neuropathy, ichthyosis, and keratoderma syndrome). This is intriguing given the fact that ichthyosis, like psoriasis, is associated with an altered skin barrier and skewing towards Th17 immune activation. We also computed the wGRS of the trios to understand the genetic influence of common psoriasis risk alleles in pediatric patients. The wGRS were derived from a GWAS meta-analysis (N >39,000), which had revealed 88 independent SNPs associated with psoriasis. These data suggested that both pediatric psoriasis cases and their unaffected parents have a greater wGRS than population healthy controls, indicating an important role for common variants in the pathogenesis of pediatric psoriasis and inheritance of these alleles from silent carrier parents. Overall, these studies suggest that both common and rare variants contribute to psoriasis in the pediatric population.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1624 Investigation of the sex chromosomes in Autism Spectrum Disorder.

Authors:

M. Mendes de Aquino[1,2], W. Engchuan[1,2], B. Thiruvahindrapuram[1,2], E. McKee[1,2], B. Trost[1,2], R. Alexandrova[1,2], T. Peixoto Leal[3], W. Li[1,2], S. Scherer[1,2,4]; [1]The Ctr. for Applied Genomics, The Hosp. for Sick Children, Toronto, ON, Canada, [2]Genetics and Genome Biology Program, The Hosp. for Sick Children, Toronto, ON, Canada, [3]Cleveland Clinic, Cleveland, OH, [4]McLaughlin Ctr. and Dept. of Molecular Genetics, Univ. of Toronto, Toronto, ON, Canada

Autism Spectrum Disorder (ASD) is a neurodevelopmental condition characterized by varying degrees of social communication deficits, restricted interests, and repetitive sensory-motor behaviours, which affects approximately 1% of the global population. However, the prevalence of ASD differs between males and females, with a commonly reported ratio of 4:1; this ratio may increase to 11:1 if Asperger's syndrome is included as a form of ASD. There have been many X-linked genes identified to carry rare ASD-associated variants in males, having even a hypothesis of a potential protective effect of having two X chromosomes and the risk effect associated with the Y chromosome. With many new resources, there is now the opportunity to investigate all aspects of genetic variations on the sex chromosomes in ASD. Here, we explored ASD genetic associations in both sex chromosomes, taking into account their differential recombination patterns and the inactivation of the X chromosome in females. We performed an X-chromosome Wide Association Study (XWAS) and a Y-haplogroup inference, followed by logistic regression for testing statistical association. These methodologies were applied to ASD whole genome sequence resources, including the MSSNG, Simons Simplex Collection and SPARK, with a total of 6,873 cases for the XWAS and 5,835 ASD-affected males for the Y chromosome analyses. To conduct the XWAS, we utilized as controls samples from HostSeq (Host Genome Sequencing Initiative; 6,456 samples) and MGRB (Medical Genome Reference Bank; 2,525 samples). Additionally, for the Y-haplogroups analysis, we included WGS data from two pediatric control cohorts; Canadian Healthy Infant and Longitudinal Development (CHILD) and Inova Health System Cohort (INOVA). We conducted both association tests using European samples and controlling for population structure. From XWAS, we identified a specific region of linkage disequilibrium (LD) in the Xp22.2 that showed significant association with ASD when examining only males (top variant with p-value of $3.57x10^{-7}$, odds ratio (OR) [95%CI]: 0.8 [0.73 - 0.87], global frequency=0.35). For the Y chromosome analysis, we did not find any major Y-haplogroups associated with ASD. However, we discovered one final haplogroup branch, belonging to the I major haplogroup, that exhibited significant association results, with a p-value of $7.27x10^{-7}$ and an OR of 3.03. The identified signals were cross-referenced with phenotype data, family information, gene expression, and gene-set differentiation. Our findings open new avenues of investigation into the genomic architecture in ASD.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1625 IPMI: Precision Medicine Genomics Project Advancing Integrated Polygenic Risk Score Development and Drug Target Discovery

Authors:

S-G. Park[1], J. Choi[2], J. Ko[3], E-K. Kwon[4], S. Lee[1], Y. Lee[3], H. Kim[4], B. Choi[1], J. Kim[4], S-S. Byun[5]; [1]Invites BioCore, Yongin, Korea, Republic of, [2]Jeju Natl. Univ. Hosp., Jeju Natl. Univ. Sch. of Med., Jeju, Korea, Republic of, [3]Invites Genomics, Jeju, Korea, Republic of, [4]Procagen, Seongnam, Korea, Republic of, [5]Seoul Natl. Univ. Bundang Hosp., Seongnam, Korea, Republic of

In this project, called IPMI (Invites Precision Medicine Initiative), the primary objective is to develop an assessment method called iPRS (Integrated Polygenic Risk Score) and identify new drug targets for various diseases. The study will involve the collection of comprehensive data from over 50,000 individuals of South Korean ancestry. IPMI initiated the project in December 2022, and has collected data from around 10,000 individuals to date. The project is expected to be completed by the end of 2025. This data includes genetic information obtained through whole genome sequencing (WGS) and a customized SNP array, clinical data obtained from medical records and surveys, and lifelog information collected from wearable devices and smartphone apps. The iPRS assessment method combines genetic and clinical information to predict an individual's risk of developing diseases like cancer or cardiovascular disease. By identifying high-risk individuals before they display symptoms, early detection and prevention can be facilitated. The IPMI project aims to develop iPRS for different diseases, including cancer, cardiovascular disease, type 2 diabetes, and neurovascular disease. Additionally, the IPMI project aims to discover new drug targets for these diseases. Drug targets are specific molecules or pathways that medications can target to treat diseases effectively. By identifying novel drug targets, the IPMI project intends to accelerate the development of new and improved treatments. Overall, the IPMI project represents a significant endeavor in the field of medicine. Through the development of iPRS and the discovery of new drug targets, it has the potential to enhance prevention, early detection, and treatment of diseases, thereby improving health outcomes worldwide. Furthermore, the IPMI project is expected to contribute to the growth of the digital healthcare market and the precision medicine industry.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1626 Japanese biobank whole genome sequencing analysis identifies hearing loss risk genes for drug discovery.

Authors:

T. Monjo[1], P. Yang[1], Y. Ling[1], T. Saito[2], S. Mugikura[2], S. Ogishima[2], F. Nagami[2], K. Kinoshita[2], S. Kuriyama[2], M. Yamamoto[2], T. Ando[1]; [1]Takeda Pharmaceutical Company Limited, Fujisawa, Kanagawa, Japan, [2]Tohoku Med. Megabank Organization, Tohoku Univ., Sendai, Miyagi, Japan

Identification of genetically validated novel drug targets is absolutely required to treat hearing loss (HL) because retrospective studies suggested that genetically supported targets were more likely to be successful in drug development. Tohoku medical megabank organization (ToMMo) is a Japanese biobank that consists of over 150,000 participants' biospecimens and deep phenotyping information (e.g. physiological measurements, questionnaire, and imaging). The partnership between ToMMo and Japanese pharmaceutical companies are proceeding a huge project that sequences whole genome of 100,000 ToMMo participants. In this presentation, we show our drug discovery strategy using genetic data and a case study that used currently available approximately 50,000 WGS data and hearing test phenotypes. Whole genome sequencing (WGS) enables us to detect rare deleterious variants related to various diseases. To identify potential risk genes to HL, we conducted association tests using 40,449 individuals' WGS data in ToMMo. To be specific, we firstly defined hearing ability phenotypes from hearing tests of left and right ears and performed gene-based association tests of rare protein-truncating variants (e.g. frameshift variant, stop gained variants; minor allele frequency < 1%) using scalable linear-mixed model (i.e. Regenie) to increase statistical power. As a result, we found significant genes that were associated to hearing test phenotypes after Bonferroni correction. Some genes were previously reported as a causal gene of hereditary HL. In conclusion, we identified the potential genetic associations with HL including known causal genes using currently available half of WGS data. Further replication with the full dataset and wet validation study would investigate the gene for potential drug development programs by generating and supporting therapeutic hypothesis of HL.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1627 Joint analysis of de novo mutations in autism spectrum disorders, schizophrenia, congenital heart disease, and other developmental disorders improves detection power and implicates shared molecular pathways and CNS processes

Authors:

M. Kealhofer, B. Riley, T-H. Nguyen; Virginia Commonwealth Univ., Richmond, VA

Background: Exome studies have previously implicated shared risk genes and pathways for autism spectrum disorders (ASD), congenital heart disease (CHD), schizophrenia (SCZ), and other developmental disorders (DD). However, an in-depth analysis for this shared information has not been conducted on current large-scale datasets of these disorders. Here, we aim to use a two-trait Bayesian integrative analysis approach to increase statistical power for gene discovery, and to better understand the biological information of genes shared between these disorders.
Methods: We used mTADA to perform joint pair-wise analysis of exome variants identified from 6,430 ASD trios, 3,154 SCZ trios, 2,645 CHD trios, and 31,058 trios with severe DD to identify shared risk genes. We used ClusterProfiler to assess if those risk genes cluster in GO and KEGG genesets. Expression weighted cell type enrichment was performed using EWCE.
Results: All 6 pairwise analyses showed evidence of overlapping DNV signal to varying degrees, with shared risk genes identified using a cut of >80% on their respective posterior probabilities. From the 6 shared risk gene sets, 5 were taken forward for post-hoc systems biology analysis (20 SCZ-ASD genes, 41 SCZ-DD genes, 9 ASD-CHD genes, 153 ASD-DD genes, and 42 CHD-DD genes) due to only 2 shared risk genes being identified in the SCZ-CHD analysis. KEGG and GO gene set analysis revealed that the identified risk gene sets clustered in processes relating to synaptic organization and signaling, histone binding and modification, and learning and cognition. Risk gene sets from pairwise analyses including CHD additionally implicated gene sets relating to response to fibroblast growth factor and adherens junction assembly. Expression weighted cell type enrichment analysis using a reference set of mouse brain single cell expression showed the genes in the SCZ-ASD, SCZ-DD and ASD-DD risk gene sets had significantly elevated expression in pyramidal cells and interneurons, with the SCZ-DD risk gene set showing additional enrichment in vascular smooth muscle cells. Signals for the DD-CHD risk gene set were less clear, with some implication of pyramidal cells but also a trend towards involvement across more cell types in the brain. Enrichment in the ASD-CHD risk gene set did not reach significance, but had similar trends to the SCZ-ASD, SCZ-DD and ASD-DD risk gene sets.
Discussion: The integrative analytic approach increased the power to detect trait-associated risk genes, as well as shared risk genes in pair-wise analyses. Post-hoc systems biology analyses identified important CNS and epigenomic processes that may be implicated in the etiologies of ASD, CHD, SCZ and DD.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1628 Lack of association between *Mir27a, PPARG, Mir146a* and *Traf6* polymorphisms and type 2 Diabetes in the Tunisian population.

Authors:

H. Dridi[1], R. Gereisha[1], H. Ben Hassine[1], L. dardour[1], E. chabchoub[1], Y. Hasni[2], A. Ben Abdelkrim[2], M. Chaied[2], R. zemni[1], F. Ben Hadj Slama[1]; [1]Faculty of medecine of sousse, Sousse, Tunisia, [2]Farhat Hached Hosp., Sousse, Tunisia

Background : Type 2 diabetes (T2D) is a prevalent metabolic disorder and multifactorial disease with several genetic factors associated. MicroRNAs (miRNAs), small noncoding RNA molecules approximately 21 nucleotides in length, have emerged as crucial regulators in numerous biological processes, including diabetes pathogenesis. Single nucleotide polymorphisms (SNPs) are potential contributors to miRNA dysregulation, as they can impair miRNA function and influence disease outcomes. Notably, miRNA 146a rs2910164 and miRNA 27a rs895819 have garnered interest due to their potential association with diabetes susceptibility. Objectives: The aim of the present study was to investigate the association of the single nucleotide polymorphisms *Mir146a* rs2910164, *Mir27a* rs895819 and their target genes respectively *TRAF6* rs540386, *PPARG* rs1801282 with Diabetes type2. Material and methods: *TRAF6* rs540386, *Mir146a* rs2910164, *Mir27a* rs895819 and *PPARG* rs1801282 genotyping were performed by mutagenically separated polymerase chain reaction (MS-PCR with newly designed primers) in 235 type2 diabetes patients and 225 age- and gender-matched controls. Results: The association analysis between these single nucleotide polymorphisms (SNPs) and type 2 diabetes in Tunisian Population revealed that *Mir146a* is not significantly associated with the disease (p=0.99, odds ratio [OR]=0.97 (0.53-1.7)), and similarly, *TRAF6* is also not significantly associated (p=0.72, OR=1.31 (0.56_3.1)). Also, there was no association between *Mir27a* rs895819 (*p*=0.33, OR=1.09 (0.55-2.16)) and *PPARG* rs1801282 (p= =0.27, OR=0.67 (0.11-4.05)) and type2 diabetes in our cohort. Conclusion: Our study revealed that *Mir146a*, *Mir27a*, *TRAF6* and *PPARG* polymorphisms are probably not susceptibility factors to type2 Diabetes in the Tunisian population. Further investigations are needed to elucidate the role of these gene polymorphisms.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1629 Lack of association between Native American genetic ancestry and the development of severe/critical COVID-19 in a Peruvian population sample.

Authors:

J. Mesia[1,2], S. Gonzales-Magallanes[1,3], E. Mamani-Pariona[1,4], R. Juscamaita[1,3], E. Tarazona-Santos[5], M. Santolalla-Robles[1]; [1]Univ. Peruana Cayetano Heredia, Lima, Peru, [2]Univ. Ricardo Palma, Lima, Peru, [3]Univ. Natl. Mayor de San Marcos, Lima, Peru, [4]Univ. Natl. Federico Villareal, Lima, Peru, [5]Univ.e Federal de Minas Gerais, Belo Horizonte, Brazil

Background: The COVID-19 pandemic has highlighted the need of studying potential risk factors that may contribute to susceptibility to infectious diseases. Genetic ancestry can influence the development of several diseases, unfortunately, most of reports does not Latino populations. Native American genetic heritage represents about 80% of the Peruvian population. This research aimed to investigate the association between global genetic ancestry and the onset of severe/critical COVID-19 disease in a sample of a Peruvian population. **Methods:** A case-control study was conducted with 239 participants where 130 subjects were clinically diagnosed with severe/critical COVID-19 (cases) and 109 subjects with asymptomatic, mild or moderate disease (controls). The genotyping was performed using the Infinium Exome-24 chip and the ADMIXTURE algorithm was used to examine the markers and estimate genetic ancestry ratios. The statistical analysis was performed in R Studio v4.1.2. **Results:** Our results did not demonstrate a statistical association between genetic ancestry and COVID-19 severity, however, we were able to establish a correlation between Native American ancestry and the presence of metabolic comorbidities, which could indirectly contribute to the development of severe/critical COVID-19. It was observed that a higher Native American ancestry, higher odd of having a metabolic diseases (OR= 1.21-53.82). The confidence interval is very wide since the number of samples is limited. These findings imply that genetic ancestry may have an effect on COVID-19 outcome through its relationship to comorbidities, however is needed a deep analysis of other socio-economical factors.**Conclusions:** Our study did not find a direct correlation between Native American ancestry and COVID-19 severity in a Peruvian population. However, the observed relationship between Native American ancestry and metabolic comorbidities provides insight into the existence of potential pathways linking them to COVID-19 severity, so additional studies would be required to elucidate possible underlying mechanisms.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1630 Large-scale analysis of patient-derived induced neurons discover early developmental changes in schizophrenia.

Authors:

B. Rowan[1], C. Seah[1], X. Wang[1], M. Breen[1], J. Friedman[1], E. Drapeau[1], K. Brennand[2], L. Huckins[2], T. Rusielewicz[3], D. Paull[3], The NYSCF Global Stem Cell Array Team, J. Buxbaum[1]; [1]Icahn Sch. of Med. at Mount Sinai, New York, NY, [2]Yale, New Haven, CT, [3]New York Stem Cell Fndn., New York, NY

Background: Schizophrenia is one of the most costly and debilitating psychiatric disorders with an incidence of 1% worldwide along with limited advances in therapeutic development in decades. Schizophrenia shows high heritability from both common and rare variants. These genetic variants and their mechanisms are poorly understood. Potential differences in ancestry have also not been explored, largely due to lack of collecting enough samples from non-European ancestral populations. Differences by sex also warrant further attention, since multiple studies point to differences in prevalence and potential treatments for schizophrenia by sex. We propose to address these issues by looking at gene expression in over 120 donors, some of admixed ancestry, with either idiopathic or 22q11-associated schizophrenia. We will look at differential expression and differential networks in cases and controls and stratify by ancestry and sex effects. **Methods**: Donor induced pluripotent stem cells (iPSCs) were induced into glutamatergic and GABAergic neurons, followed by RNAseq. In idiopathic cases, we performed differential gene expression (DEG) via limma voom, with a weighted least square linear regression model for GABAergic neurons and a linear mixed model was used for glutamatergic neurons due to cloning. We performed enrichment analysis, and then compared results to Common Mind results in bulk and single cell datasets. We performed weighted gene coexpression analysis (WGCNA) in cases and controls separately and then compared differential expression of modules detected in each one. We also performed WGCNA in a combined case-control dataset. **Results**: DEG analyses were negative for both derived cell types. However, we found statistically significant by multiple correction enrichment for similar genes differentially expressed in Common Mind bulk data and Molecular Signature Database (MSigDB) known schizophrenia risk genes. WGCNA detected some sets of genes more expressed in cases than controls and vice versa. In the combined case-control results, WGCNA detected 2 modules of statistically significant (> 0.05) without multiple testing corrections. One of these modules was also statistically significant for sex differences after multiple test corrections. **Discussion**: We found network module DEGs between cases and controls via WGCNA as well as enrichment for DEGs in Common Mind bulk tissue. We will repeat these analyses in stratified groups of sex, ancestry, and ancestry by sex. Then, we will perform eQTL, local ancestry eQTL, secondary RNA analyses, and compare 22q11 to idiopathic schizophrenia cell lines, resulting in potentially novel therapeutics.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1631 Large-scale genome-wide association study using genomic structural equation modeling provides insights into metabolic syndrome.

Authors:

S. Park[1], S. Kim[2], B. Kim[1], D. Kim[1], Y. Ahn[1], H. Kim[1], M. Song[1], J. Kim[1,3], S-H. Jung[4], C. Cho[1], T. Yu[5], W. Myung[3,6], H-H. Won[1,7]; [1]Samsung Advanced Inst. for Hlth.Sci. and Technology (SAIHST), Sungkyunkwan Univ., Seoul, Korea, Republic of, [2]Broad Inst. of MIT and Harvard, Cambridge, MA, [3]Seoul Natl. Univ. Bundang Hosp., Seongnam, Korea, Republic of, [4]Univ. of Pennsylvania, Philadelphia, PA, [5]Wonkwang Med. Ctr., Wonkwang Univ. Sch. of Med., Iksan, Korea, Republic of, [6]Seoul Natl. Univ., Seoul, Korea, Republic of, [7]Samsung Med. Ctr., Sungkyunkwan Univ. Sch. of Med., Seoul, Korea, Republic of

Introduction: Metabolic syndrome (MetS) is a condition where an individual possesses a combination of multiple metabolic traits such as diabetes, dyslipidemia, hypertension, insulin resistance, and obesity. It is known to be associated with an increased risk of cardiovascular diseases. The genetic component of individual MetS risk factors was widely studied, but their unitary effect on MetS was less focused. We utilized a structural equation modeling (SEM)-based multivariate analysis (i.e., Genomic SEM) to elucidate the complex genetic architecture of MetS.
Methods: We collated publicly available large-scale genome-wide association studies conducted in European ancestry for seven MetS risk factors (i.e., body mass index, waist circumference, type 2 diabetes, hypertension, fasting glucose, triglycerides, and high-density lipoprotein) with sample sizes ranging from 151,188-1,253,277. By combining MetS risk factors through exploratory and confirmatory factor analysis, we constructed a hierarchical factor model that represents the genetic architecture of MetS. The association between single-nucleotide polymorphism (SNP) and MetS was computed by leveraging the genetic covariances between the MetS risk factors.
Results: We identified 1,650 lead SNPs that span across 939 genomic risk loci associated with MetS with an effective sample size of 1,384,348. The estimated SNP-based heritability was 11%. These genetic signals were significantly enriched in neurons and brain tissues. The genome-wide significant MetS SNPs were mapped to the genes through three gene mapping strategies available on FUMA (positional, eQTL, chromatin interaction), and combined with MAGMA gene-based analysis results, we pinpointed 886 genes relevant to MetS. To further prioritize the genes, we conducted Summary-based Mendelian Randomization in brain tissue using the BrainMeta database (n=2,865) and identified 43 significantly associated genes after Bonferroni's correction, where two genes, *FEZ2* and *STRA13*, were further replicated in GTEx (n=205).
Conclusion: Our findings provide new insights into the genetic architecture of MetS. We suggest two genes where their expression level is causally associated with MetS. These genes may contribute to multiple MetS risk factors and may serve as potential therapeutic targets.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1632 Leveraging >8 million patient records across University of California to understand common variable immunodeficiencies.

Authors:

R. Mester, A. Stephens, A. Chin, R. Johnson, S. Knyazev, B. Pasaniuc, M. Butte; Univ. of California, Los Angeles, Los Angeles, CA

Common variable immunodeficiency (CVID) is a rare primary immunodeficiency disorder characterized by defective antibody production and increased susceptibility to recurrent infections. It affects individuals of diverse genetic backgrounds and typically manifests during adolescence or early adulthood. Although the precise etiology remains elusive, recent studies have highlighted the involvement of multiple genes associated with immune system regulation and B-cell function. CVID is a heterogeneous disorder with various genetic and environmental factors contributing to its development. Diagnosing CVID proves to be a formidable task due to its complex nature and heterogeneous presentation. In 2022, Johnson et al introduced PheNet, an algorithm that utilizes a curated list of 197 confirmed CVID cases at UCLA to identify additional patients with CVID from electronic health record (EHR) data. In this work, we introduce a cohort of 323 clinician-identified CVID patients from 5 health systems across the University of California. This case cohort is 68% female and made up of patients with an average age of 56. Of these patients, 73% self-identified as European, 8% identified as Hispanic or Latino, 3% identify as Asian, and 0.6% identify as Black or African American. We use this large, gold-standard case cohort, to demonstrate the heterogeneity of CVID presentation both between and across health systems. We find that across the health systems, these CVID patients have an average of 2.6 immune-related, 4.5 infection-related, and 0.8 other PheCodes that PheNet demonstrated are useful in identifying CVID. Additionally, we assess the generalizability of PheNet to identify potential new CVID cases across the 8.6 million patients in the University of California health system. This population is 54% female with an average age of 46. In this population, 3.6 million patients self-identify as European, 1.4 million patients identify as Hispanic or Latino, 400,000 patients identify as Black or African American, and over 700,000 patients identify as Asian. Finally, we introduce extensions to the PheNet algorithm with the goal of increasing its sensitivity for identifying CVID patients across health systems.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1633 Leveraging a multi-population GWAS meta-analysis to prioritize genes underlying risk loci for central obesity: the Genetic Investigation of ANthropometric Traits (GIANT) Consortium.

Authors:

E. Wilson[1], D. Kim[2], V. Diez-Obrero[3], S. Vedantam[4,5], E. Bartell[4,5,6], A. Pozarickij[7], Z. Fairhurst-Hunter[7], E. Marouli[8], K. Young[2], M. Graff[2], J. Arias[9], G. Chittoor[10], S. Berndt[9], T. Winkler[11], C. Lindgren[12,13], R. Loos[3], K. North[2], R. Walters[7], K. Mohlke[1], A. Justice[10], Genetic Investigation of ANthropometric Traits (GIANT) Consortium; [1]Dept. of Genetics, Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, [2]Dept. of Epidemiology, Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, [3]Novo Nordisk Fndn. Ctr. for Basic Metabolic Res., Univ. of Copenhagen, Copenhagen, Denmark, [4]Dept. of Endocrinology, Boston Children's Hosp., Boston, MA, [5]Broad Inst. of MIT and Harvard, Cambridge, MA, [6]Dept. of Genetics, Harvard Med. Sch., Boston, MA, [7]Clinical Trial Service Unit & Epidemiological Studies Unit (CTSU), Nuffield Dept. of Population Hlth., Big Data Inst. Building, Univ. of Oxford, Old Road Campus, Roosevelt Drive, Oxford, United Kingdom, [8]William Harvey Res. Inst., Barts and The London Sch. of Med. and Dentistry, Queen Mary Univ. of London, London, United Kingdom, [9]Div. of Cancer Epidemiology and Genetics, Natl. Cancer Inst., Bethesda, MD, [10]Dept. of Population Hlth.Sci., Geisinger Hlth.System, Danville, PA, [11]Dept. of Genetic Epidemiology, Univ. of Regensburg, Regensburg, Germany, [12]Wellcome Trust Ctr. for Human Genetics, Univ. of Oxford, Oxford, United Kingdom, [13]Oxford Big Data Inst., Li Ka Shing Ctr. for Hlth.Information and Discovery, Univ. of Oxford, Oxford, United Kingdom

GWAS for central obesity have identified many risk loci, but the underlying mechanisms largely remain to be determined. The GIANT Consortium performed GWAS meta-analyses for waist-to-hip ratio adjusted for body mass index (WHR) in 1.18 million individuals. We identified conditionally distinct signals (p<5e-9) in 18 meta-analyses performed separately and together by sex and population group; we then combined signals across analyses based on linkage disequilibrium. We identified 1,009 signals, 49% of which showed significant sex effect heterogeneity (FDR 5%) and 63% of which have been previously reported in GWAS for a non-anthropometric trait. We replicated an enrichment for signals near genes highly expressed in adipose and adipocytes, and also identified 30 additional enriched tissues (e.g. heart, skeletal, endocrine, and digestive systems). We used three approaches to link the signals to target genes: Polygenic Priority Score (PoPS), Data-driven Expression Prioritized Integration for Complex Traits (DEPICT), and colocalization with blood and adipose eQTL. The colocalization analyses included a subset of 659 WHR signals and were performed using two approaches: first, using only raw WHR summary statistics (marginal data), and second, after isolating conditionally distinct WHR signals. Analysis with PoPS identified 353 genes and with DEPICT identified 501 genes. Colocalization (coloc PPH4 >= 0.8) with GTEx subcutaneous adipose eQTL (n=491 samples) identified 79 and 121 genes that colocalized with marginal and distinct WHR signals, respectively, and with eQTLGen blood eQTL (n=31,684) identified 80 and 105 genes with marginal and distinct WHR signals, respectively. Colocalization with distinct GWAS signals identified up to 50% more genes than with marginal data, highlighting the value of analyzing signals separately. The larger number of genes identified using eQTL from adipose vs. blood despite a 64-fold smaller sample size may reflect the relevance of adipose to WHR. Together, these methods identified 901 unique genes. Fewer than 2% were prioritized by all three methods; while some of these 13 genes are known to impact central obesity (e.g. *TBX15* and *RSPO3*), others are at newly identified WHR loci but are associated with related traits. For example, *IGF2* is a well-known type 2 diabetes (T2D) locus that also plays a role in metabolic regulation and obesity, while *TMEM132C* may play a role in insulin secretion and T2D. In conclusion, the 1,009 WHR signals identified by our ongoing GWAS are enriched in a wide range of tissues, and the use of three gene prioritization methods identified 901 candidate genes that may influence the development of central obesity.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1634 Leveraging machine learning derived liver fat predictions from multiple data modalities in the UK Biobank for genetic discovery in NAFLD

Authors:

H. Somineni, S. Mukherjee, J. Pei, K. Guo, D. Light, S. Satapati, D. Lloyd, P. Philip, D. Koller, C. O'Dushlaine, T. Soare; Insitro, South San Francisco, CA

Proxy phenotypes, phenotypes that are associated with a disease of interest, where the association is suspected to be due to shared biology, allow for the utilization of genetic data from large population cohorts to gain a better understanding of the genetic architecture of complex diseases. Nonalcoholic fatty liver disease (NAFLD), liver with more than 5.5% fat content, is a leading risk factor for end-stage liver disease and is associated with a range of cardio-metabolic conditions. To better understand the genetic architecture of NAFLD, we used machine learning (ML) to predict liver fat content using three different data modalities measured in the UK Biobank: magnetic resonance imaging (MRI; n=25,474 participants), blood-based clinical and anthropometric markers (biomarkers; n=262,927), and plasma metabolites (n=82,138). Genome-wide association analyses identified over 800 independent signals, significantly expanding the catalog of likely genetic contributions to liver fat accumulation. Among these signals were well-known associations, including PNPLA3, TM6SF2 and GCKR, with consistent effect estimates confirming established biology. By comparing our results to publicly available summary statistics from clinically-defined NAFLD case-control status (clinical NAFLD, 49 signals) or MRI-derived proton density fat fraction (PDFF; 92 signals), we observed that the MRI modality was most informative (57% precision and 41% recall when compared to clinical NAFLD; 93% precision and 28% recall with PDFF). Plasma metabolites were least informative (8% precision and 20% recall with clinical NAFLD; 8% precision and 14% recall with PDFF). Lastly, liver fat content from biomarkers yielded a 76% recall value, but only had a 5% precision when compared to clinical NAFLD, and 58% recall and 7% precision when compared to PDFF. Genetic correlation analysis indicated a strong correlation between liver fat content from across modalities (ranging from 0.79 to 0.97) and with clinical NAFLD (ranging from 0.73 for metabolite-derived liver fat to 0.87 for MRI-derived liver fat). Overall, these findings demonstrate the value of leveraging ML-based trait predictions across orthogonal data sources to identify potentially novel genetic associations.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1635 Lipidomics profiling and plasma triglyceride concentration in African populations.

Authors:

A. Bentley, A. Doumatey, J. Zhou, K. Meeks, E. Heuston, C. Rotimi, A. Adeyemo; NIH, Bethesda, MD

Studies using metabolomic approaches may help elucidate the biochemical mechanisms underlying associations between serum triglyceride concentration (TG) and cardiometabolic traits (CMT), but such studies in diverse populations are lacking. To characterize the lipidomics profile associated with TG concentration, we conducted an untargeted metabolomics study of 308 Nigerian adults. Samples were assayed on the Metabolon© platform and the 480 metabolites of the Lipids Super Pathway were analyzed. Regression models were used to assess the association of metabolites with TG and high TG (TG>150 mg/dl [highTG]). Models were adjusted for age, sex, and type 2 diabetes status (T2D) and with and without adjustment for BMI or waist circumference. Statistical significance was set at FDR q<0.01. There were 99 metabolites associated with TG, with the strongest associations for diacylglycerols, such as palmitoyl-oleoyl-glycerol (β 0.67, q=5.4E-40). There were 56 significant associations with highTG, also led by diacylglycerols, including oleoyl-oleoyl-glycerol (β 0.66, q=1.4E-38). Overrepresentation analysis identified significant enrichment of diacylglycerols, monoacylglycerols, glycerophospholipids, and sphingophospholipids among the TG-associated metabolites. Most of the significant associations (91%) replicated in a cohort of 199 Kenyan adults. Next, we conducted analyses stratified by T2D, given known perturbations in TG with T2D. While most associations were similar across strata, there were 4 significant associations for fatty acyl carnitines among those without T2D that were not observed among those with T2D or in unstratified analyses. We subsequently assessed the association of lead TG metabolites with CMT with which TG is known to be associated. These metabolites were most strongly associated with LDL, total cholesterol, and measures of adiposity, but patterns of association differed across the metabolites, suggesting the TG-related pathways underlying these traits differ. We conducted mediation analyses for metabolites that were associated with CMT and found that 81% of the TG-CMT associations were partially or completely mediated by these metabolites. For example, after controlling for the mediating effect of N-stearoyl-sphinganine (a dihydroceramide), TG was no longer significantly associated with BMI or waist circumference. In summary, in this lipidomics profile study, variation in TG concentration is associated with both TG and non-TG lipids classes. Differing patterns of association between lead TG metabolites and CMT suggest different TG-related pathways may underlie known relationships between TG and CMT.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1636 Localization of polygenic signal in Alzheimer's disease through the integration of cell-type functional annotations and deep learning models.

Authors:

C. Lakhani[1], J-S. Lin[1], A. Das[2,1], T. Raj[3], D. Knowles[1,2]; [1]New York Genome Ctr., New York, NY, [2]Columbia Univ., New York, NY, [3]Icahn Sch. of Med. at Mount Sinai, New York, NY

The advent of increasingly well-powered Alzheimer's disease (AD) GWAS has made it clear that AD is a polygenic disease. Tools such as stratified LD score regression (*Finucane et al. 2015*) have enabled researchers to study the polygenic signal of disease by quantifying the contribution of functional annotations such as cell-type specific enhancers and promoter annotations to the heritability of AD. In another direction, deep learning methods have been increasingly used in genomics to train models which can predict functional genomics profiles as a function of genomic DNA sequence. We utilized a recent deep learning model, Enformer (*Avsec et al. 2021*), to predict variant transcription factor binding disruption (TF-VEP) scores for hundreds of TFs. However, a key challenge in AD genetics is that most models do not have TF-VEP scores for AD-relevant cell types (e.g. myeloid cells such as microglia and peripheral monocytes). In our study, we constructed functional annotations intersecting TF-VEP scores from ENCODE cell lines with active enhancer and promoter regions (*van de Geijn et al. 2020*) for 4 putatively relevant AD cell types (astrocytes, oligodendrocytes, microglia, and neurons) (*Nott et al. 2019*) and then quantified the extent to which these annotations contributed to AD polygenicity for four recent AD GWAS (*Kunkle (2019)*, *Jansen (2019)*, *Wightman (2021)*, and *Bellenguez (2022)*). Among these 4 cell types, we found microglia enhancers and promoters in open chromatin regions consistently captured the largest proportion of heritability (p_h^2) ($p\_h^2_{Kun}$ = 0.44, $p\_h^2_{Jan}$ = 0.77, $p\_h^2_{Wigh}$ = 0.18, $p\_h^2_{Bell}$ = 0.40). We also found further localizing these regions by integrating TF-VEP scores reduced the proportion of heritability only slightly but did significantly increase the percentage change in the enrichment of heritability (e_h^2) ($p\_h^2_{Kun}$ = 0.30, $e\_h^2_{Kun}$ = 39%, $p\_h^2_{Jan}$ = 0.67, $e\_h^2_{Jan}$ = 73%, $p\_h^2_{Wigh}$ = 0.15, $enrich\_h^2_{Wigh}$ = 64%, $prop\_h^2_{Bell}$ = 0.29, $enrich\_h^2_{Bell}$ = 94%). We also trained a novel transformer-based deep learning model on microglia chromatin accessibility data in order to capture variants with the potential to create new regulatory regions outside of those currently measured. We found the model had good performance on held-out validation data (Pearson correlation 0.45). Our analysis demonstrated how predictions from deep learning models not trained on AD-relevant cell types can be combined with relevant AD data to better localize the polygenic signal in AD GWAS. Such annotations can also be used as strong biological priors to better guide functionally-informed fine-mapping and rare variant analysis.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1637 Loci for Urinary Microbiota Abundance in a Multi-Ethnic Cohort of Children with Urinary Tract Infections and Vesicoureteral Reflux.

Authors:

M. Verbitsky, D. L. Bivona, P. Khosla, Y. Gupta, A. Khan, H. Park, K. Kiryluk, S. Sanna-Cherchi, J. Barasch, C. L. Mendelsohn, A-C. Uhlemann, A. G. Gharavi; Columbia Univ., New York, NY

Background: Gut microbiota and microbiome GWAS have been previously reported. However, host genetic factors associated with the human urinary microbiome composition, of special interest in recurrent urinary tract infections (UTIs) and genitourinary malformations, remain unknown.
Methods: We performed a microbiota GWAS (mGWAS) in a cohort of 278 children with urinary tract infections with or without vesicoureteral reflux (VUR), from the Randomized Intervention for Children with Vesicoureteral Reflux (RIVUR) and Careful Urinary Tract Infection Evaluation (CUTIE) studies, based on 16S rRNA gene sequencing from urine samples and whole-genome blood DNA genotyping. Given the limitation of our cohort sample size (N = 278), performing global ancestry-specific analyses is not feasible for the less represented and admixed groups and can eliminate the power to detect associations in the largest group (European ancestry, EUR), while analyzing the whole cohort can lead to spurious results, even when correcting for global ancestry. We, therefore, used a local ancestry-aware regression model implemented with Rfmix and Tractor software.
Results: Our initial mGWAS had identified a genome-wide significant association with the zero-truncated relative abundances of Pseudomonas on chr10 ($P = 2x10^{-9}$) and a suggestive association with Clostridia on Chr3 ($P = 5x10^{-8}$), in the whole cohort, correcting for global ancestry with 10 principal components, among other covariates. The top SNPs in these loci were on or near genes associated with immune surveillance, inflammation, and genitourinary tract development and disease (*CXCL12* and *ROBO1*). We further confirmed both these signals with local ancestry-aware mGWAS on the EUR background haplotype tracts extracted from the whole cohort ($P = 2.76x10^{-8}$ and $P = 2.06x10^{-7}$, respectively). In addition, phenome-wide association studies on the UK Biobank and eMERGE cohorts of the top SNPs showed suggestive associations with functional disorders of the bladder ($P = 4x10^{-3}$), urinary obstruction ($P = 2x10^{-3}$), and UTI ($P = 8x10^{-3}$).
Conclusion: We report the first combined urine microbiota analysis and mGWAS in children with UTI and VUR. We used a local ancestry-aware regression model to augment power and identified associations of bacterial taxa's relative abundances with human host genetic factors on the European ancestral background.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1638 Long-chain polyunsaturated fatty acids aberrations in bipolar disorder: Whole genome sequencing analysis

Authors:

E. Czeisler, S. P. Smieszek, R. Torres, P. Sweeney, C. Xiao, C. M. Polymeropoulos, G. Birznieks, M. Polymeropoulos; Vanda Pharmaceuticals Inc., Washington, DC

Long-chain polyunsaturated fatty acids (PUFAs) affect immune responses and inflammation and can be converted to potent inflammatory mediators. The omega (n)-3 and omega (n)-6 PUFA metabolic pathways participate in inflammatory processes and are linked to bipolar disorder (BD) pathophysiology. The omega (n)-3 PUFA helps treat inflammatory conditions, while arachidonic acid, an omega (n)-6 PUFA, causes inflammation. Prior GWAS analyses identified fatty acid desaturase polymorphisms conferring susceptibility to BD across several studied cohorts. In this study, we examined PUFA levels in BD patients in relation to whole genome sequencing (WGS) data. **S**equencing samples were obtained from a multicenter, randomized, double-blind, placebo-controlled, parallel-group study of iloperidone in patients with acute manic or mixed episodes associated with bipolar I disorder. WGS analysis consisted of 456 samples. Linear models on PUFA levels were adjusted for principal components (PCs), sex and age. PUFA panel was analyzed on a subset of 313 samples. We report mean elevated levels of (n)-6 PUFAs, arachidonic acid (mean value = 1487 nmol/mL) and linoleic acid (mean value = 925 nmol/mL). The total polyunsaturated and w6 were higher than the normal range. Vaccenic acid (mean value = 236 nmol/mL) and arachidic acid (mean value = 37 nmol/mL) were lower than the normal range. (n)-6 and (n)-3 PUFAs had a low correlation with BMI. The baseline PUFA levels led to further genetic analysis. We tested the association between variants in FADS genes and levels of (n)-6 PUFAs. We report rs174576 (intron variant; effect allele A, MAF 0.34) located on *FADS2*, which was also a previously reported top hit in the meta-analysis of the GWAS ($P = 1.34 \times 10^{-10}$, OR = 1.13), and rs174548. These variants were statistically and significantly associated with reduced D5D in a linear manner. A whole genome association analysis (linear model) was conducted on (n)-6 PUFAs correcting for age, sex, and PCs. Variants within CABLES2, rs11698818, MAF 0.09, ($P = 1.13 \times 10^{-8}$), exonic variant CABLES2: NM_031215: exon9: c.C1227G:p.G409G ($P = 1.13 \times 10^{-8}$), and variants within *KCNMA1* gene, which encode α-subunit of the large conductance calcium-sensitive potassium channel amongst multiple other loci, were reported. *FADS1/2* genes are expressed in the liver and catalyze the desaturation steps in synthesis of n-3 and n-6 PUFAs. Variants within are associated strongly with variation in blood PUFA levels, likely affecting the expression activities. We confirmed several existing loci and delineated potentially novel associations modifying the blood PUFA levels. These may inform of aberrant pathways driving the BD phenotype.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1639 Longitudinal analysis of 4,187 participants in the Atherosclerosis Risk in Communities Study reveals new insights into the determinants of incident clonal hematopoiesis

Authors:

M. Uddin[1], S. Saadatagah[2], A. Niroula[1], B. Yu[3], W. Hornsby[1], S. Ganesh[1], K. Lannery[1], A. Shuermans[1], M. Honigberg[4], A. Bick[5], B. L. Ebert[6], C. M. Ballantyne[2], P. Natarajan[4]; [1]Broad Inst., Cambridge, MA, [2]Baylor Coll. of Med., Houston, TX, [3]Univ. of Texas Hlth.Sci. Ctr. at Houston, Houston, TX, [4]Massachusetts Gen. Hosp., Boston, MA, [5]Vanderbilt, Nashville, TN, [6]Dana-Farber Cancer Inst., Boston, MA

Clonal hematopoiesis (CH), a phenomenon in which blood cells are predominantly derived from a single mutated hematopoietic stem cell, is associated with diverse aging-related diseases, including hematologic malignancy and atherosclerotic cardiovascular disease (ASCVD). Clonal hematopoiesis is common with advancing age, but the factors contributing to the development of CH among older adults are largely unknown. To identify environmental and genetic determinants of incident CH, we analyzed 8,374 whole-exome sequences from 4,187 participants without hematologic malignancy from the Atherosclerosis Risk in Communities Study with a median (range) follow-up of 21 (5-27) years between two sequencing time points. Around 59% of participants were female, 23% were African American, and the mean (range) age was 55.5 (45-83) years at baseline and 75.8 (67-90) years at follow-up visits. We identified 576 and 1,302 clones at variant allele fraction ≥2% in 457 and 1,047 participants at baseline and follow-up visits, respectively. During the follow-up, 735 (~20%) out of 3,730 participants without prevalent CH developed incident CH. Clonal diversity reduced with advancing age, with increased incident CH in *TET2*, *ASXL1*, *TP53*, and splicing factor genes (*SF3B1*, *U2AF1*, *SRSF2*, *ZRSR2*). Age at baseline was associated with a higher incidence of CH. However, ASCVD and other traditional risk factors for ASCVD, including smoking, body mass index, hypertension, type 2 diabetes, high-density lipoprotein cholesterol (HDL-C), and non-HDL-C, were not significantly associated with the incidence of CH. We showed that germline variants associated with prevalent CH, such as *SMC4*, *TERT*, *TCL1A*, and *SETBP1*, are also associated with the incidence of CH. Our comprehensive longitudinal assessment yields new insights into the factors promoting incident CH among older adults.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1640 Longitudinal gene expression changes associated with liver measures in Hispanic/Latino population at risk for metabolic-associated fatty liver disease

Authors:

R. Roshani[1], H-H. Chen[1], E. Frankel[1], H. Highland[2], E. Gamazon[3], W. Zhu[1], L. Petty[1], J. Below[4], K. North[2], J. McCormick[5], S. Fisher-Hoch[6], M. Graff[7]; [1]Vanderbilt Univ., Nashville, TN, [2]Univ North Carolina at Chapel Hill, Chapel Hill, NC, [3]VUMC Clare Hall, Univ. of Cambridge, Nashville, TN, [4]Vanderbilt Univ Med Ctr., Nashville, TN, [5]UTH SPH Brownsville, Brownsville, TX, [6]Univ. of Texas Sch. of Publ. Hlth., Brownsville, TX, [7]Univ North Carolina, Chapel Hill, NC

Metabolic-associated fatty liver disease (MAFLD) is a heterogeneous condition with variable severity and progression. Globally, MAFLD affects approximately 25% of adults, with the highest prevalence in Hispanic/Latino populations. Poorly controlled cardiometabolic risk factors (CMRF), genetic factors, lifestyle, and social determinants of health can influence MAFLD progression. However, longitudinal omics studies of MAFLD to understand mechanisms are limited. To interpret genome-wide association studies (GWAS) signals and identify biomarkers, it is crucial to investigate transcriptomics associated with liver disease progression, particularly in populations disproportionately affected by liver disease and underrepresented in biomedical research. This project aims to examine longitudinal gene expression changes related to liver measures in the Cameron County Hispanic Cohort (CCHC), an extensively phenotyped, randomly recruited cohort of Mexican Americans residing at the US-Mexico border with high rates of MAFLD. This study focused on a group of 927 CCHC participants with longitudinal transient elastography phenotyping (FibroScan) to measure liver stiffness, and controlled attenuation parameter (CAP) to measure hepatic fat. Clinical chemistries were used to derive biomarkers: fibrosis-4 index (FIB-4), aspartate aminotransferase/platelet index (APRI). We employed linear mixed-effect regression models to estimate the effects of liver measures on gene expression using up to 1294 measures, adjusting for covariates such as age, sex, and 10 Probabilistic Estimation of Expression Residuals (PEER) factors. We used a false discovery rate corrected p-value (FDR) to account for multiple testing. We identified 800 differentially expressed genes, including several genes within loci that have previously been implicated in GWAS of MAFLD and related traits. These genes include *DDX60L* ($FDR_{FIB\text{-}4}$ = 0.038), which has been implicated for alanine aminotransferase levels, *PARVB* ($FDR_{FIB\text{-}4}$ = 0.043), which has been implicated for measures of hepatic fat, alanine aminotransferase, and both adult and pediatric MAFLD, and *NDRG2* ($FDR_{FIB\text{-}4}$ = 0.028), which has been implicated for pediatric nonalcoholic steatohepatitis. Our results provide evidence of the functional effect driving those GWAS signals, including mapping the effector gene. We also performed sensitivity analyses stratified by the number of visits, to distinguish expression effects driven by cross-sectional and longitudinal effects, while controlling for BMI allowed identification of liver-associated expression effects independent of obesity measures.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1641 Long-read RNA-sequencing identifies novel protein coding transcripts of genes with implications for inherited and complex cardiac disease

Authors:

R. Dore[1], S. Garcia-Ruiz[1], E. Gustavsson[1,2], M. Ryten[1,2]; [1]Dept. of Genetics and Genomic Med. Res. & Teaching, UCL GOS Inst. of Child Hlth., London, United Kingdom, [2]NIHR Great Ormond Street Hosp. BioMed. Res. Ctr., Univ. Coll. London, London, United Kingdom

There has been considerable progress in the identification of genes causing inherited cardiac diseases, yet many individuals remain undiagnosed. In part, this could be explained by incomplete annotation of genes causally linked to these disorders leading to inaccurate variant interpretation. To test this hypothesis, we analysed long-read RNA-sequencing data originating from 16 individuals and nine cardiac regions (ENCODE4). We focused on 105 genes implicated in monogenic disease, as defined by Genomics England PanelApp. Among these genes, 67 met our inclusion criteria, requiring a minimum expression level of RPM≥20 in at least 80% of samples. Subsequent filtering for transcripts mapping to these 67 genes (≥ 2 full-length reads per sample) identified 245 unique transcripts of which 140 (57.1%) were putatively novel (Ensembl v105). Of these, 12 transcripts from six genes were predicted to be protein coding (*BAG3* = 1, *CASQ2* = 1, *DES* = 7, *MYH7* = 1, *TNNC1* = 1 and *SLC40A1 = 1*). These were generated through novel combinations of known splice junctions (n = 3), and through novel splice sites (n = 9). Notably, the novel *CASQ2* transcripts contain a novel coding exon detected in human mass spectrometry data from cardiac tissue. A common variant previously associated with left ventricular strain, left ventricular thickness and atrial fibrillation falls within this novel exon. Upon re-interpretation, this variant is now predicted to cause an amino acid substitution and an increased inclusion of this exon. Expression of the novel exon was also enriched in atrial fibrillation samples. Colocalisation analysis between *CASQ2* expression and atrial fibrillation implicated a lead intronic variant with the closest coding region being the novel exon. In fact, *CASQ2* is already a candidate target for arrhythmia treatments. Additionally, a novel exon in *DES*, also detected in human mass spectrometry data, was demonstrative of an Alu-exonisation event. Whilst this resulted in the loss of a key binding domain, the novel exon was found to be depleted in dilated cardiomyopathy samples. Despite the small sample set, this project demonstrates, through the examples of *CASQ2* and *DES*, that long-read transcriptomics improves the interpretation of pathogenic variants and provides key insights into complex human disease.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1642 Long-read sequencing show Japanese myotonic dystrophy type 2 patients carry a haplotype different from the European founder haplotype

Authors:

T. Matsuura[1], K. Tachikawa[2], T. Kimura[3], R. Kitao[4], S. Ura[5], Y. Sanpei[6], H. Funasaka[6], Y. Yamano[2], Y. Kochi[7], L. Ranum[8], K. Ohno[9], S. Mitsuhashi[2]; [1]Jichi Med. Univ., Shimotsuke, Japan, [2]St. Marianna Univ. Sch. of Med., Kawasaki, Japan, [3]Asahikawa Med. Ctr., Natl. Hosp. Organization, Asahikawa, Japan, [4]Natl. Hakone Hosp., Natl. Hosp. Organization, Odawara, Japan, [5]Japanese Red Cross Asahikawa Hosp., Asahikawa, Japan, [6]Akita Univ. Graduate Sch. of Med., Akita, Japan, [7]Med. Res. Inst., Tokyo Med. and Dental Univ., Tokyo, Japan, [8]Univ Florida, Gainesville, FL, [9]Nagoya Univ Grad Sch Med, Nagoya, Japan

Objective: Myotonic dystrophy type 2 (DM2) is a subtype of myotonic dystrophy caused by the expansion of a tetranucleotide CCTG repeat in intron 1 of the cellular nucleic acid-binding protein (*CNBP*) gene1. The aim of this study is to characterize the origin and features of expanded repeats in Japanese DM2 patients. **Methods:** Over the last 15 years, we have identified 7 DM2 patients from 5 unrelated families in Japan. We used Nanopore long-read sequencing (LRS) to investigate the following measures: (1) Repeat length of the expanded repeat, (2) sequence of the repeat tract, (3) adjacent DNA modifications, and (4) the phase of SNPs flanking the repeat. **Results:** LRS revealed that Japanese DM2 patients share a common haplotype, which differs from European DM2 patients. Simultaneous detection of the *CNBP* repeat region revealed that each read has different number of repeats, suggesting somatic mosaicism of the expanded repeat allele. The sequence within the repeat showed interruptions of the CAGA repeat, as recently reported in European DM2 patients. We also detected the CpG methylation in the native DNA molecules, which were not substantially altered in normal and expanded alleles, as previously reported. **Conclusions:** Japanese DM2 patients have a unique founder haplotype. Regardless of origin, the similar clinical features found in DM2 patients from Japan and other parts of the world are likely attributed to the pathogenic effects of the repeat expansion. The study demonstrates the usefulness of LRS for analyzing expanded repeats in a single sequencing experiment.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1643 Long-term cardiovascular risk according to genetic risk for hypertensive disease during pregnancy, lifestyle, and metabolic health

Authors:

S-H. JUNG[1], H. Kim[2], J-S. Yun[3], M. Shivakumar[1], Y. Jung[4], S. Kim[5], C-W. Park[2], J. Park[2], J. Jun[2], D. Kim[1,6], S. Lee[2,7]; [1]Univ. of Pennsylvania, Philadelphia, PA, [2]Seoul Natl. Univ. Coll. of Med., Seoul, Korea, Republic of, [3]St. Vincent's Hosp., Coll. of Med., The Catholic Univ. of Korea, Seoul, Korea, Republic of, [4]Guro Hosp., Coll. of Med., Korea Univ., Seoul, Korea, Republic of, [5]Univ. of Ulsan Coll. of Med., Asan Med. Ctr., Seoul, Korea, Republic of, [6]Inst. for BioMed. Informatics, Univ. of Pennsylvania, Philadelphia, Philadelphia, PA, [7]Innovative Med. Technology Res. Inst., Seoul Natl. Univ. Hosp., Seoul, Korea, Republic of

Recent studies have reported that genetic risk for hypertensive disorder during pregnancy (HDP) is associated with atherosclerotic cardiovascular disease (ASCVD) later in life. However, the impact of lifestyle and metabolic syndrome on the risk of ASCVD on genetic risk for HDP has not been evaluated despite these factors being modifiable risk factors. The aim of this study was to assess the long-term association between HDP, genetic risk, metabolic syndrome (MetS), lifestyle, and ASCVD. A total of 164,406 women of European descent from the UK Biobank were included. The genetic risk for HDP was determined using a genome-wide polygenic risk score (PRS) derived from a large-scale genome-wide association study (FinnGen Biobank; 13,071 HDP cases and 177,808 controls) using PRS-CS. The study population was divided into two generic risk groups according to PRS (Low risk, <20th percentile; High risk, >80th percentile). The incident ASCVD risk was evaluated according to genetic risk, lifestyle, and MetS. Lifestyle was categorized according to the presence of four modifiable lifestyle components (Favorable, ≥3 factors; Intermediate, two factors; Unfavorable, ≤1 factor) and MetS was categorized into two groups according to the number of metabolic components (Ideal [0-3 components] and poor [>=4 components]). Among the study subjects, 2,423 (1.5%) had a history of HDP, and 8,936 (5.4%) was diagnosed with incident ASCVD during follow-up. Individuals at high genetic risk for HDP had a 25.0% higher risk of ASCVD than those with low genetic risk (HR 1.250; 95% CI, 1.161-1.345; $P=2.89\times10^{-9}$). In both groups with low and high genetic risks for HDP, the risk of ASCVD was reduced by up to 48.2% (HR 0.518; 95% CI, 0.694-0.387; $P=1.07\times10^{-5}$) through ideal MetS and favorable lifestyle. Although a higher the genetic risk for HDP increases the overall risk of ASCVD, maintaining healthy lifestyles and MetS can reduce ASCVD risks regardless of genetic risk.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1644 † Looking Back on GWAS Using Hybrid Artificial Intelligence

Authors:

S. D. Crosby, C. E. Storer, R. A. Barve, R. Head; McDonnell Genome Inst., Washington Univ., St. Louis, MO

Multifactorial disorders (e.g. T2 diabetes, cardiovascular disease, etc) also known as complex traits or polygenic disorders, arise from a combination of genetic and environmental factors. A genome wide association study (GWAS) is employed to compare the genomes of individuals with a disorder to those without, to identify genetic variants that are more prevalent in affected individuals. These variants may be located in genes, or as is more often the case, in non-coding regions of the genome. GWAS have been limited not only by design and population size but also by analytical methods. Statistical cut-offs incompletely address the problems of false positives and negatives. Here we present a meta-analysis of 47 **autism** *GWASs (1321 gene associations) using* **CompBio***, a platform we created to facilitate parallel multi-omic analysis. CompBio is novel hybrid AI platform, which combines natural language processing, conditional probability analysis, and machine learning CompBio employs the entirety of PubMed abstracts and millions of full-text papers to create "knowledge maps" out of ALL unique terms found in the biomedical corpus. CompBio simultaneously evaluates all variant-associated genes in a given* input *list. This "holistic" analysis, as opposed to independent observations, allows CompBio to identify both signal and noise components of the input list with great accuracy. The* output *consists of closely related clusters of genes and concepts; these clusters are called* **Themes***. Themes transcend traditional pathways in that they represent not only molecular pathways, but also physiologic processes, cell types, diseases, etc. based on concept relationships. In the present example, we employed the NHGRI-EBI Catalog of Human Genome-wide Association Studies (https://www.ebi.ac.uk/gwas/) to identify "mapped genes" which either include or are nearby variants associated with autism. Of the 1165 mapped genes, 479 clustered within CompBio Knowledge Maps. 44% of the 50 most significant Themes identified were related to neurodevelopment, neurophysiology, or neuropathology. We were reassured to see CompBio validated, in that most previously identified physiologic processes appeared in our results. Interestingly, we also observed the emergence of novel themes in the output, suggesting undetected facets which may be related to better-understood aspects of this disease.*

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1645 Loss-of CFHR5 conveys protection from Age-related macular degeneration

Authors:

M. Reeve[1,2], S. Loomis[3], Y. Okunuki[3], J. Yang[3], C. Palmer[3], A. Khalifa[3], D. Ben-Isvy[1,2,4,5], J. Fu[1], C. Liao[1], X. Zhao[4], F. Finngen[2], M. Talkowski[4,1], J. Makela[6], M. Kurki[7,2], Z. Zheng[7], K. Connor[3], M. Daly[4,1,2], H. Runz[8]; [1]Broad Inst., Cambridge, MA, [2]Finnish Inst. of Molecular Med., Helsinki, Finland, [3]Biogen, Cambridge, MA, [4]Massachusetts Gen. Hosp., Boston, MA, [5]Harvard Med. Sch., Boston, MA, [6]Finnish Biobank Cooperative (FinBB), Turku, Finland, [7]Broad Inst. of MIT and Harvard, Boston, MA, [8]Biogen Inc., Cambridge, MA

Introduction: Age-related macular degeneration (AMD) is the leading cause of blindness in the elderly. Efficient treatments exist for wet AMD, but therapeutic options for the more common dry form remain limited. Genetic variants in the complement factor H (CFH) locus explain ~25% of dry and wet AMD risk, which has led to clinical trials testing whether increasing CFH might be therapeutic. In a recent meta-analysis of FinnGen and UKBiobank we identified a FIN-enriched frameshift variant within CFHR5 (1:196994128:C:CAA, rs565457964, CFHR5fs) associated with protection from AMD in a presumably CFH-independent manner. Here, we further explored the potential of CFHR5 as a drug target for AMD.
Methods: We performed detailed fine-mapping of the CFH/CFHR5 region and initiated a sample recall study from FinnGen on 200 dry AMD cases and 200 controls, enriched for carriers of CFHR5fs. Serum levels of CFH, CFHR1-5 and ~7000 other proteins were measured using the Somascan assay.
Results: Fine-mapping of the CFH/CFHR5 locus in FinnGen R10 identified two significant AMD associations that were conditionally independent from the two established CFH signals. Both associations could be traced to low-frequency variants that lead to a protein-coding change in CFHR5 and were associated with protection from AMD. Common structural variant maps were explored to test the impact of these variants on disease and proteomic signatures. Carriers of the previously described CFHR5fs had reduced levels of CFHR5 with each additional protective allele ($p<3x10-54$) consistent with complete loss-of-function in CFHR5fs homozygotes. CFHR5fs carriers also showed lower levels of CFHR2 ($p=7.0x10-33$) and CFHR4 (6.5x10-11), but not CFH, CFHR1 or CFHR3 ($p>0.05$).
Discussion: Through fine-mapping and functional analysis we identify two independent protective coding variants that reveal previously unknown complexities at the CFH/CFHR5 locus and establish reduction of CFHR5 as a promising therapeutic strategy for AMD. **Note:** M. P. Reeve and S. Loomis are shared first authors and M. Daly and H. Runz jointly supervising last authors.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1646 Low carnitine palmitoyltransferase 1 activity is associated with narcolepsy type 1 and other hypersomnia

Authors:

T. Miyagawa[1,2], Y. Shigematsu[3], M. Shimada[1,4,2], Y. Honda[1], K. Tokunaga[4,2], M. Honda[1,5]; [1]Tokyo Metropolitan Inst. of Med. Sci., Tokyo, Japan, [2]The Univ. of Tokyo, Tokyo, Japan, [3]Univ. of Fukui, Fukui, Japan, [4]Natl. Ctr. for Global Hlth.and Med., Tokyo, Japan, [5]Japan Somnology Ctr. and Seiwa Hosp., Tokyo, Japan

Narcolepsy type 1 (NT1), a rare hypersomnia of central origin characterized by excessive daytime sleepiness and cataplexy, is associated with metabolic abnormalities, but their etiology remains largely unknown. Palmitoyltransferase 1B (*CPT1B*) gene and abnormally low serum acylcarnitine levels have been linked to NT1. We measured levels of individual acylcarnitines and evaluated CPT1 activity in patients with NT1 and other hypersomnia to clarify the details of altered fatty acid metabolism. Blood samples from 57 NT1 patients, 51 other hypersomnia patients, and 61 healthy controls were analyzed. The levels of 25 major individual acylcarnitines were determined and the C0/(t[C16] + t[C18]) ratio was used as a CPT1 activity marker. We further performed a transcriptome analysis (RNA sequencing (RNA-seq) using independent blood samples from 73 NT1 patients and 63 healthy controls to examine the relevance of fatty acid metabolism. NT1-specific changes in CPT1 activity and in expression of related genes were investigated. CPT1 activity was lower in patients with NT1 ($p = 0.00064$) and other hypersomnia ($p = 0.0014$) than in controls. Regression analysis revealed that CPT1 activity was an independent risk factor for NT1 (OR = 1.68; $p = 0.0031$) and for other hypersomnia (OR = 1.64; $p = 0.0042$). There was a significant interaction between obesity (BMI <25, ≥ 25) and the genotypes of NT1 associated SNP rs5770917 in *CPT1B* such that nonobese NT1 patients without risk allele had better CPT1 activity ($p = 0.0089$). The RNA-seq showed that expression levels of carnitine-acylcarnitine translocase (*CACT*) and *CPT2* in carnitine shuttle were lower in NT1 ($p = 0.0000073$ and $p = 0.0046$, respectively). Gene set enrichment analysis (GSEA) and pathway analysis using the RNA-seq data also revealed that several pathways related to carnitine shuttle and fatty acid metabolism were significantly associated with NT1. These results indicated that abnormal fatty acid metabolism is involved in the pathophysiology of NT1 and other hypersomnia.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1647 Machine learning-based penetrance of clinical variants

Authors:

I. Forrest, G. Rocheleau, D. Jordan, B. Petrazzini, H. Vy, G. Nadkarni, J. Cho, R. Do; Icahn Sch. of Med. at Mount Sinai, New York, NY

Accurate assessment of disease risk of genetic variants, known as penetrance, is a significant challenge in clinical genetics. Integration of large-scale phenotype data, genetic sequence data, and artificial intelligence holds promise in quantifying variant penetrance. Here, we constructed machine learning models for 10 genetic conditions using 1,347,298 participants with electronic health records (EHRs) from a large health system, then applied the models to an independent set of 28,953 individuals with linked EHR and exome data. Resulting disease probability scores were used to evaluate machine learning-based penetrance (ML penetrance) for 1,648 rare variants across 31 disease-predisposition genes that were pathogenic (P), benign (B), uncertain (VUS), or novel loss-of-function (LoF). The models achieved mean area under the receiver-operating-characteristic curves of 0.85 and 0.84 in validation and holdout sets, respectively. ML penetrance was variable within each variant class but greatest for P variants (median [IQR], 0.52 [0.38]) and novel LoF variants (0.48 [0.36]), then VUS (0.46 [0.31]), and smallest for B variants (0.21 [0.081]). Rare variants (allele frequency<0.001) had a higher ML penetrance than more common variants (median [IQR], 0.46 [0.32] versus 0.29 [0.073]; $P=8.3\times10^{-48}$). Notably, ML penetrance was associated with disease-relevant clinical outcomes such as chronic kidney disease in individuals with polycystic kidney disease variants (OR=1.11 per 0.1 increase in ML penetrance, 95% CI 1.08-1.14; $P=2.9\times10^{-13}$) and myocardial infarction in individuals with familial hypercholesterolemia variants (OR=1.02 per 0.1 increase in ML penetrance, 95% CI 1.01-1.03; $P=2.4\times10^{-3}$). Compared to conventional case-versus-control phenotypes and penetrance estimates, ML penetrance provided more refined quantitative estimates of disease risk. Furthermore, ML penetrance aided VUS and novel LoF variant interpretation by delineating those with high versus low penetrance; individuals with highly penetrant variants exhibited worse clinical trajectories over time. For example, carriers of highly penetrant *KCNQ1* VUS had prolonged QTc and bradycardic heart rates on electrocardiogram versus carriers of B variant or weakly penetrant *KCNQ1* VUS. By leveraging ML and deep clinical phenotyping, we systematically evaluated variant penetrance for a wide range of genetic diseases, thereby streamlining variant prioritization, enhancing risk assessment, and augmenting genetic discovery. This study presents a blueprint for precision medicine approaches that employ ML to accurately quantify disease risk of clinical variants at scale.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1648 Maternal genetic nurture and risk of childhood neuropsychiatric disorders.

Authors:

B. Mahjani[1], A. Kepinska[1], M. Caballero[1], A. Batuure[1], L. Cohen[1], B. Devlin[2], A. Buxbaum Grice[1], C. Hultman[3], M. Janecka[1], L. Klei[4], H. Larsson[5], C. Mahjani[1], A. Reichenberg[1], S. Sandin[1], J. Buxbaum[1], D. Grice[1]; [1]Icahn Sch. of Med. at Mount Sinai, New York, NY, [2]Univ. of Pittsburgh, Pittsburgh, PA, [3]Karolinska Inst.t, Stockholm, Sweden, [4]WPIC/UPMC, Pittsburgh, PA, [5]Karolinska, institutet, Sweden

Studies have shown that maternal factors are associated with an elevated risk of neurodevelopmental disorders in children. Understanding the causes of these associations is complex; however, the association can arise from factors such as the direct impact of maternally transmitted genotypes on the offspring's phenotype and through maternal effect. In quantitative genetics, maternal effect encompasses the influences on the offspring phenotype that stem from the indirect effect of maternal genotypes, referred to as maternal genetic nurture (dynastic effect) and/or the maternal environment.

Using data from the Swedish national registries, we estimated direct additive genetic and maternal effect contributions to the liability of obsessive-compulsive disorder (OCD), Tourette syndrome (TS), and attention-deficit/hyperactivity disorder (ADHD). We also examined the role of assortative mating and specific maternal factors, including maternal age, maternal smoking during pregnancy, maternal psychiatric history, and gestational age.

For OCD, our findings revealed that direct additive genetics accounted for 35% of the total variance in risk, and maternal genetic nurture accounted for another 7%. Moreover, our results indicated an association between maternal age, maternal smoking during pregnancy, and the risk of OCD in children. We observed evidence for substantial assortative mating among individuals with OCD, meaning that individuals with OCD chose a partner with OCD more frequently than expected under a random mating pattern. We show that failing to account for maternal effect or assortative mating, if present, leads to overestimates of the impact of direct additive genetic effect. Regarding TS, we estimated 60.7% direct additive genetic effect, 4.8% maternal genetic nurture, and 0.5% environmental maternal effect. For ADHD, our preliminary results suggest 66.1% direct additive genetic effect and 14.3% maternal genetic nurture. We also observed evidence for substantial assortative mating among individuals with ADHD.

The risk of developing OCD, TS, and ADHD arises from a complex interplay of genetic factors (nature) and environmental influences (nurture). Our findings indicate that maternal genotypes play an important role in determining the risk of neurodevelopmental disorders in their children, not only through direct genetic transmission but also through environmental pathways.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1649 Matrix factorization analysis of cross-trait associations identifies platelet crit as a new causal or pleiotropic effect for primary open angle glaucoma

Authors:

A. Segre[1,2,3], J. Ohmura[4,5], K. Smith[6,3], C. Kim[6,3], J. Rämö[7,8], A. R. Hamel[1,3], W. Yan[9], Y. Zhao[10,2], A. Monavarfeshani[9], T. van Zyl[11,12], Q. Liang[13], FinnGen, A. Palotie[8], E. Rossin[5], T. Elze[10,2], N. Zebardast[10,2], R. Chen[14], J. R. Sanes[9], M. S. Udler[6,3], J. L. Wiggs[1,2,3]; [1]Ocular Genomics Inst. and Dept. of Ophthalmology, Massachusetts Eye and Ear, Boston, MA, [2]Dept. of Ophthalmology, Harvard Med. Sch., Boston, MA, [3]Broad Inst. of Harvard and MIT, Cambridge, MA, [4]Harvard-MIT Div. of Hlth.Sci. and Technology Program, Cambridge, MA, [5]Massachusetts Eye and Ear, Harvard Med. Sch., Boston, MA, [6]Ctr. for Genomic Res., Massachusetts Gen. Hosp., Boston, MA, [7]Broad Inst. of MIT and Harvard, Cambridge, MA, [8]Inst. for Molecular Med. Finland (FIMM), Helsinki Inst. of Life Sci. (HiLIFE), Univ. of Helsinki, Helsinki, Finland, [9]Harvard Univ., Cambridge, MA, [10]Dept. of Ophthalmology, Massachusetts Eye and Ear, Boston, MA, [11]Yale Sch. of Med., New Haven, CT, [12]Regeneron Pharmaceuticals, Tarrytown, NY, [13]Baylor Coll. of Med., Houston, TX, [14]Baylor Coll. of Med., Pearland, TX

Primary open angle glaucoma (POAG) is a heterogeneous disease whose underlying molecular mechanisms and pathophysiology are not well understood. While elevated intraocular pressure (IOP) is a major risk factor, 40% of glaucoma patients have normal IOP levels. Over 130 genomic loci have been associated with POAG from large European and cross-ancestry genome-wide association study (GWAS) meta-analyses. We hypothesize that subsets of POAG-associated variants will affect genes in different or overlapping biological processes associated with various sets of human traits involved in glaucoma pathogenicity. Thus, we applied Bayesian nonnegative matrix factorization (bNMF), a soft clustering method that emulates physiological relevance by allowing variants to belong to more than one pathway, to 133 independent POAG variants and their associations with traits relevant to POAG development. Endophenotypes were selected based on known or putative pathogenic mechanisms of POAG, e.g. vascular, neurodegenerative, metabolic, and mitochondrial, and having available GWAS summary statistics. A standardized association matrix of 133 POAG variants by 28 endophenotypes, accounting for trait direction of effect on POAG risk and normalized to GWAS sample size, was prepared. bNMF identified 4 clusters with 5 to 39 POAG variants each (converging to N=4, 83.8% of 1000 iterations). Each cluster was represented by unique trait and direction of effect combinations. The strongest effects on POAG risk resulted from increased IOP and decreased corneal hysteresis (cluster 3), and decreased platelet crit (PCT, platelet volume-%) and diastolic blood pressure (cluster 2). These results were replicated in an independent study (Mass General Brigham biobank) of up to 54,000 individuals, associating cluster partitioned polygenic risk scores with cluster-defining traits, correcting for age, sex, and top 10 genotype PCs (P<1E-3). PheWAS using the UK biobank and FinnGen study identified additional cluster-related traits (P<1E-08), such as reduced white blood cell counts for cluster 2. To gain cellular insight, genes mapped to POAG loci in each cluster via e/sQTL colocalization analysis were tested for enrichment in specific ocular cell types, highlighting ciliary and iris fibroblasts in the anterior segment for cluster 3 (IOP), and fibroblasts and vascular cells in the optic nerve head and retinal astrocytes for cluster 2 (PCT) (FDR<0.1). This work proposes new disease mechanisms associated with POAG risk, such as reduced PCT and blood pressure potentially affecting optic nerve blood flow, that may inform therapy development and personalized diagnosis and treatment.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1650 Medical and surgical complications in paediatric patients with achondroplasia followed in a specialty skeletal dysplasia centre.

Authors:

E. Martin, Y. Wang, L. Dupuis, M. Bouchard, A. Howard, **R. Mendoza-Londono**; The Hosp. for Sick Children, Toronto, ON, Canada

Achondroplasia is the most prevalent skeletal dysplasia. It is typically caused by a single gain of function missense variant in the fibroblast growth factor receptor 3 (*FGFR3*) gene. Increased signalling via FGFR3 leads to abnormal bone, joint, and cartilage development, causing short stature, skeletal dysplasia, and distinctive craniofacial features. In addition, patients with achondroplasia can experience medical and surgical complications requiring a multidisciplinary management approach. Novel treatments are being developed aimed at mitigating the effect of the activating variant; however, primary endpoints have focused primarily on stature and have not documented an impact on other complications. **Objective**: To determine the prevalence and management of the medical and surgical complications associated with achondroplasia in patients referred to The Hospital for Sick Children, Canada's largest children's hospital. **Methods**: We conducted a retrospective chart review and collected longitudinal data on complications, referrals for evaluation, and medical and surgical interventions in a cohort of 60 patients. **Results**: Breathing and airway issues were the most frequent complications. Sleep apnea affected 87% of patients. Fifty-three percent required tonsillectomy/adenoidectomy surgery, and 18% required continuous positive airway pressure (CPAP) treatment. Middle ear effusions requiring myringotomy tubes were documented in 48% of patients. Neurosurgical complications, including foramen magnum stenosis and cervicomedullary compression requiring surgery, were seen in 37% of patients. Orthopaedic concerns included genu varum treated by guided growth in 23% of patients. This study documents the frequency of complications in patients with achondroplasia and provides a benchmark to compare the efficacy of management strategies. Gaining a better understanding of the clinical presentation and prevalence of non-stature related complications in achondroplasia will allow for better monitoring of outcomes as new therapies are introduced. It will also guide counselling of new families regarding the prevalence of these complications and the most effective treatment strategies, ultimately leading to a better quality of life for individuals with achondroplasia.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1651 Mendelian randomization analyses identify causal associations between human gut microbiome composition and intelligence.

Authors:

S. Yao[1], J-Z. Han[2], Y. Guo[2], T-L. Yang[1,2]; [1]Natl. and Local Joint Engineering Res. Ctr. of Biodiagnosis and Biotherapy, The Second Affiliated Hosp., Xi'an Jiaotong Univ., Xi'an, China, [2]Key Lab. of BioMed. Information Engineering of Ministry of Ed., BioMed. Informatics & Genomics Ctr., Sch. of Life Sci. and Technology, Xi'an Jiaotong Univ., Xi'an, China

Intelligence is a robust predictor of educational and socioeconomic achievement and broadly implies lifestyle behaviors and health resource advantages across the lifespan. Growing evidence indicates that dynamic changes in the gut microbiome can affect intelligence; however, whether the relationships are causal is unknown. Randomized controlled trials (RCTs) have the potential to establish causal relationships. Nevertheless, most RCTs are expensive and time-consuming, and more importantly, gut microbiome composition cannot be randomly allocated in practice. To disentangle the poorly understood causal relationship, we performed a bidirectional two-sample Mendelian randomization analysis using genetic variants from the largest available genome-wide association studies of gut microbiota composition (N = 18,340) and intelligence (N = 269,867). We observed causal evidence indicating a risk effect of the *Oxalobacter* on intelligence (β = -0.032; 95% CI, -0.049 to -0.015; $P = 1.88\times10^{-4}$) and a protective effect of the *Fusicatenibacter* on intelligence (β = 0.051; 95% CI, 0.023 to 0.079; $P = 3.03\times10^{-4}$) after multiple testing corrections. Considering the close relationship between brain volume and human intelligence, we performed a two-step MR analysis and found that the effect of *Fusicatenibacter* on intelligence was partly mediated by regulating the brain volume, with a mediated proportion of 26.7%. In the other direction, with genetic liability for intelligence as exposure, we performed reverse MR analyses to explore the causal effect of intelligence on the gut microbiome abundance and did not find significant evidence. Our findings may help reshape our understanding of the microbiota-gut-brain axis and development of novel intervention approaches for preventing cognitive impairment.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1652 Mendelian randomization analyses suggest a causal role for circulating GIP levels in homeostatic model assessment (HOMA)-derived measures of beta-cell function and insulin sensitivity.

Authors:

K. Meeks[1], A. Bentley[1], T. Assimes[2], N. Franceschini[3], A. Adeyemo[1], C. Rotimi[1], A. Doumatey[1]; [1]NIH, Bethesda, MD, [2]Stanford Univ Sch Med., Palo Alto, CA, [3]Univ North Carolina at Chapel Hill, Chapel Hill, NC

Certain cytokines have been implicated in the development and progression of type 2 diabetes (T2D). However, studies on their role in T2D in humans are scarce. In a sub-Saharan African cohort, we evaluated associations between 11 circulating cytokines and T2D and tested for causal relationships using Mendelian randomization (MR) analyses. The participants were from Ghana, Nigeria, and Kenya, and included 2,276 individuals with T2D and 2,790 non-T2D individuals from the Africa America Diabetes Mellitus study. We fitted logistic regression models adjusted for age, sex, body mass index (BMI), and country of origin to regress levels of 11 cytokines (adipsin, leptin, visfatin, PAI-1, GIP, GLP-1, ghrelin, resistin, IL-6, IL-10, IL-1RA) on T2D. Linear regression models using the same covariates were fitted with homeostatic modelling assessments of insulin sensitivity (HOMA-S) and β-cell function (HOMA-B) as dependent variables among non-T2D individuals. Circulating GIP and IL-1RA levels were associated with T2D. Nine of the 11 cytokines (all except GLP-1 and IL-6) were associated with HOMA-S, HOMA-B, or both among non-T2D individuals. We subsequently used 35 genetic variants previously associated with at least one of these 11 cytokines among non-T2D individuals as instrumental variables in univariable and multivariable MR analyses. Statistical significance was set at 0.0045 (0.05/11 cytokines). Two-stage least squares MR analysis provided evidence for a causal effect of GIP on HOMA-S and HOMA-B both in univariable (HOMA-S β =-0.78; P-value=1.11×10^{-11} and HOMA-B β=0.67; P-value=1.47×10^{-9}) and multivariable analyses (HOMA-S β=-0.67; P-value=1.88×10^{-6} and HOMA-B β=0.59; P-value=1.88×10^{-5}). Inverse variance weighted MR analysis provided evidence for a causal effect of adipsin on T2D (multivariable OR=1.83, P-value=9.79×10^{-6}), though these associations were not robust in the MR-Egger and weighted median sensitivity analyses. Subsequent mediation analyses revealed that the effect of GIP on HOMA-S and HOMA-B was not mediated through BMI. Our findings indicate that circulating GIP levels are independently causal for reduced insulin sensitivity and increased β-cell function in non-T2D individuals, but not for T2D status. This is concordant with studies reporting impairment of GIP's glucose-dependent insulinotropic effect among T2D individuals. These observations suggest that circulating levels of GIP could be a promising early biomarker for T2D risk or a therapeutic target. Indeed, emerging clinical studies are exploring the therapeutic potential of GIP in combination with other gastrointestinal peptides such as GLP-1.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1653 Mendelian Randomization Analysis of the Effects of Alcohol Use on Cancer Risk

Authors:

H. Zhou[1], B-Z. Yang[2], H. Kranzler[3], J. Gelernter[4]; [1]Yale Univ., Orange, CT, [2]Yale Univ Sch Med, West Haven, CT, [3]Univ. of Pennsylvania, Philadelphia, PA, [4]Yale Univ Sch. of Med., West Haven, CT

Background/significance/hypothesis: Cumulative epidemiological evidence shows that an estimated 5.5% of all cancers are attributable to alcohol use, totaling 770,000 cases annually. Alcohol consumption is associated with the development of cancer of the oral, breast, liver, and colon. GWASs have identified risk genes for alcohol-related traits and cancers. However, the relationship between alcohol use and different kinds of cancer remains unclear. **Methods**: To provide adequate power to investigate the causal relations between two alcohol-related phenotypes and different cancers, we used the GWAS summary statistics for problematic alcohol use (PAU, Zhou et al., PMID: 36747741, N=903,147) and drinks per week (DPW, Liu et al., N=941,280). We used Mendelian Randomization (MR) and summary statistics from GWAS of several cancers, including breast, colorectal, esophageal and oral/pharyngeal cancers. Inverse-variance weighting (IVW), weighted median, MR-Egger, and MR-PRESSO were implemented for MR inference. **Results**: We observed nominally significant causal effects of PAU on esophageal cancer (IVW p=0.01) and significant effects (correcting for 8 tests) on breast (IVW $p=2.6\times10^{-3}$) and oral cancer risks (IVW $p=4.4\times10^{-6}$). We found no causal effects of DPW on cancers. **Conclusions**: There are few MR studies on this topic in the literature, mainly due to underpowered GWAS of alcohol-related traits and cancers. This study finds evidence of the causal effects of PAU on the risk of specific cancer types. These results are limited to European ancestry; further studies in underrepresented, understudied, and underreported populations are needed.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1654 Mendelian randomization in approximately 1 million stuttering cases and controls show significant causal effects of genes and illuminate causal relationships between hormonal, behavioral, and psychiatric traits and stuttering liability

Authors:

A. Scartozzi[1], D. Pruett[2], H. Polikowsky[1], D. Shaw[1], H-H. Chen[3], L. Petty[1], A. Petty[2], E. Lowther[4], Y. Yu[5], 2. Research Team[6], H. Highland[7], C. L. Avery[8], K. Mullan Harris[8], R. Gordon[2], J. Beilby[9], K. Z. Viljoen[4], R. Jones[1], C. Huff[10], S. Kraft[11], J. Below[12]; [1]Vanderbilt Univ., Nashville, TN, [2]Vanderbilt Univ. Med. Ctr., Nashville, TN, [3]Academia Sinica, Taipei, Taiwan, [4]Curtin Univ., Perth, Australia, [5]UT MD Anderson Cancer Ctr., Houston, TX, [6]23andMe, Inc., Sunnyville, CA, [7]Univ North Carolina at Chapel Hill, Chapel Hill, NC, [8]Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, [9]Curtin Univ., Perth, Austria, [10]U.T. MD Anderson Cancer Ctr, Pearland, TX, [11]Wanye State Univ., Detroit, MI, [12]Vanderbilt Univ Med Ctr., Nashville, TN

Developmental stuttering is a common speech disorder characterized by prolongations, blocks, and repetitions of speech sounds. Stuttering genetic risk factors are complex and involve both familial and population-level variation. Furthermore, stuttering is associated with a variety of comorbid traits. Despite progress in the field, the genetic factors affecting stuttering risk and comorbidity are unknown.

The current study uses 23andMe, Inc. genome-wide association study (GWAS) summary statistics of self-reported stuttering to better understand the causal pathways between stuttering and previously reported comorbid traits, and identify potential causal genes.

To determine if comorbid traits associated with stuttering harbor a causal or horizontal pleiotropic relationship, we performed sex-specific Mendelian randomization with 18 traits and stuttering. We found that the genetic risk of BMI, chronotype, walking pace, suicide ideation, and testosterone levels showed causal effects on stuttering liability, and the genetic risk of stuttering showed causal effects on depression ($p < .05$). These results are consistent with several studies suggesting that males and females who stutter report elevated symptoms of depression compared to their fluent counterparts.

To identify potential causal genes associated with stuttering, we performed sex-specific joint-tissue expression imputation followed by Mendelian randomization. For female stuttering, after multiple test correction, we identified 92 unique causal genes across all tissues. Only three of these genes, *NMUR2*, *MMAB*, *DCC*, were reported at genome-wide significant loci in the original 23andMe GWAS of stuttering (Polikowsky et al., 2023) and are involved in GPCR signaling, metabolic processes, and axon guidance. For male stuttering, we identified 24 causal genes across all tissues. Only *MMAB* was identified in Polikowsky et al. 2023, and ZMAT4 was a top hit in an independent analysis, Shaw et al. 2021, and is involved in nucleic acid binding.

Future studies will examine neuroendophenotypes associated with these potentially causal stuttering genes. Understanding the genetic relationships between stuttering and associated traits, as well as potential causal genes may inform clinical manifestations of stuttering that differ between sexes.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1655 Mendelian randomization study of body mass index and all-cause and vascular-metabolic mortality in 125,000 Mexican adults

Authors:

L. Gnatiuc Friedrichs[1], J. Alegre-Díaz[2], E. Trichia[1], R. Ramirez-Reyes[2], D. Aguilar-Ramirez[3], R. Clarke[1], N. Staplin[1], J. Torres[3], R. Wade[1], F. Rivas[2], M. Hill[4], R. Collins[5], J. Emberson[1], J. Berumen[6], P. Kuri-Morales[7], R. Tapia-Conyer[2]; [1]NDPH and MRC PHRU, Univ. of Oxford, Oxford, United Kingdom, [2]Natl. Autonomous Univ. of Mexico, Mexico City, Mexico, [3]NDPH, Univ. of Oxford, Oxford, United Kingdom, [4]CTSU and MRC PHRU, Univ. of Oxford, Oxford, United Kingdom, [5]Univ. of Oxford, Oxford, United Kingdom, [6]Univ. Natl. Autónoma de México, México, Mexico, [7]Inst. Tecnológico y de Estudios Superiores de Monterrey, Monterrey, Mexico

Background: Mendelian randomization (MR) studies are widely used to estimate the causal relevance of adiposity for vascular-metabolic diseases but most previous studies have been conducted in European ancestry populations. Using data from the Mexico City Prospective Study (MCPS), we used MR to estimate the causal relevance of body mass index (BMI) for all-cause and vascular-metabolic mortality in an admixed Mexican population.
Methods: 124,880 MCPS participants with genetic data and aged 35-74 years at recruitment were included. 767 single nucleotide polymorphisms associated with BMI were selected from the GIANT Consortia and, together with their weights, used to create an instrumental variable for BMI. The one-sample ratio method was used to estimate the causal effect of BMI on all-cause and vascular-metabolic mortality at ages 35-74 years, with Cox regression used to estimate sex-specific hazard ratios (HR) per 5 kg/m^2 higher BMI after adjustment for age-at-risk and genetic ancestry. Other MR approaches were used in sensitivity analyses including use of non-linear MR to examine linearity.
Results: The genetic instrument explained 3% of the variation in baseline BMI in men (F-statistic 130) and 5% of the variation in women (F-statistic 420). Over a median of 19 years follow-up there were 11,837 deaths at ages 35-74 years including 6634 vascular-metabolic deaths (3221 vascular, 1908 renal, 1006 hepatobiliary and 499 acute diabetic crises). Each 5 kg/m^2 higher genetically-predicted BMI was associated with a doubling in all-cause mortality (HR 2.05, 95% CI 1.84-2.29) and more than a doubling in vascular-metabolic mortality (HR 2.40, 2.07-2.78). For the vascular-metabolic deaths, each 5 kg/m^2 higher genetically-predicted BMI was associated with a doubling in vascular death (HR 1.94, 1.58-2.39) but more than a quadrupling in renal death (HR 4.04, 3.08-5.31) and death from an acute diabetic crisis (HR 5.69, 3.37-9.61). For deaths from hepatobiliary disease the HR was 1.42 (0.96-2.08). Hazard ratio estimates were generally larger in younger than in older people and among those with a higher proportion of Indigenous ancestry, but were similar in men and women. Much of the effect of BMI on mortality was explained by its effect on diabetes and the remaining (ie, residual) effects of BMI on mortality were comparable in individuals with and without diabetes. Non-linear MR analyses showed no departure from linearity and the results were consistent using alternative MR approaches.
Conclusion: These findings indicate that the causal relevance of higher levels of adiposity for death from cardiometabolic and other causes is particularly strong in Mexican adults.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1656 Metabolome-wide association study of obesity in a high-risk pediatric cohort

Authors:

A. Apsley, K. Passero, J. Noll, I. Shalev, M. Hall; The Pennsylvania State Univ., University Park, PA

Childhood obesity is a nationwide problem and is a risk factor for many later-life cardiometabolic disorders such as type-2 diabetes, hypertension and dyslipidemia. While it is well-established that the experience of child maltreatment (CM) is a risk factor for obesity, it is unclear whether the biological mechanisms underlying risk for obesity are unique for individuals who have experienced CM. The human urine metabolome, the complete collection of small molecules found in human urine, contains a plethora of information about an individual's physiological state and is a prime biological domain for investigating the molecular mechanisms of childhood obesity in this high-risk population. Using data from a pediatric cohort from the Pennsylvania State University's Child Health Study (CHS; N=408; N_{CM}=347; $N_{Comparison}$=61; 51.5% Female; Age=11.4±1.4; BMI=21.6±5.8), we sought to determine the metabolomic signatures of childhood obesity and to uncover whether these signatures are different for obese children having experienced CM vs. comparison youth. We used liquid chromatography coupled with mass spectrometry (LC-MS) to identify and quantify an average of 165±8.2 metabolites per sample. Samples were spiked with appropriate internal standards (e.g., stable isotope-labeled amino acids/nucleosides) that were later used for correcting for batch effects and for normalizing instrument responses. We used a metabolome-wide approach to detect individual metabolites associated with BMI and found nine metabolites were associated with BMI with $p < 0.05$. Three of these metabolites also had nominally significant CM and BMI interaction terms (Vanillylmandelic acid, 5-Hydroxylysine and L-Carnitine), indicating potential modifying mechanisms of CM on these metabolome-BMI associations. Further analysis using weighted gene co-expression network analysis (WGCNA) framework identified two groups of metabolites that were significantly correlated with BMI. Using a metabolite pathway analysis on the WGCNA identified groups, we found an overrepresentation of metabolites in the "Aminoacyl-tRNA biosynthesis pathway" (adjusted p-value=5.05E-08), which was enriched in individuals with higher BMI. These results demonstrate the value of evaluating the urine metabolome to uncover the biological mechanisms of obesity in a high-risk population.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1657 Metabolomic signatures of cognitive impairment in the Canadian Longitudinal Study on Aging

Authors:

S. Selber-Hnatiw[1,2,3], K. Trajanoska[1,2,3], V. Mooser[1,2,3], S. Zhou[1,2,3]; [1]McGill Univ., Montreal, QC, Canada, [2]McGill Genome Ctr., Montreal, QC, Canada, [3]McGill Canada Excellence Res. Chair in Genomic Med., Montreal, QC, Canada

Background and Objectives:
Dementia and Alzheimer's disease (AD) are cognitive disorders that represent major disparities in healthcare as there are no effective interventions nor robustly established biomarkers for disease detection. As such, it is important to focus on the earliest stages of disease onset, such as identifying modifiable circulating molecules (e.g., metabolites) that can be probed for controlling AD development, maintaining cognitive function, and mitigating early symptoms of dementia.
Methods:
We used data from 10,000 Canadian Longitudinal Study on Aging participants with 1,458 serum metabolite measurements with scores from in-depth cognitive tests to identify metabolites with large effects on cognitive function. We constructed three cognitive composite scores using cognitive tests related to general cognitive function, memory (i.e., verbal episodic memory and delayed recall), and executive function (i.e., executive and set switching, and verbal fluency). Log transformed, normalized metabolite data was assembled, and values with SD > 5 and > 50% missingness were removed, totalling 1,093 metabolites. Linear regression analyses were conducted to examine the association of the post-QC metabolites with each composite score, accounting for sex, age, and hours since consuming food or drink.
Results:
We identified 200 metabolites associated with at least one of the three composite scores following Bonferroni correction ($p < 1.52 \times 10^{-5}$). Several findings were supported in the literature as protection from cognitive decline and AD or biomarkers of early AD pathology, including carotene diol (beta = 0.34, $p = 3.56 \times 10^{-37}$, CI = 0.283-0.399), beta-cryptoxanthin (beta = 0.316, $p = 7.74 \times 10^{-27}$, CI = 0.259-0.374), and long, branched-chain dicarboxylic acids (beta = 0.248, $p = 3.88 \times 10^{-17}$, CI = 0.190-0.306). Other metabolites of interest included sphingomyelins (beta = 0.143, $p = 2.863 \times 10^{-06}$, CI = 0.083-0.203), which are critical in myelination and cognitive development, and 4-hydroxyphenylacetate and lactosyl-N-palmitoyl-sphingosine (d18:1/16:0), which appeared to be solely associated with executive function, suggesting the importance of focusing on a specific biological pathway of cognitive function.
Conclusions:
In this study, we have identified several metabolites associated with cognitive function and AD biology. The findings of this work may contribute to early AD and dementia management and understanding new pathophysiological mechanisms. Future work will focus on validating these metabolites by estimating their causal effects on cognitive decline and combining them with proteomic signatures.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1658 Meta-Prediction Reveals Polygenic Scores are the Principal Features Distinguishing Prospective Risk Reduction Profiles in Coronary Artery Disease

Authors:

S-F. Chen[1], A. Torkamani[2], H. Javedani Sadaei[1], S. Lee[1]; [1]Scripps Res., La Jolla, CA, [2]Scripps Translational Sci. Inst., La Jolla, CA

Coronary artery disease (CAD) remains the leading cause of mortality and morbidity worldwide. Recent advancements in large-scale genome-wide association studies have sparked renewed interest in polygenic risk scores (PRSs), highlighting the potential of genetic risk in clinical prevention and therapeutic selection. However, the current clinical utility of PRS models is limited to simply identifying high-risk populations based solely on the top percentiles of genetic likelihood. While some studies have attempted integrative prediction using genetic and non-genetic factors, many of these studies have been cross-sectional and lacked actionable recommendations. Our primary objective in this study was to address the limitations of current approaches by integrating genetic risk factors in a prospective prediction framework. We achieved this through a machine learning (ML) model capable of identifying interactions between risk factors and interventions, facilitating the identification of personalized risk reduction strategies.Thus, we present an omnigenic meta-prediction framework, employing an explainable ML approach, which effectively captures CAD risk subgroups, primarily distinguished by degree of genetic risk, with differing benefits from standard interventions. Our initial model development considered over 4000 predictive features, including demographic data, lifestyle factors, physical measurements, laboratory tests, medication usage, diagnoses, and PRSs. To drive our meta-prediction approach, we stratified the UK Biobank into two primary cohorts: 1) a prevalent CAD cohort which was employed to develop baseline predictive models for contributing risk factors and diagnoses, and 2) an incident CAD cohort used to train the final CAD incident risk prediction model. The final model included 200 predictive features, including meta-features derived from models trained on the prevalent risk cohort, that significantly enhanced 10-year incident CAD risk prediction. The final model achieved an AUC of 0.81 or F1-score of 0.67, outperforming prior models and able to identify interventions that provided the most benefit at an individual level.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1659 Microbic Risk Score to mapping metabolic networks in childhood obesity.

Authors:

N. Vergara Lope Gracia, M. Pedone; Univ. of Florence, Florence, Italy

Childhood obesity is a major global public health problem. In 2023, WHO estimated that one in three primary school-aged children is living with obesity or overweight**.** Obesity is a chronic metabolic disease influenced by the interaction of genetic and modifiable factors such as gut microbiota (GM). In a healthy status, the GM participates in the regulation of multiple physiological functions of the host. Herein, gut dysbiosis is unfavourable for the host and putatively participates in the pathogenesis of obesity. In this study, we propose a mathematical method to compute the individual's genome content based on the information theory to quantify the risk of a complex disease and offer a trail to predicting future metabolic paths associated with childhood obesity. Our analysis yields 48 Mexican children aged 6-10 years categorised as normal-weight, overweight and obese. The children's information includes biochemical measurements and bacterial DNA from human stool samples. We compute (1) the Microbic Risk Score (MRS) by summing the trait-associated alleles across genetic loci nurturing the genetic information content of bacterial signatures between individuals. Subsequently, we employ (2) a Bayesian Gaussian graphical model with covariates (GGMx) to estimate a metabolic network among established biochemical risk factors. The GGMx formulation allows both the strength and sparsity pattern of the precision matrix (controlling the network pattern) to change with the MRS, resulting in population-level and subject-specific graphs. The findings will indicate that the bacterial genomic information may help to evaluate potential variations in global genomic architecture between overweight/obese and normal-weight children. According to the estimated network, our goal is to provide that the genetic profile from MRS will elucidate a major implication to be associated with metabolic pathways in child populations. These results may suggest that the genetic and non-genetic factors are involved to disrupt energy homeostasis, increasing lipid synthesis/storage or dietary behaviour. Overall, our study will enhance our understanding of the role of GM in the pathogenesis of obesity and highlights the necessity for nuanced study designs in the association between biochemical patterns and genetic information.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1660 Microglial genes show significant difference in transcriptional burst kinetics between minor cognitive impairment and Alzheimer's disease cases

Authors:

B. Jin[1], W. Bush[2]; [1]Case Western Reserve Univ., Cleveland, OH, [2]Case Western Reserve Univ, Cleveland, OH

Microglial function is highly associated with neurodegenerative diseases. As innate immune cells in the central nervous system, microglia play critical roles in maintaining brain homeostasis. However, overactive microglia potentially cause synaptic loss, which is one of the pathologies of Alzheimer's disease (AD). Single-cell RNA-seq in microglia from AD mice has uncovered AD-associated microglia subtypes; however, the gene expression patterns associated with these subtypes are not replicated in human microglia.A reason is that human microglia have highly variable gene expression profiles. Such variability is not necessarily from differences in microglia sub-cell types. Instead, human genes are highly bursting in nature and exhibit large heterogeneity between alleles and within a population of cells. Moreover, increased transcriptional heterogeneity in humans is associated with aging, which is one of the biggest risk factors for AD.We hypothesize that transcriptional bursts are different for critical genes in AD cases compared to mild cognitive impairment individuals (MCI). We use a statistical model to quantify the heterogeneity of human microglia due to bursting gene expression. In detail, we used single-cell RNA-seq on microglia with 8 AD cases and 4 MCI cases from the ROSMAP project to estimate single chromosome transcriptional burst kinetics (burst size, and burst frequency) transcriptome-wide. We identified two genes, C1QC and ACTB, with significant differences in burst size between AD and MCI. Both C1QC and ACTB show larger burst sizes in MCI than in AD cases (p-value for C1QC: 0.00504; p-value for ACTB: 0.00511; p-value for candidate gene SPP1: 0.0359). Consequently, C1QC and ACTB show a higher expression level variability among microglia cells in MCI cases than AD. We found that even at the bulk level, such cell-cell variability can contribute to inter-sample gene expression variability. With an independent pseudo bulk RNA-seq data from single nucleus RNA-seq on 25 samples (16 AD, 9 Controls), the inter-sample variability in AD is significantly different for ACTB (AD:1.513; Control: 26.38) and SPP1 (AD:0.341, Control: 6.553). C1Q is associated with synaptic and neuron loss by microglial phagocytosis, whose expression increases with aging, especially in the hippocampus. Both highly variable expressions of C1QC and ACTB in MCI than AD indicate a globally elevated microglia state, increased mobility of microglia in MCI, and promoted microglial phagocytosis. The active microglia physiological states undergo potential functional change with AD progression and are prime targets for further experimental validation.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1661 Minimal evidence of statistical epistasis between height-associated SNPs

Authors:

A. Eliasen[1,2,3,4], L. Yengo[5], S. Vedantam[1,2], E. Marouli[6], E. Bartell[1,2], G. Lettre[7,8], Y. Okada[9], A. Wood[10], P. Visscher[5], J. Hirschhorn[1,2], GIANT Consortium height working group; [1]Dept. of Pediatrics, Div. of Endocrinology, Boston Children's Hosp. and Harvard Med. Sch., Boston, MA, [2]Med. and Population Genetics, Broad Inst. of MIT and Harvard, Cambridge, MA, [3]COPSAC, Copenhagen Prospective Studies on Asthma in Childhood, Herlev and Gentofte Hosp., Univ. of Copenhagen, Copenhagen, Denmark, [4]Deparment of Bio and Hlth.Informatics, Technical Univ. of Denmark, Kgs. Lyngby, Denmark, [5]Inst. for Molecular BioSci., The Univ. of Queensland, Brisbane, Australia, [6]Barts and The London Sch. of Med. and Dentistry Queen Mary Univ. of London, London, United Kingdom, [7]Université de Montréal, Montreal, QC, Canada, [8]Montreal Heart Inst., Montreal, QC, Canada, [9]Osaka Univ. / The Univ. Tokyo / RIKEN, Osaka, Japan, [10]Coll. of Med. and Hlth., Exeter Univ., Exeter, United Kingdom

Adult height is the classic model polygenic trait. It is highly heritable, with an estimated broad-sense heritability of 80%. Recently, a GWAS by the GIANT consortium of height in 5.4 million individuals identified 12,111 independent SNPs, which explain >90% of the SNP-based heritability for height due to common variants. Similar to other polygenic traits, height exhibits a difference between the broad and narrow-sense heritability. There are several potential explanations for this discrepancy, one of which is non-additive interactions between SNPs, a form of epistasis. Few examples of epistasis have been discovered and verified for human traits. The paucity of detected epistasis signals has largely been attributed to the statistical power required to detect these interactions, which is limited in human studies, and the possibility of more complex epistasis (beyond simple pairwise interactions).
We leveraged the results from the largest GWAS of height to date to quantify the contribution of certain types of genetic interactions to the genetic architecture of height. We first tested for interaction effects between all pairwise combinations of the recently identified 12,111 height-associated SNPs. There was no evidence of interaction between any of the variants. Additionally, we observed 3 interactions between sex and the height-associated SNPs at variants near *NKX2-1*, *KIAA0319*, and *CYP19A1*.
To look for more global evidence of interactions, we constructed a PRS of height and compared PRS-predicted height versus inverse-normalized height in the UK Biobank. We found significant deviations for individuals in the upper tail of the PRS (p-value = 2.21e-04), where individuals were taller than expected, suggesting the possibility of non-additive interactions at the taller extremes.
Lastly, we created gene-set-specific height PRS to search for evidence of interactions within and between gene sets (gene sets defined using DEPICT; SNPs assigned to genes based on proximity). To mitigate inflation due to cis-acting effects, we generated an odd and even PRS for each gene set based on chromosomes. Overall, no significant within or between pathway interactions were found. There was suggestive enrichment of between and within gene set interactions based on individuals with extreme height PRS. In summary, we found no statistical evidence of SNP-by-SNP interactions between height-associated SNPs. Guided by the recent GIANT GWAS results, we found 3 examples of height-associated interactions between genetic variants and sex. There was no overall evidence of genetic interactions between or within gene sets.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1662 Minor genetic overlap between rheumatoid arthritis and myocardial infarction.

Authors:

A. Öberg Sysojev[1], L. Alfredsson[1], L. Klareskog[1], The SRQ Biobank Group, G. N. Silberberg[1], S. Saevarsdottir[1,2], L. Padyukov[1], P. Magnusson[1], J. Askling[1], H. Westerlind[1]; [1]Karolinska Inst.t, Stockholm, Sweden, [2]deCODE Genetics Inc, Reykjavik, Iceland

Background: Patients with rheumatoid arthritis (RA), a chronic inflammatory disease affecting 0.5-1% of the Western population, are at increased risk of cardiovascular disease (CVD), including myocardial infarction (MI). This increased risk can be partly attributed to an increased prevalence of traditional cardiovascular risk factors, partly to the chronic and systemic inflammation of the disease and partly to interplay with various treatments. However, we and others have recently implicated familial factors as potentially contributory, indicating that also underlying genetic effects may predispose individuals to both RA and CVD. Nevertheless, targeted attempts at quantifying such shared genetic predisposition has yet to be undertaken.
Objectives: We attempted to quantify the genetic overlap between RA and MI by estimating the genetic correlation between the two traits.
Methods: We conducted a genome-wide association study (GWAS) in 26,637 Swedish RA cases and controls taken from the Epidemiological Investigation of RA study (EIRA), the Swedish Rheumatology Quality register biobank (SRQb) and the Swedish Twin Register (STR). The resulting RA GWAS data was paired against publicly available GWAS data on MI. Estimation of genome-wide genetic correlation was done through linkage disequilibrium score regression (LDSC) while estimates of local genetic correlation was obtained via local analysis of variant associations (LAVA). Furthermore, to assess the genetic overlap between RA and various traditional CVD risk factor traits, we estimated the genome-wide genetic correlation between RA and body adiposity, LDL-cholesterol, blood pressure, CRP-levels as well as smoking traits, using publicly available GWAS data.
Results: After quality control, our RA GWAS contained 25,826 individuals and ~5 million SNPs of which ~4.1 million overlapped the reference panel and the two GWASs. Genome-wide genetic correlation was estimated to rg = 0.13 (95%CI -0.03 - 0.29). Of 2495 distinct genomic regions tested for a genetic overlap, six showed significant local genetic correlation after correction for multiple testing, of which none were within the major genetic risk loci for RA, the MHC region on chromosome six. In analyses contrasting RA with CVD risk factor traits, no significant correlations were observed in any pairing, with the highest absolute genetic correlation being between RA and CRP (rg = 0.13, 95%CI -0.04 - 0.29).
Conclusions: Our findings indicate that any genetic overlap between RA and MI is minor. Furthermore, genetic overlap between RA and CVD risk factor traits seems unlikely to provide a major contribution to the increased risk of CVD observed in RA.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1663 miRNA & mRNA gene networks in Major Depression in a large postmortem brain sample

Authors:

Z. Taylor[1], J. Drake[1], A. Denham[1], S-A. Bacanu[2], J. Shin[3], J. Kleinman[2], T. Hyde[3], V. I. Vladimirov[1]; [1]Univ. of Arizona, Phoenix, AZ, [2]Virginia Inst. for Psychiatric and Behavioral Sci., Virginia Commonwealth Univ., Richmond, VA, [3]Lieber Inst. for Brain Dev., Johns Hopkins Univ., Baltimore, MD

Major Depression (MDD) is a disorder characterized by low mood and anhedonia that affects roughly one out of every six adults worldwide. While GWAS studies have identified genome-wide significant variants associated with MD much work remains in elucidating the functional impact of such variants. Complimentary research into MDD's underlying neuropathology involves identifying transcriptome changes associated with major depression in postmortem brain tissues. MicroRNA (miRNA), a class of small non-coding RNA with gene regulatory functions and high expression in the brain, has been studied in relation to neuropsychiatric phenotypes. Here, we used miRNA-Seq to compare the expression of over 1000 miRNAs between 150 MDD patients and 150 matched controls in the subgenual anterior cingulate cortex (sACC) and Amygdala. We performed differential expression analysis (DEA) of individual miRNAs, weighted gene co-expression analysis (WGCNA) of miRNA and mRNA expression, and miRNA/mRNA correlation-based analyses to identify miRNAs with a converging role in the etiology of MDD. We identified differentially expressed miRNAs (at FDR of 5%) including several previously associated with MDD and/or associated with neurodevelopment or other psychiatric illness. Our network analyses detected both significant miRNA and mRNA modules associated with MDD at the Bonferroni corrected $p \leq 0.05$. Using miRNA target site and gene enrichment analyses further identified gene ontologies that may constitute regulatory targets of the differentially expressed miRNAs. Ours is the largest to date postmortem brain miRNA expression study of major depression, and our ongoing analyses provides solid evidence of the importance of miRNA as a contributing factor to the development of MDD.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1664 Mitochondrial DNA heteroplasmy and incident cardiovascular disease

Authors:

Y. Hong[1], W. Shi[1], D. Puiu[2], N. Pankratz[3], N. Lake[4], M. Lek[4], E. Guallar[5], D. Arking[1]; [1]Johns Hopkins Univ Sch. of Med., Baltimore, MD, [2]Johns Hopkins Univ. Dept. of BioMed. Engineering, Baltimore, MD, [3]Univ. of Minnesota, Minneapolis, MN, [4]Yale Univ. Sch. of Med., New Haven, CT, [5]Johns Hopkins Univ Sch. of Publ. Hlth., Baltimore, MD

Background: Variants of mitochondrial DNA (mtDNA) may exist in heteroplasmy and have been associated with increased risk of mortality and cancer. Previous studies suggest an increased risk of cardiovascular disease (CVD) with mitochondrial dysfunction, represented as low mtDNA copy number (mtDNA-CN). Better understanding of the mtDNA heteroplasmy on CVD will provide insight into the molecular physiology of disease development and progression.**Methods:** We used the MitoHPC pipeline to quantify mtDNA heteroplasmy and mtDNA-CN from whole genome sequencing data in 194,871 participants in the UK Biobank. We defined a heteroplasmy allele frequency of 5% as heteroplasmic. We functionally characterized mtDNA single nucleotide variants using a novel constraint-based score, mitochondrial local constraint score sum (MSS). We evaluated the associations of MSS, mtDNA-CN, and incident CVD, defined as fatal or non-fatal myocardial infarction (MI) or ischemic stroke, using Cox proportional hazards models stratified by center and adjusted for age, sex, smoking, body mass index, blood pressure, blood lipid and glucose levels, and history of hypertension and diabetes in 191,204 participants without prevalent CVD. For mtDNA-CN, we further adjusted for neutrophil and platelet counts. We performed additional analysis stratified by MSS for each complex/region in the mtDNA. Lastly, we performed analysis separately for MI and ischemic stroke. **Results:** During a median (1^{st} and 3^{rd} quartiles) follow-up of 12.6 (11.8, 13.8) years, there were 8340 incident CVD events (5665 MI and 3206 ischemic stroke cases). 30.4% of the study population had 1 or more heteroplasmies. We found that higher MSS was associated with CVD (adjusted hazard ratio [aHR] for a 1-unit increase in MSS 1.16; 1.04, 1.28). On the other hand, mtDNA-CN was not associated with CVD (aHR for 10-copy increase in mtDNA-CN 1.00; 0.98, 1.02) after adjusting for neutrophil and platelet counts. Additionally, there was no interaction between MSS and mtDNA-CN on the risk of CVD (p for interaction 0.11). When stratified by complex / region, higher MSS in Complex I increased the risk of CVD by 31% (aHR 1.31; 1.09, 1.57) but not in other complexes / regions. In addition, higher MSS was associated with incident MI (aHR for a 1-unit increase in MSS 1.24; 1.09, 1.40), whereas it was not associated with incident ischemic stroke (aHR for a 1-unit increase in MSS 0.96; 0.81, 1.15). For MI, higher MSS in Complex I and rRNA regions was associated with higher incidence. **Conclusion:** These results indicate that mitochondria may have a functional role in the development of CVD, particularly MI, and MSS may serve as a biomarker for CVD risk.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1666 † Multi-ancestry genetic analysis of long COVID in 23andMe Research Participants

Authors:

N. Chaudhary, C. Weldon, P. Nandakumar, J. Shelton, 23andMe Research Team, M. Holmes, S. Aslibekyan; 23andMe, Sunnyvale, CA

Background: Long COVID (LC), a post-acute sequelae of SARS-CoV2 infection, is a heterogeneous condition with poorly understood etiology and limited insights into its genetic architecture. We identified genetic risk factors predisposing individuals with a history of acute SARS-CoV2 infection to LC in a large-scale study of 23andMe, Inc research participants . Methods: Consented and genotyped 23andMe adult research participants who had previously reported SARS-CoV2 infection completed web-based surveys on LC diagnosis, LC-related symptoms and impact on daily activities, initial disease severity, and vaccination status. Cases were defined via self-report of LC experience or diagnosis; controls had a history of acute SARS-CoV2 infection but did not self-identify as having LC. We conducted GWAS adjusting for age, sex, and ancestry-specific principal components among European, African American and Latinx populations. We then conducted multi-ancestry meta-analysis. Variants that met a genome-wide statistical threshold (P<5*10-08) were functionally characterized. We estimated genetic correlations of LC with phenotypes with similar symptomatology using LD score regression. Results: Among the 171,025 participants that responded to the LC survey, 50,393 cases had self-reported LC and 120,632 were identified as controls. The mean age of participants was 45 (SD=15) years. Compared to controls, LC was more common among females (77% vs 64% in controls), and individuals with LC had more co-morbidities. Twenty-six percent of individuals with LC reported a moderate to severe impact on daily activities. Our GWAS in the European population yielded four independent loci: hits on chromosome 3 were related to depression and anxiety disorder, and hits on chromosome 6 were related to COPD as well as to immune function. The immune-related loci remained genome-wide significant (P-value for the index variant=8.8x10-10) in the multi-ancestry analysis. LC showed strong genetic correlations with chronic fatigue (rg=0.71), fibromyalgia (rg=0.70), chronic pain (rg=0.69), and dysautonomia (rg=0.54). Conclusion: We conducted the largest GWAS of LC to date and identified multiple genetic associations. We also report evidence of a shared genetic architecture between LC and several chronic disease phenotypes.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1667 Multi-ancestry genetic architecture of heart failure subtypes.

Authors:

C. Liu[1], Q. Hui[1,2], Q. Wells[3], J. M. Gaziano[4,5], P. Wilson[1,6], the Million Veteran Program, L. Phillips[1,6], J. Joseph[7,8], Y. Sun[1,2]; [1]Emory Univ. Rollins Sch. of Publ. Hlth., Atlanta, GA, [2]Atlanta VA Hlth.care System, Decatur, GA, [3]Vanderbilt Univ. Sch. of Med., Nashville, TN, [4]VA Boston Hlth.care System, Boston, MA, [5]Harvard Med. Sch., Boston, MA, [6]Emory Univ. Sch. of Med., Atlanta, GA, [7]VA Providence Hlth.care System, Providence, RI, [8]The Warren Alpert Med. Sch. of Brown Univ., Providence, RI

Heart failure (HF) affects an estimated 6.2 million patients in the US, and is associated with high morbidity, mortality, and healthcare cost. Clinical subtypes are currently defined by the left ventricular ejection fraction (LVEF), including two major subtypes - HF with reduced ejection fraction (HFrEF) with LVEF ≤ 40%, and HF with preserved ejection fraction (HFpEF) with LVEF ≥ 50%. Genetic factors contribute to HF risk, supported by an estimated heritability of 26%. The genetic architecture of HFrEF and HFpEF can be different given their distinct pathophysiology and risk profile. Previous large genome-wide association studies (GWAS) mainly focused on all-cause HF. The genetic architecture of HF subtypes was studied only among European ancestry. Therefore, we performed a multi-ancestry GWAS meta-analysis to identify HF-associated loci, focusing on clinical subtypes of HFrEF and HFpEF. Ancestry-specific GWAS of all-cause HF, HFrEF and HFpEF were conducted among European, African, Hispanic and Asian American participants from the Million Veteran Program (MVP), and European and African American participants from the BioVU study. Under the assumption of additive genetic model, logistic regression models were used, with adjustment of age, sex, and top 10 principal components derived from each ancestry group. Meta-analysis was then performed using the Genome-Wide Association Meta-Analysis (GWAMA) software, which implements random-effects model that allows heterogeneity in effect sizes. Using a total of 200,070 all-cause HF cases and 2,076,466 controls of European, African, Hispanic, Asian ancestries from the MVP, BioVU, and a recent multi-ancestry GWAS, we identified 136 genome-wide significant all-cause HF loci, including 93 novel loci. Using 38,781 HFrEF cases, 38,163 HFpEF cases and 526,135 controls of European, African, Hispanic, Asian ancestries from the MVP and BioVU, we identified 34 novel loci for HFrEF including *MAP3K7CL* and *SPI1*, and 2 novel loci for HFpEF (*ARSG, LOC107985511*). Two loci (*FTO, ARSG*) were shared between the two subtypes. Using Genotype-Tissue Expression (GTEx) data, the gene expression analysis suggested that the HFrEF loci are enriched in heart tissues such as coronary artery and aorta, while HFpEF loci are enriched in tissues related to the endocrine system, including the pituitary and adrenal gland. These findings from this diverse population shed light on the differences in the underlying molecular pathways of HFrEF and HFpEF. Future studies are warranted to further characterize the less-understood HFpEF phenotypes and expand the genetic discoveries among diverse populations.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1668 Multi-ancestry genome-wide analysis of circulating D-dimer.

Authors:

J. Nicholas[1], M. Brown[2], J. Huffman[3], M. Sabater-Lleal[4], P. S. De Vries[2], M-H. Chen[5,6], F. Thibord[5,6], A. P. Reiner[7], A. Morrison[2], N. Smith[8,9,10], L. M. Raffield[1]; [1]Dept. of Genetics, Univ. of North Carolina, Chapel Hill, Chapel Hill, NC, [2]Human Genetics Ctr., Dept. of Epidemiology, Human Genetics, and Environmental Sci., Sch. of Publ. Hlth., The Univ. of Texas Hlth.Sci. Ctr. at Houston, Houston, TX, [3]Palo Alto Veterans Inst. for Res., Boston, MA, [4]Unit of genomics of Complex Disease, Inst. d'Investigació Biomèdica Sant Pau (IIB SANT PAU), Barcelona, Spain, [5]Div. of Intramural Res., Population Sci. Branch, Natl. Heart Lung and Blood Inst., Framingham, MA, [6]The Framingham Heart Study, Framingham, MA, [7]Div. of Publ. Hlth.Sci., Fred Hutchinson Cancer Ctr., Seattle, WA, [8]Dept. of Epidemiology, Univ. of Washington, Seattle, WA, [9]Kaiser Permanente Washington Hlth.Res. Inst., Kaiser Permanente Washington, Seattle, WA, [10]Seattle Epidemiologic Res. and Information Ctr., Dept. of VA Office of Res. and Dev., Seattle, WA

D-dimer is a peptide product of fibrinolysis and clinical biomarker of diseases involving activated coagulation, including venous thromboembolism and ischemic cardiovascular disease. Circulating D-dimer levels are heritable and heterogeneous across populations, with, for example, higher levels observed in African (AFR) ancestry relative to European (EUR) ancestry populations. To date, our understanding of genetic contributors to D-dimer variation has been limited to the European ancestry population. Here, we performed the largest, most diverse genome-wide analysis of circulating D-dimer [total n=46,031: including 36,688 EUR, 7,397 AFR, 1,321 Hispanic, and 654 East Asian participants]. We performed both single variant and gene-based aggregate rare variant tests in whole genome sequences (WGS) from 14,334 participants within the Trans-Omics for Precision Medicine (TOPMed) program. For single variant tests, we then meta-analyzed TOPMed-imputed genotype results from 31,697 participants within the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) consortium using a fixed-effects, p-value based method. Single-variant analysis results revealed 5 known genetic loci associated (P<5E-9) with D-dimer, including the AFR-driven *HBB* sickle-cell locus, and 2 new signals, at *F2* and *PROCR*, respectively. The lead variant at *PROCR* is a common (EAF=10%) intergenic variant in near-perfect correlation (r^2>0.99) with rs867186, a missense variant in *PROCR.* Variants at this locus have previously been associated with circulating measures of Protein C, Factor VII, and Factor XI, as well as thrombotic disease risk. Conditional analyses in TOPMed identified multiple variants distinctly associated with D-dimer at the *F5* locus and the fibrinogen gene locus**.** Sex-differentiated effects were observed at the *F3* locus, where effect-size heterogeneity tests in TOPMed revealed the lead variant (rs2022309) had a significantly higher effect-size for D-dimer association in females compared to males (P=1.57E-04). Gene-based aggregate tests further implicate *FGL1* missense variants (P=7.46E-06) in D-dimer regulation. Together, these loci provide new targets for functional work to disentangle mechanisms regulating fibrin production and fibrinolysis.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1669 Multi-ancestry genome-wide association study of broad psychiatric symptom domains in the Adolescent Brain Cognitive Development Study.

Authors:

M. E. Kumagai[1], L. M. Hernandez[1], M. J. Gandal[2]; [1]Univ. of California, Los Angeles, Los Angeles, CA, [2]Univ. of Pennsylvania, Philadelphia, PA

Psychiatric symptoms and disorders in adolescence are often highly heritable, yet few genome-wide association studies (GWAS) have been conducted in this population. To address this critical gap, here we ran GWAS on a broad range of quantitative psychiatric traits using the Child Behavior Checklist (CBCL)—a widely used parent-report questionnaire indexing child behavior across a number of psychiatric domains. Data were acquired for 10,612 multi-ancestry youth between the ages of 9-11 who participated in the Adolescent Brain Cognitive Development Study (European: EUR, N=6,654; African American: AFR N=2,168; Admixture American: AMR N=1,790). Within ancestry heritability and GWAS analyses were conducted using GCTA and FastGWA-mlm with covariates including age, sex, genotyping array batch, and the first 10 ancestry principal components. Multi-ancestry fixed effect meta-analyses were then performed using METAL. We identified a genome-wide significant association between the CBCL Thought Problems Scale (i.e. psychosis symptoms) at chromosome 8p21.2 (index SNP rs113595871, $p=3.4x10^{-8}$). Concordantly, Thought Problems Scale exhibited significantly heritability across ancestries—38.5%, 66.4%, and 54.6% for AFR, AMR, EUR groups, respectively ($p<0.001$)—indicating substantial genetic contribution to this childhood behavioral trait. Genetic correlations showed significant positive genetic correlations between the Thought Problems Scale and all other CBCL subscales ($p<0.001$), except for the Anxiety/Depression Scale ($p=0.38$). Finally, functional genomic follow up analyses identified specific enrichment of Thought Problems Scale GWAS results with gene expression in the pituitary ($p<0.001$), suggesting potential involvement of the endocrine system in the biological processes underlying adolescent thought problems. Together, these findings enhance our understanding of the genetic underpinnings of adolescent psychiatric symptoms and underscore the significance of considering diverse ancestry groups in investigating genetic associations with behavioral traits.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1670 Multi-ancestry genome-wide association study of prostate-specific antigen levels identifies novel loci and improves cross-population prediction

Authors:

T. Hoffmann[1], R. Graff[2], R. Madduri[3], A. Rodriguez[4], A. Wang[5], R. Klein[6], B. Pierce[7], L. Tong[8], W. Blot[9], J. Long[10], T. Rebbeck[11], J. Lachance[12], C. Andrews[13], I. Agalliu[14], Y. Jiang[2], S. Berndt[15], K. Schaffer[10], M. Machiela[15], P. Goodman[16], C. Till[17], I. Thompson[18], H. Lilja[19], S. Van Den Eeden[20], S. Chanock[21], J. Mosley[10], D. Conti[22], C. Haiman[23], A. Justice[24], L. Kachuri[25], J. Witte[26]; [1]UCSF, Millbrae, CA, [2]UCSF, San Francisco, CA, [3]Argonne Natl. Lab., Naperville, IL, [4]Argonne Natl. Lab., Lincolnwood, IL, [5]Univ. of Southern California, Pasadena, CA, [6]Icahn Sch. of Med. at Mount Sinai, New York, NY, [7]Univ. of Chicago, Chicago, IL, [8]Univ Chicago, chicago, IL, [9]vanderbilt Univ. Med. center, nashville, TN, [10]Vanderbilt Univ. Med. Ctr., Nashville, TN, [11]Univ Pennsylvania Sch Med, Philadelphia, PA, [12]Georgia Inst. of Technology, Atlanta, GA, [13]Harvard Sch. of Publ. Hlth., Boston, MA, [14]Albert Einstein Coll. of medicine, Bronx, NY, [15]Natl. Cancer Inst., Rockville, MD, [16]Fred Hutchinson Cancer research Ctr., Seattle, WA, [17]Fred Hutchinson Cancer research Ctr., Seattle, UT, [18]cristus Santa Rosa Med. Ctr. Hosp., San Antonio, TX, [19]Mem. Sloan Kettering Cancer Ctr., New York, NY, [20]Kaiser division of research, Oakland, CA, [21]Natl. Cancer Inst, Rockville, MD, [22]Univ of Southern California, Los Angeles, CA, [23]USC, Santa Cruz, CA, [24]Veterans Admin. Connecticut Hlth.care system, West Haven, CT, [25]Stanford Univ., San Francisco, CA, [26]Stanford Univ., Stanford, CA

Prostate-specific antigen (PSA) is encoded by KLK3 and abundantly secreted by the prostate gland with a miniscule fraction (10-6) normally circulating in blood, but levels increase with prostate cancer. Most countries recommend against universal population screening for prostate cancer with prostate-specific antigen (PSA) testing due to potential for overdiagnosis and overtreatment of non-aggressive disease. Heritability of PSA in blood ranges between 30% and 40%, suggesting that polygenic variation is an important modulator of basal PSA levels in blood. Adjusting measured PSA levels for an individual's genetically determined baseline PSA could remove variation unrelated to cancer risk and improve overall test accuracy. Previous work investigating genetic determinants of PSA has focused largely on European ancestry individuals and has identified 128 independent variants, but many more PSA-associated loci remain to be discovered. Thus, we conducted the largest ever multi-ancestry genome-wide association study (GWAS) of PSA levels in 296,754 men without prostate cancer (211,342 European ancestry; 58,236 African ancestry; 23,546 Hispanic/Latino; and 3,630 Asian ancestry) across 9 cohorts (Million Veteran Program, 96.5%).

We identified 319 independent variants ($p<5x10^{-8}$), 185 of which were novel. Most of the novel variants had evidence of replication in an independent sample of 95,768 men. Of the novel variants, 47 reached genome-wide significance only in the multi-ancestry analysis (compared to individual ancestry-specific analysis), including rs372203682 in LMTK2, a gene implicated in spermatogenesis. A signal in the androgen receptor (AR) gene that was detected in the multi-ancestry analysis (rs184476359, $P_{Meta}=3.4x10^{-10}$) was primarily driven by the African ancestry population ($P_{AFR}=5.8x10^{-9}$). Combining GWAS results from the discovery and replication meta-analyses (N=392,522) identified an additional 112 novel variants. A total of 21% of discovery identified variants and 17% of meta-analysis identified variants met a Bonferroni significance level in the PRACTICAL consortium prostate cancer GWAS (107,247 cases; 127,006 controls).

The variance in PSA levels explained by a multi-ancestry genome-wide polygenic risk score (PRS-CSx) derived from our discovery GWAS in multiple independent cohorts was up to 16.9% (95% CI=16.1%-17.8%) in European ancestry, 9.5% (95% CI=7.0%-12.2%) in African ancestry, 18.6% (95% CI=15.8%-21.4%) in Hispanic/Latino, and 15.3% (95% CI=12.7%-18.1%) in Asian ancestry. Altogether, these results further our efforts towards more personalized and equitable prostate cancer screening.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1671 † Multi-ancestry genome-wide meta-analysis of 56,241 individuals identifies *LRRC4C* and*LHX5-AS1* and nominates ancestry-specific loci *PTPRK*, *GRB14*, and *KIAA0825* as novel risk loci for Alzheimer's disease: the Alzheimer's Disease Genetics Consortium (ADGC)

Authors:

A. Naj[1], C. Reitz[2], F. Rajabli[3], G. Jun[4], P. Benchek[5], G. Tosto[2], J. Sha[1], K. Bazemore[1], C. Zhu[4], N. Kushch[3], W-P. Lee[1], J. Haut[1], K. Hamilton-Nelson[3], N. Wheeler[5], Y. Zhao[1], J. Farrell[4], J. Chung[4], M. Grunin[5], Y. Leung[1], P. Kuksa[1], D. Li[4], E. Lucio Da Fonseca[3], J. Mez[4], E. Palmer[5], J. Pillai[6], R. Sherva[4], Y. Song[5], X. Zhang[4], E. Greenfest-Allen[1], T. Iqbal[1], O. Pathak[1], O. Valladares[1], A. Kuzma[1], B. Kunkle[3], W. Bush[5], L-S. Wang[1], L. Farrer[4], J. Haines[5], R. Mayeux[2], M. Pericak-Vance[3], G. Schellenberg[1], Alzheimer's Disease Genetics Consortium (ADGC); [1]Univ. of Pennsylvania, Philadelphia, PA, [2]Columbia Univ., New York, NY, [3]Univ. of Miami, Miami, FL, [4]Boston Univ., Boston, MA, [5]Case Western Reserve Univ., Cleveland, OH, [6]Cleveland Clinic, Cleveland, OH

Previous Alzheimer's disease (AD) genetic studies lack ancestral diversity, limiting the ability to detect risk variants more prevalent in non-European ancestry groups. Here we constructed and analyzed the largest multi-ancestry collection to date of directly genotyped genome-wide association study (GWAS) datasets from the Alzheimer's Disease Genetics Consortium (ADGC) to identify novel shared and ancestry-specific AD susceptibility loci, evaluate genetic architecture in different ancestries, and explore potential functional effects. These data included TOPMed R5-imputed genotypes from 37,382 non-Hispanic White (NHW), 6,728 African American (AFA), 8,899 Hispanic (HIS), and 3,232 East Asian (EAS) individuals. We performed a two-stage analysis: (1) single-variant association using score-based logistic regression for population-based datasets and GLMM for family-based studies, adjusting for onset/exam age, sex, population substructure principal components, and *APOE* ε4 genotype, followed by within-ancestry fixed-effects meta-analysis using METAL; and (2) cross-ancestry meta-analysis using the Han-Eskin random-effects (RE_{HE}) model in METASOFT. Secondary analyses included functional annotation, differential expression follow-up, and pathway analysis using EnrichR. We identified 13 loci with genome-wide significant (GWS; $P \leq 5\times10^{-8}$) cross-ancestry associations. These include known AD loci at/near *CR1*, *BIN1*, *TREM2*, *CD2AP*, *PTK2B*, *CLU*, *SHARPIN*, *MS4A6A*, *PICALM*, *ABCA7*, *APOE* and two novel loci at 11p12 (*LRRC4C*) and 12q24.13 (*LHX5-AS1*). We identified three GWS ancestry-specific loci, *PTPRK* ($P=2.4\times10^{-8}$) and *GRB14* ($P=1.7\times10^{-8}$) in HIS and *KIAA0825* ($P=2.9\times10^{-8}$ in NHW). Functional follow-up revealed the lead *LRRC4C* SNP, rs12576934, is a significant splicing quantitative trait locus (sQTL) for *LRRC4C* in brain frontal cortex ($P=5.89\times10^{-4}$) in GTEx expression data. Pathway analysis implicated multiple amyloid regulation pathways (strongest: negative regulation of APP catabolism, $P_{adjusted}=1.6\times10^{-4}$) and the classical complement pathway ($P_{adjusted}=1.3\times10^{-3}$). In summary, cross-ancestry analyses identified novel AD susceptibility loci in/near *LRRC4C* and *LHX5-AS1*, both genes with known roles in neuronal development, and several novel ancestry-unique loci found in biologically plausible pathways, including *PTPRK*, which encodes a protein tyrosine phosphatase involved in neuronal development and outgrowth, and *GRB14*, which encodes a negative modulator of insulin receptor activity. These highlight the value of data from traditionally underrepresented populations for gene discovery in AD, even with smaller sample sizes.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1672 Multi-ancestry phenome-wide association analyses of genetic liability for problematic alcohol use

Authors:

R. Kember[1], J. Johnson[2], K. Johnston[3], H. Lee[4], J. Xu[3], L. Davis[5], J. Smoller[4], T. Mallard[4], L. Huckins[3], S. Sanchez-Roige[6], H. Kranzler[1], J. Gelernter[3], H. Zhou[3]; [1]Univ. of Pennsylvania Perelman Sch. of Med., Philadelphia, PA, [2]Icahn Sch. of Med. at Mount Sinai, New York, NY, [3]Yale Sch. of Med., New Haven, CT, [4]Massachusetts Gen. Hosp., Harvard Med. Sch., Boston, MA, [5]Vanderbilt Univ. Med. Ctr., Nashville, TN, [6]Univ. of California San Diego, La Jolla, CA

Problematic alcohol use (PAU) and alcohol use disorder (AUD) are major causes of death and disability worldwide, in part because they increase risk for other mental and medical illnesses. While PAU may potentiate the course of comorbid conditions, there is accumulating evidence suggesting common genetic architectures for these disorders. We conducted a phenome-wide association study (PheWAS) to improve our understanding of the relationship between genetic liability for PAU and comorbid conditions. We used a cross-ancestry genome-wide association study (GWAS) of PAU (a phenotype based on a meta-analysis of AUD, alcohol dependence, and AUDIT-P [Alcohol Use Disorders Identification Test-Problem score, a measure of problematic drinking], N=1,079,947) for discovery. We calculated polygenic risk scores (PRS) using PRS-continuous shrinkage (PRS-CS) for PAU (in EUR; N=131,500) and AUD (in AFR; N=27,494) in four independent datasets (Vanderbilt University Medical Center's Biobank [BioVU], Mount Sinai [BioMe], Mass General Brigham Biobank [MGBB] and Penn Medicine Biobank [PMBB]) from the PsycheMERGE Network. We performed a PheWAS in each dataset, followed by a meta-analysis across all sites. In the EUR cross-site meta-analysis, after controlling for multiple comparisons, 58 of the 1,493 tested phenotypes were significantly associated with the PAU PRS, and 78 were significantly associated with the AUD PRS. Patterns of associations were mostly consistent across sites, although there was significant heterogeneity in the strength of associations. The most common associations were with substance use or mental disorders (e.g., tobacco use disorder, mood disorder, posttraumatic stress disorder), respiratory traits (e.g., chronic airway obstruction), and neurological conditions (e.g., chronic pain). Phenotypes potentially arising from alcohol use were also identified, such as alcoholic liver damage, and chronic liver disease and cirrhosis. In the AFR cross-site meta-analysis, 3 of the 793 phenotypes were significantly associated with the AUD PRS, all substance related (alcoholism, alcohol-related disorders, and tobacco use disorder). Covarying for AUD in the target samples reduced the significance, but not the pattern, of associations, with most phenotypes remaining significantly associated following correction. Genetic liability for PAU and AUD is associated with a range of disorders across the phenotypic spectrum, even in the absence of an AUD diagnosis. This suggests that there are shared pathways of risk for these common comorbid conditions.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1673 Multi-ancestry polygenic risk scores improve prediction of PTSD symptom severity and trajectory

Authors:

M. Sendi[1], Y. Zhao[2], X. An[2], D. A. Pizzagalli[1], K. Ressler[1], AURORA group, N. Daskalakis[1]; [1]Harvard Med. Sch./ McLean Hosp., Boston, MA, [2]The Univ. of North Carolina, Chapel Hill, NC

Post-traumatic stress disorder (PTSD) can affect those who experience trauma, and genetic variants for PTSD have been identified in different ancestry populations. But, the potential of polygenic risk score (PRS) for predicting genetic risk for PTSD and its role and interaction with demographic factors in predicting different PTSD trajectories are still unclear. Here, we aim to understand the role of multi-ancestry PRS in predicting PTSD symptoms and trajectories and their interaction with demographic factors, including age, sex, years of education, body mass index, and socioeconomic conditions. Data of 2637 individuals seeking emergency department care after traumatic stress exposure were analyzed as part of the AURORA study. Participants aged 18 to 75 sought care within 72 hours of a traumatic event at selected sites. Participants were phenotyped at pre-trauma (PRE), two weeks (WK2), eight weeks (WK8), three months (M3), six months (M6), and twelve months (M12) post-trauma using the PCL-5 scale. We calculated PRS using three genome-wide association study datasets: PRS_{EUR} (European), PRS_{AFR} (African), and PRS_{META} (European and African). Next, the null model (covariates only) R^2 values were compared to the full model (PRS and covariates) ones to evaluate PRS as a predictor of PCL-5 at different time points. Also, participants were grouped into resilient PTSD (Res-PTSD), moderate PTSD (Mod-PTSD), recovery PTSD (Rec-PTSD), and continuous PTSD (Con-PTSD) trajectories. Next, a binary classification method was employed to classify trajectory pairs based on demographic variables, genetic-based ancestry, and PRS data, with significant features identified using a feature learning approach. Including PRS_{META} at all time points ($p<0.001$) led to a significant improvement in the model, with the full model consistently exhibiting higher R^2 values compared to the null model. Furthermore, PRE_{AFR} contributed to the model improvement by increasing the R^2 values of the full model compared to the null model at PRE, WK2, WK8, and M6 ($p<0.05$ to $p<0.01$). On the other hand, PRS_{EUR} did not contribute to the improvement of the model. The integration of demographic, ancestry, and PRS information yielded a significantly better prediction of the PTSD trajectory than what would be expected by chance. Notably, certain demographic variables (including age, sex, and education level) and PRS_{META} were the most effective predictors of the PTSD trajectory. In conclusion, multi-ancestry PRS outperformed single-ancestry ones in predicting PTSD symptoms, and a combination of demographic factors and PRS_{META} effectively predicted different PTSD recovery trajectories.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1674 Multi-ethnic meta-analysis GWAS of autism uncovers four novel genes.

Authors:

J. Glessner[1,2], M. Khan[1], X. Chang[1], Y. Liu[1], F. Mentch[1], C. Kao[1], M. March[1], J. Connolly[1], H. Hakonarson[1,2]; [1]Childrens Hosp. of Philadelphia, Philadelphia, PA, [2]Univ. of Pennsylvania Perelman Sch. of Med., Philadelphia, PA

We report the results of a Genome-Wide Association Study (GWAS) meta-analysis of Autism Spectrum Disorder (ASD) that has identified several significant novel genes associated with the disorder. The study was conducted as part of an ongoing autism cohort analysis at the Center for Applied Genomics (CAG) program at Children's Hospital of Philadelphia (CHOP), which uses SNP array genotyping and imputation to identify genetic variants that may contribute to complex diseases. Our analysis included a total of 7,714 individuals with ASD and 19,620 controls of EUR, AFR, AMR, EAS, and SAS ancestry. We used imputation referencing Trans-Omics for Precision Medicine (TOPMed) data to increase the number of genetic variants tested, and then performed a meta-analysis across multiple race and sex stratified cohorts (using tools SAIGE and METAL) to identify genetic variants associated with ASD. Our results revealed several significant genes that have not previously been linked to ASD. Specifically, we identified three genes that showed a strong association with ASD in our meta-analysis: FOXL3-OT (p=1.87E-8), MCPH1-AS1(p=2.69E-8), and DDAH1 (p=2.97E-8). We also identified a fourth gene, DNAH14 (p=3.52E-7), that was borderline significant but still showed a strong trend towards association with ASD. FOXL3-OT is an antisense RNA gene located on chromosome 1p31.3 that has been implicated in various biological processes, including cell differentiation and development. DDAH1 is involved in the regulation of nitric oxide synthesis, which is important for neuronal development and function, including neuronal synaptogenesis and synaptic pruning. MCPH1-AS1 is an antisense RNA gene that has been linked to microcephaly, a condition characterized by a small head circumference that is associated with cognitive impairment. Finally, DNAH14 encodes a protein involved in the movement of cilia, which are essential for the development and function of various organs, including the brain. While the exact mechanisms by which these genes contribute to ASD are still unclear, our findings suggest that they may play a role in the development and function of the brain. Further studies will be needed to fully understand the functional consequences of the genetic variants we have identified. In conclusion, our GWAS meta-analysis has identified several novel genes associated with ASD in ethnically diverse patients, providing new insights into the biological basis of this complex disorder. Our results may ultimately lead to the development of new diagnostic and therapeutic strategies for individuals with ASD.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1676 Multi-modal characterization of type 2 diabetes in Project Baseline beyond clinical diagnosis

Authors:

A. Breschi, Y. Wang, S. Short, W. Luk, P. Kheradpour, P. Cimermancic, G. J. Tong, J-P. Martin, M. Liu, L. Cao, D. Liu, A. Han, K. A. Drake, C. C. Kim; Verily Life Sci., South San Francisco, CA

Around 11% of the US adult population, 28 million people, have been estimated to have type 2 diabetes, with 3% more, 8.5 million people, estimated to be undiagnosed. The main clinical diagnostic tools are measurements of glycated hemoglobin (HbA1c) in blood and blood glucose level, fasting or in response to a glucose oral tolerance test. Understanding the disease at the molecular level is paramount to improve diagnosis and prevention. The Project Baseline Health Study longitudinally performed deep phenotypic, clinical and molecular profiling for 2,502 volunteers.Comprehensive profiling included demographics, medical history, laboratory tests, echocardiograms, physical assessments, lifestyle monitoring through a proprietary smart watch and multi-omics assays. Of these, 957 participants had sufficient clinical and proteomics data from their initial visit. We integrated information from the participants' medical history and laboratory tests to determine their diabetes status, so we could identify biomarker plasma proteins of diabetes. By integrating clinical and molecular data, we trained a logistic regression classifier that better discriminated between diabetic and normoglycemic individuals as compared to clinical and proteomics data alone. We identified 85 high confidence differentially expressed proteins in diabetics. Of these, 71 were upregulated in diabetics; are involved in lipid transport, inflammation, and blood coagulation; and appeared to be transcribed in distinct liver zones. By clustering the participants on their clinical and molecular profiles, we noticed that inflammation-related proteins were also present in overweight normoglycemic participants with diabetic-like proteomics profile, and that this profile was associated with worse physical performance and cardiac health as measured by orthogonal signals. We observed that proteomics data can be used to predict diabetes status with balanced accuracy over 85% when used in combination with clinical data. We showed that molecular data can enhance diabetes phenotyping obtained with clinical data, suggesting that a more holistic approach to diabetes diagnosis, beyond HbA1c and glucose measurements, might help identifying people with diabetes or metabolic alterations, and target them with personalized treatments or interventions.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1677 Multiomic association network: Capturing cardiovascular disease complexity with the genomic, metabolomic, drug treatment, and disease features.

Authors:

M. Aguilar[1], M. A. Hall[2]; [1]Penn State Univ., State College, PA, [2]The Pennsylvania State Univ., State College, PA

In the US, the prevalence of cardiovascular disease (CVD) in adults ≥ 20 years of age was 48.6% overall (127.9 million in 2020). CVD encompasses various phenotypes: stroke, heart disease, hypertension, diseases of the veins, peripheral artery disease, and others. Previous genome-wide association studies revealed that CVD has a complex genetic architecture that can be understood by analyzing molecular components underlying the trait. Metabolomics (a collection of small quantified molecules) studies intermediate molecular traits to be used as markers of CVD. Commonly prescribed drugs for CVD have metabolic on- or off-target effects that require mapping; this is crucial in generating new drug efficacy and repurposing hypotheses. Thus, the following domains, metabolomic, drug treatment, and disease phenotypes were evaluated in our study to supplement our understanding of complex diseases beyond their genetic basis. The statistically significant results from association tests between domain features can be organized into a network to generate new hypotheses about the features' roles in disease development. We used data from the LUdwigshafen RIsk and Cardiovascular Health (LURIC) study because it investigated the metabolic and genetic risk factors of CVD, and each sample had data on all four domains of interest (drugs, metabolites, and genomic single nucleotide polymorphisms (SNPs) used to build our association network. We ran nine regression analyses cumulatively yielded 3,254 statistically significant results (total tests: 183,520,500). Each regression analysis had a Bonferroni corrected significance threshold ($\alpha = 0.05$) divided by the number of outcomes (5×10^{-8} divided by the number of outcomes for GWAS). The network was analyzed with clustering coefficient, betweenness centrality, and PageRank algorithms to identify network hubs. Dyslipidemia disease was the top-ranked feature in the network (1054 significant results, i.e., 732 in-degrees and 322 out-degrees) according to betweenness centrality and PageRank measurements; this indicated that dyslipidemia was central to CVD and had associations across the genomic, metabolic, and drug domains. New genome-metabolite associations related to collagen synthesis and cell proliferation may describe CVD's coronary and vascular structural changes. When targeting genes is not feasible as a therapy, an association network offers new hypotheses on targetable molecular traits. Results show how a high-dimensional search space can be narrowed to identify features for future multi-factor risk models of complex diseases.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1678 Multi-omic integration uncovers biological pathways underlying HIV viral load

Authors:

K. Sullivan[1], I. Miller[1], B. C. Quach[2], C. Willis[3], M. R. Garvin[1], A. Townsend[4], P. Kruse[4], M. Lane[4], D. Kainer[1], M. Cashman[1], K. Xu[5], B. E. Aouizerat[6], D. B. Hancock[2], D. A. Jacobson[1], E. O. Johnson[7]; [1]Oak Ridge Natl. Lab., Oak Ridge, TN, [2]RTI Intl., Durham, NC, [3]RTI Intl., Carrboro, NC, [4]Univ. of Tennessee-Knoxville, Knoxville, TN, [5]Yale Sch. of Med., New Haven, CT, [6]New York Univ., New York, NY, [7]RTI Intl., Duluth, MN

While antiretroviral therapy (ART) has significantly improved disease prognosis in people living with HIV (PLWH), substance use in these individuals can increase HIV/AIDS viral load and accelerate disease progression. It is critical to understand the biological mechanisms underlying HIV viral load to understand how to prevent substance use from exacerbating HIV disease progression. Here, we integrated multi-omic datasets and used machine learning network biology tools to identify biological mechanisms associated with HIV viral load across 10 cohorts: the Veterans Aging Cohort Study (VACS; n = 2,465), Swiss HIV Cohort (n = 198), and 7 cohorts of whole-genome sequencing data used to obtain an MHC locus fine mapping reference panel (n = 21,546). We integrated three genes from HIV set point viral load GWAS using fine mapping from WGS data (*HLA-A*, *HLA-B*, *HLA-C*), 258 differentially expressed genes in CD4+ T-cells associated with HIV viral load from the Swiss HIV Cohort, 143 genes based on differential DNA methylation status from the VACS cohort, and 8 genes known to affect the pharmacokinetics of anti-HIV drugs (*ABCC2*, *CYP2A6*, *CYP2B6*, *CYP2C9*, *CYP2C19*, *CYP3A4*, *SLCO1B1*, and *UGT1A1*). Using our network biology tool GRIN, we retained 194 genes related to HIV viral load based on their network topological connectivity to one another. Next, we used our recently developed Functional Partitioning tool to identify functional network relationships among our multi-omic viral load genes, based upon hierarchical clustering of multiplex network random walk-based vector comparisons among genes in the gene set. We identified three large clades of highly inter-connected genes based on these network topology-based vector comparisons. Within the first clade, we identified genes known to be involved in cell cycle checkpoints, DNA repair, and epigenetic modifications to histones. In the second clade, we identified genes known to be involved in mitochondrial function and oxidative stress. In the third clade, we identified genes known to be involved in viral replication and host antiviral response. Across all clades we identified several host genes known to physically interact with HIV viral proteins. The empirically derived, densely connected, network is highly enriched for gene sets linked to non-AIDS defining cancers. Together, we demonstrate that multi-omic integration using network biology tools can identify biological pathways and create a conceptual model of the mechanisms underlying HIV viral load that would not have been identified with the use of a single omics data type.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1679 Multi-omics analyses to prioritise genes at migraine risk loci.

Authors:

A. Ghaffar, The International Headache Genetics Consortium, D. Nyholt; Queensland Univ. of Technology, Brisbane, Australia

Migraine—a painful, throbbing headache disorder that can cause nausea, vomiting, and sensitivity to light and sound—is the most common brain disorder, affecting almost 14% of the adult population. Genome-wide association studies (GWAS) have proven successful in identifying migraine risk loci. However, most of the GWAS risk variants are in non-coding regions, hence they act by altering the regulatory mechanism such as gene expression. Epigenetic mechanisms such as methylation have also been reported to be associated with migraine. Utilizing data from the largest known migraine GWAS (Hautakangas et al. 2022), which comprised 102,084 migraine cases and 771,257 controls and identified 171 independent single nucleotide polymorphism (SNP) risk loci significantly associated with migraine, we performed an imputation-based transcriptome-wide association study (TWAS), to identify genetically regulated gene expression (eQTLs) associated with migraine. We used SMultiXcan (a gene-based TWAS approach that leverages the correlation of eQTLs across tissues) and prediction models generated from 49 GTEx tissues to impute differential expression associated with migraine for 21,647 genes. Of the 21,647 genes, 244 were study-wide significant at the Bonferroni adjusted threshold of $P < 2.31 \times 10^{-6}$, of which 212 genes were within ±1 Mb of one of the 171 independent SNP risk loci from Hautakangas et al. (2022). We also performed a methylome-wide association study (MWAS) which integrates GWAS and DNA methylation (mQTL) data to discover the genetically regulated DNA methylation sites (CpGs) associated with migraine. We used EstiMeth and prediction models from blood to impute differential methylation associated with migraine for 86,518 CpGs. Of the 86,518 CpGs, 258 were significant at the Bonferroni adjusted threshold of $P < 5.78 \times 10^{-7}$. The 258 CpGs mapped to 115 genes, of which 227 CpGs mapping to 99 genes were within ± 1 Mb of an independent risk SNP from Hautakangas et al. (2022). Notably, 43 genes at 33 loci were implicated by both the TWAS and MWAS analysis, suggesting genome-epigenome-transcriptome mechanistic relationships. Colocalization analyses of the migraine GWAS risk SNPs with the associated eQTL and mQTL within these genes will be completed by October 2023, to further characterize the likely functional consequence of these migraine risk loci.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1680 Multi-omics approaches reveal key molecular signature of severe obesity

Authors:

L. Petty[1], H-H. Chen[2], P. Sharma[3], H. G. Polikowsky[4], J. Seo[3], M. Y. Anwar[3], D. Kim[3], M. Graff[3], K. L. Young[3], W. Zhu[4], K. Karastergiou[5,6], D. M. Shaw[4], A. E. Justice[7], L. Fernandez-Rhodes[8], M. Krishnan[3], A. Gutierrez[9], P. Mccormick[10], P. Gordon-Larsen[3], M. Lee[11], H. M. Highland[3], E. R. Gamazon[1,12], N. J. Cox[1], S. K. Fried[5,6], S. P. Fisher-Hoch[13], J. B. McCormick[13], K. E. North[3], J. E. Below[1]; [1]Vanderbilt Univ. Med. Ctr., Nashville, TN, [2]Academia Sinica, Taipei, Taiwan, [3]Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, [4]Vanderbilt Univ., Nashville, TN, [5]Boston Univ. Sch. of Med., Boston, MA, [6]Icahn Sch. of Med. at Mount Sinai, New York, NY, [7]Geisinger Hlth., Danville, PA, [8]The Pennsylvania State Univ., University Park, PA, [9]The Univ. of Texas Hlth.Sci. Ctr. at Houston, Houston, TX, [10]Univ. of London, London, United Kingdom, [11]The Univ. of Texas Hlth.Sci. Ctr. at Houston Sch. of Publ. Hlth., Brownsville Regional Campus, Brownsville, TX, [12]Univ. of Cambridge, Cambridge, United Kingdom, [13]Univ. of Texas Hlth.Sci. Ctr. at Houston Sch. of Publ. Hlth., Brownsville Regional Campus, Brownsville, TX

Background: Severe obesity [body mass index (BMI) $\geq$ 40 kg/m^2] is a driver of many cardiometabolic diseases, and disproportionately impacts marginalized populations, including Hispanic/Latinos, yet we know little about its underlying mechanistic pathways. Multi-omic analyses can link genetic variation and disease, highlighting pathways for targeted therapeutic intervention, but have not been fully leveraged for studying severe obesity.
Methods: We deployed integrative multi-omics approaches in whole blood to enable discovery of new genes involved in severe obesity in Mexican Americans from the Cameron County Hispanic Cohort (CCHC). First, using RNA-sequencing data for 49 severe obesity cases and 81 controls (BMI < 25 kg/m^2) from CCHC, we assessed differential expression using DESeq2 and performed Mendelian randomization to estimate causality. We replicated our top findings in an independent sample of 52 cases and 59 controls from CCHC. We validated our transcriptomic results using proteomic data for 49 cases and 42 controls from the CCHC and explored the specificity of detected effects by leveraging gene expression data from abdominal subcutaneous adipose tissue from 19 community volunteers from New York City, New York. Lastly, we tested for enrichment of associations of genetically regulated expression of our identified genes with obesity-related diseases in a large electronic health record-linked biobank, BioVU, N=70k.
Results: We identified 124 significantly differentially expressed genes after false discovery rate correction (FDR <0.05), of which 33% replicated, and 22% of those measured showed differential protein abundance associated with severe obesity. Top genes included *C1RL*, *IL4R*, *RGS16*, *OSM*, and *SDC2*. Twenty-six of the differentially expressed genes showed correlation with BMI in abdominal subcutaneous adipose tissue in a diverse independent sample. Transcripts with causal effects on severe obesity also showed enrichment for causal impacts on sequelae of obesity, including hyperlipidemia and obstructive sleep apnea, in BioVU.
Conclusion: We provide compelling evidence for genes whose expression is associated with severe obesity in an underrepresented and disproportionately impacted population, observing highly concordant effects in an independent replication, generalization of effects in abdominal subcutaneous adipose with BMI, and translation effects in the proteome. Our findings illuminate new mechanisms of risk and consequences of severe obesity, revealing novel targets for prevention, intervention, and treatment.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1681 Multi-omics in nasal epithelium reveals three axes of dysregulation for asthma risk in the African Diaspora populations

Authors:

B. Szczesny[1], M. Boorgula[2], M. Campbell[3], R. Johnson[4], K. Kammers[5], I. Ruczinski[6], S. Chavan[2], C. Figueiredo[7], C. N. Rotimi[8], C. R. Landis[9], H. Watson[10], N. Hansel[1], I. Yang[11], C. Olopade[12], C. Ober[13], A. Liu[11], E. Kenny[14], M. Taub[15], K. Barnes[2], R. Mathias[16]; [1]Johns Hopkins Univ., Baltimore, MD, [2]Univ. of Colorado Denver, Denver, CO, [3]Univ. of Colorado, Aurora, CO, [4]Univ. of Colorado Anschutz Med. Campus, Aurora, CO, [5]Johns Hopkins Univ Sch. of Med., Baltimore, MD, [6]Johns Hopkins Bloomberg Sch. of Publ. Hlth., Baltimore, MD, [7]Federal Univ. of Bahia, Salvador, Brazil, [8]Natl. Human Genome Res. Inst., NIH, Bethesda, MD, [9]Caribbean Inst. for Hlth.Res., The Univ. of the West Indies, Bridgetown, Barbados, [10]The Univ. of the West Indies, Queen Elizabeth Hosp., Bridgetown, Barbados, [11]Univ. of Colorad Denver, Denver, CO, [12]Univ. of Chicago Med., Chicago, IL, [13]Univ. of Chicago, Chicago, IL, [14]Icahn Sch. of Med. at Mount Sinai, New York, NY, [15]Johns Hopkins, Baltimore, MD, [16]Johns Hopkins Univ, Baltimore, MD

Asthma, a complex chronic lung disease affecting the airways, has striking disparities, but the molecular underpinning of these differences is poorly understood and minimally studied. A major goal of the Consortium on Asthma among African-ancestry Populations in the Americas (CAAPA) is to use multi-omics to address this gap. RNASeq and DNA methylation data (DNAm) were generated from nasal epithelium (NE) in subjects recruited from 7 sites (US: Baltimore, Washington DC, Chicago, Denver; international: Salvador Brazil, Barbados, and Nigeria). Current asthma cases (N=253) were compared to never-asthma controls (N=283) to identify DEGs. Network analyses were performed with Ingenuity Pathway Analysis (IPA) and weighted gene co-expression network analysis (WGCNA). DNAm data were used to identify eQTMs for DEGs limiting to CpGs ≤5kb from the transcription start site or within enhancer regions identified by promoter-capture HiC in bronchial epithelial cells. CpGs from significant eQTMs were tested for differential methylation by asthma to assess the contributions of expression and methylation in asthma risk. All models were adjusted for ancestry, sampling site, and appropriate latent factors.
We identified 389 DEGs; the top DEG, FN1, was downregulated in asthma cases (q=3.26x10-9) and encodes fibronectin which plays a role in wound healing. Other top 10 DEGs are highly relevant for asthma: SNTG2 (q = 1.12x10-4) is the target of multiple miRNAs related to asthma; PPP1R9A expression (q=7.60x10-5) was previously found to be influenced by IL-13 in mouse lung; and SPTBN1 (q=1.12x10-4) plays a role in mediating TGFβ signaling. IPA revealed upstream regulators relevant for immune response (IL4; p=7.25x10-10 and TGFβ1; p=5.47x10-8) and drug response (dexamethasone; p=4.31x10-10 and fluticasone propionate; p=9.42x10-8). The top three WGCNA modules implicate networks related to immune response (CEACAM5; p=9.62x10-16 and CPA3; p=2.39x10-14) and wound healing (FN1; p=7.63x10-9). Multi-omic analysis identified FKBP5 as a key contributor to asthma risk, whereby the association between nasal epithelium gene expression is mediated through methylation and is associated with increased use of ICS. FKBP5 is a co-chaperone of glucocorticoid receptor signaling and known to be involved in drug response in asthma.
Our analyses reveal genes and networks differentially expressed in NE of asthma cases of African ancestry in CAAPA. Importantly, this work reveals molecular dysregulation on three axes - increased Th2 inflammation, decreased capacity for wound healing, and impaired drug response - that may play a critical role in asthma within the African Diaspora.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1682 † Multi-organ genetic causal connections inferred from imaging and clinical data through Mendelian randomization

Authors:

J. Shu[1], R. Zheng[2], C. Copana[1], B. Li[1], Z. Fan[3], X. Yang[4], Y. Yang[3], X. Wang[1], Y. Li[1], B. Xi[1], T. Li[5], H. Zhu[6], B. Zhao[7]; [1]Purdue Univ., West Lafayette, IN, [2]First Hosp. of Shanxi Med. Univ., Taiyuan, China, [3]Univ. of Pennsylvania, Philadelphia, PA, [4]Purdue Univ., Lafayette, IN, [5]Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, [67]Univ. of Pennsylvania, Dept. of Statistics and Data Sci., Philadelphia, PA

Functional and morphological architectures of major human organs have been well characterized using imaging biomarkers. Nevertheless, deciphering the causal relationships between imaging biomarkers and major clinical outcomes, as well as understanding the causal interplay across multiple organs, remains a formidable challenge. Mendelian randomization (MR) presents a framework for inferring causality by using genetic variants as instrumental variables. Here we report a systematic multi-organ MR analysis between 402 imaging biomarkers and 88 clinical outcomes. We identified 488 genetic causal links for 62 diseases and 130 imaging biomarkers from 9 organs, tissue, or systems, including the brain, heart, liver, kidney, lung, pancreas, spleen, adipose tissue, and skeleton system. We prioritized crucial intra-organ causal connections, such as the bidirectional genetic links between Alzheimer's disease and brain function, as well as inter-organ causal effects, such as the adverse impact of heart diseases on brain health. Our findings uncover the genetic causal links spanning multiple organs, offering a more profound understanding of the intricate relationships between organ imaging biomarkers and clinical outcomes. We have made our database publicly available and developed a browser framework to facilitate the exploration of MR findings of multi-organ images (http://mr4mo.org/).

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1683 Multiple Viruses Detected in Human DNA are Associated with Alzheimer Disease Risk

Authors:

M. Tejeda[1], J. Farrell[1], Z. Congcong[1], L. Wetzler[1], K. Lunetta[2], W. S. Bush[3], E. R. Martin[4], W. Li-San[5], G. Schellenberg[5], M. Pericak-Vance[4], J. Haines[3], L. Farrer[1], R. Sherva[1]; [1]Boston Univ. Sch. of Med., Boston, MA, [2]Boston Univ SPH Crosstown Ctr, Rm 313, Boston, MA, [3]Dept. of Population & Quantitative Hlth.Sci., Cleveland Inst. for Computational Biology, Case Western Reserve Univ. Sch. of Med., Cleveland, OH, [4]John P. Hussman Inst. for Human Genomics and Dr. John T. MacDonald Fndn. Dept. of Human Genetics, Miller Sch. of Med., Univ. of Miami, Miami, FL, [5]Penn Neurodegeneration Genomics Ctr., Dept. of Pathology and Lab. Med., Univ. of Pennsylvania Perelman Sch. of Med., Philadelphia, PA

Background: Multiple infectious agents have been identified as risk factors for Alzheimer's disease (AD). Viruses can cause inflammatory damage and potentially increase the formation of the AD hallmark proteins amyloid-β protein and hyperphosphorylated tau. We and others previously reported association between AD and the quantity of whole genome and whole exome sequence reads mapping to herpes simplex type 1 (HSV-1), human herpes virus 6B (HHV-6B) and human papilloma virus 71 (HPV-71). Here, we conducted GWAS using these viruses to identify genetic associations that 1) affect viral load and 2) modify the effect of viral load on AD risk. **Methods:** DNA sequence reads that did not align to the human genome in whole exome and whole genome sequences generated by the Alzheimer Disease Sequencing Project from 57,000 AD cases and controls from multiple populations (European ancestry (EA), Caribbean Hispanic (CH), and African American) were mapped to 318 viral reference sequences and quantified. After quality control, there were 4,298 sequences derived from brain (AD cases= 3,404, controls= 894) and 40,156 sequences derived from blood (AD cases=15,612, controls=24,544). Two GWAS models for each virus were tested: 1) viral load ~ SNP + sex + age + tissue source and 2) AD ~ SNP + viral load + SNP*viral load + sex + age + tissue source. **Results**: We observed genome-wide significant (GWS) associations between HSV-1 presence and *HLA-DQA1* SNP rs28654242 (OR=3.01, P=2.15x10^{-8}) with HPV-71, *POTEE* SNP rs199606922 (OR=2.03,P=1.91x10^{-8}), and *USP49* SNP rs201585090 (OR=0.26, P=1.28x10^{-8}) , and between HHV-6B and *CASS4* SNP rs6014724 (Beta=5.52, P=1.26x10^{-4})and rs16941239 near *FOXF1* (Beta=6.93, P=5.95x10^{-4}), all in CHs. The *CASS4* and *FOXF1* SNPs were previously associated with AD risk. We also observed, in EAs, GWS interactions between HSV-1 presence and rs201217710 in *ABCC2* (OR= 0.15, P=3.28x10^{-8}), rs113432200 in *KMT2C* (OR=0.54, P=2.11x10^{-8}), and rs62345265 in *FRG1* (OR=0.65, P=5.56x10^{-11}) predicting AD risk. **Conclusions:** These findings provide further evidence for the role of herpes viruses in the pathogenesis of AD and highlight the importance of genetic factors in modulating these associations.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1684 Multi-trait analysis characterizes the genetics of thyroid function and identifies causal associations with clinical implications

Authors:

A. Teumer[1,2], R. B. T. Sterenborg[3,4], I. Steinbrenner[5], Y. Li[5], M. N. Bujnis[6], T. Naito[7], E. Marouli[8], M. Medici[4], ThyroidOmics Consortium; [1]Univ. Med. Greifswald, Greifswald, Germany, [2]DZHK (German Ctr. for Cardiovascular Res.), partner site Greifswald, Greifswald, Germany, [3]Dept. of Internal Med. Radboud Univ. Med. center, Nijmegen, Netherlands, [4]Academic Ctr. for Thyroid Diseases, Erasmus Med. Ctr., Rotterdam, Netherlands, [5]Faculty of Med. and Med. Ctr., Univ. of Freiburg, Freiburg, Germany, [6]The Univ. of Utah, Salt Lake City, UT, [7]Osaka Univ. Graduate Sch. of Med., Suita, Japan, [8]William Harvey Res. Inst., Barts and The London Sch. of Med. and Dentistry, Queen Mary Univ. of London, London, United Kingdom

Thyroid function tests based on thyrotropin (TSH) are among the most frequently ordered biochemical tests for assessing thyroid dysfunction, a common disorder with a prevalence of ~5% in the general population. Thyroid hormones including thyroxine and triiodothyronine (T3) play a key role in cellular growth, development and metabolism. Genetic factors are responsible for up to 58-71% of the variation in TSH and free thyroxine (FT4). We conducted a genome-wide association study meta-analysis of thyroid function in up to 271,040 individuals of European ancestry, including reference range TSH, FT4, free and total T3, and proxies for metabolism (T3/FT4 ratio). Secondary analyses included colocalization with mRNA levels using GTEx, polygenic risk score analysis on clinical endpoints, and Mendelian randomization on thyroid function-related outcomes including thyroid cancer. We revealed 198 loci associated with TSH (84 novel), 84 loci associated with FT4 (45 novel), and 29 novel loci for the T3 related traits. The loci explained 14.1%, 6.0%, 10.9% and 1.1% of the variation in TSH, FT4, total and free T3 concentrations, respectively. Genetic correlations strongly suggest that TSH associated loci reflect the thyroid function determined by the active thyroid hormone T3, whereas FT4 as well as free and total T3 loci reflect the thyroid hormone metabolism. Colocalization with FT4 revealed genes expressed in various peripheral tissues, where TSH-associated genes were predominantly expressed in the thyroid including multiple genes with a known role in the TSH signaling cascade and thyroid hormone synthesis, such as *PDE8B, PDE10A, TPO* and *GLIS3*. Polygenic risk score and Mendelian randomization analyses showed the effects of genetically determined variation in thyroid function on various clinical outcomes, including cardiovascular risk factors and diseases, autoimmune diseases, and cancer.

Our study increased the number of loci discovered for the most important thyroid function parameters, improved our understanding of the genes altering mRNA expression in different tissues and their contributing effects in various pathways influenced by thyroid function parameters. Furthermore, it provides a comprehensive overview of the effects of genetically determined variation in thyroid function on many known and newly suggested clinical outcomes. This could foster possibilities for using genetics in prevention and diagnosis, and to identify candidates for therapeutic targets to reduce the burden of thyroid diseases.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1685 Multi-trait analysis of GWAS identifies 52 novel genomic loci associated with circulating polyunsaturated fatty acids

Authors:

H. Xu[1], J. Xue[1,2], Y. Sun[3], K. Ye[1,4]; [1]Dept. of Genetics, Univ. of Georgia, Athens, GA, [2]Coll. of Life Sci., Wuhan Univ., Wuhan, China, [3]Dept. of Genetics, Univ. of Goergia, Athens, GA, [4]Inst. of Bioinformatics, Univ. of Georgia, Athens, GA

BACKGROUND: Polyunsaturated fatty acids (PUFAs) play crucial roles in a wide variety of physiological processes and have been implicated in the development and prevention of various diseases, such as cardiovascular diseases and psychiatric disorders. Despite the identification of hundreds of genomic loci for circulating PUFA (cPUFA) levels through single-trait genome-wide association studies (GWAS), these loci account for only a small fraction of cPUFA variance. **METHODS:** To address this limitation, we employed a multi-trait analysis of GWAS (MTAG) by leveraging GWAS summary statistics of 11 cPUFAs (N=114,999) from UK Biobank. The 11 PUFA traits include total PUFAs, omega-3, omega-6, docosahexaenoic acids (DHA), linoleic acid (LA), and their relative percentages in total fatty acids. To prioritize genes related to cPUFAs and gain insights into the biological roles of significant loci, we performed functional annotation and gene set enrichment analysis using FUMA. Additionally, we conducted a colocalization analysis of MTAG summary statistics with HyPrColoc to identify genomic regions and potential causal variants shared among cPUFA traits. **RESULTS:** Our MTAG analyses identified 420 genomic loci that are significantly associated with 11 cPUFAs, including 52 novel discoveries. Functional annotation and gene set enrichment analysis identified 1,947 candidate genes, which are enriched in lipid-related pathways ($P < 3.2\times10^{-6}$) and highly expressed in the liver (FDR adjusted $P < 0.05$). Furthermore, our pairwise colocalization analysis detected 2,054 genomic regions (posterior probability > 0.7) colocalized among multiple cPUFAs. We identified 540 unique SNPs that could potentially serve as shared causal variants, indicating a pleiotropic effect of certain loci on multiple cPUFA traits. The highest number of shared loci was observed between omega-6 and PUFA, totaling 124, whereas the omega-6 to omega-3 ratio and PUFA% exhibited the fewest shared loci, with only 9. Notably, a few pairs, such as DHA% and PUFA, had 5 colocalized regions with the same effect directions and 6 with opposite directions. This finding helps explain the observed low genetic correlation between these traits, as the genetic effects canceled each other out when estimating the overall correlation across the genome. **CONCLUSION:** Our multi-trait GWAS identifies 52 novel genomic loci associated with 11 circulating levels of PUFAs. These findings enhance our understanding of the genetic basis of PUFAs and pave the way for further investigation into their associations with other diseases.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1686 Multi-trait investigation of inflammation-associated DNA methylation among people with HIV

Authors:

J. Chen[1], Q. Hui[2], B. Titanji[1], K. So-Armah[3], M. Freiberg[4], A. Justice[5], K. Xu[6], X. Zhu[7], M. Gwinn[1], V. Marconi[1], Y. Sun[2]; [1]Emory Univ., Decatur, GA, [2]Emory Univ., Atlanta, GA, [3]Boston Univ., Boston, MA, [4]Vanderbilt Univ., Nashville, TN, [5]Yale Univ., New Haven, CT, [6]Yale Sch. Med., New Haven, CT, [7]Case Western Reserve Univ., Shaker Heights, OH

Inflammation is one of the primary causes for the excess morbidity and mortality among people with HIV (PWH). A handful of epigenome-wide association studies (EWAS) have suggested that inflammation is associated with DNA methylation (DNAm) among PWH. Although many inflammatory markers may share a common physiological pathway, current EWAS usually study one trait at a time. Multi-trait EWAS that analyze multiple inflammatory markers can further improve statistical power in detecting DNAm and help provide an interpretation of shared epigenetic architecture between inflammatory markers. We conducted a multi-trait EWAS investigation of three inflammatory markers including soluble CD14, D-dimers, and interleukin 6 in a subset of PWH in the Veteran Aging Cohort Study that collected biomarker data 2005-2007 (n = 920). The study population are all males with an average age of 52 years old, and 85% self-reported as Black. EWAS summary statistics for individual markers were obtained by linear mixed models using DNA methylation as the outcome, inflammatory marker as the exposure, and chip ID as random effect to adjust for batch effect. All models were adjusted for concurrent age, race/ethnicity, viral load, antiretroviral therapy, smoking, BMI, diabetes, hepatitis B/C viral infections, and cell type proportions. We applied both methods CPASSOC and OmniTest to EWAS summary statistics for individual markers. CPASSOC is a meta-analysis method accounting for the correlations induced by correlated traits and overlapping samples. OmniTest is an aggregated Cauchy association test that combine p-values from multiple tests including the minimum of the p-values approach, generalized Berk-Jones test and the generalized higher criticism test. CPASSOC and OmniTest identified 189 and 157 inflammation-associated DNAm sites including 112 overlapping DNAm sites. Among the total 234 DNAm sites identified, 132 had not been reported as significant in any single-marker EWAS. We showed that CPASSOC tends to identify DNAm sites that are associated with all 3 inflammatory markers while OmniTest prioritizes strong signals from single-marker EWAS. Top DNAm sites were mapped to inflammation-related genes such as *IFITM1*, *PARP9*, *NLRC5*, *PSMB8* and *STAT1*. These genes were significantly enriched in pathways such as “type I interferon signaling” and “immune response to virus”. In conclusion, our results show that multi-trait EWAS can increase statistical power to identify inflammation-associated DNAm sites, genes, and pathways. These DNAm sites suggest molecular mechanisms underlying the accelerated aging and the elevated chronic disease burden observed among PWH.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1687 Mutations in Cardiomyopathy Genes are Enriched in Pediatric Postural Orthostatic Tachycardia Syndrome

Authors:

H-Q. Qu[1], J. Glessner[1], X. Chang[1], F. Mentch[1], Y. Liu[1], M. Bakay[1], J. Boris[2], H. Hakonarson[1,3]; [1]Children's Hosp. of Philadelphia, Philadelphia, PA, [2]Jeffrey R. Boris, MD LLC., Philadelphia, PA, [3]The Perelman Sch. of Med., Univ. of Pennsylvania, Philadelphia, PA

Introduction: Postural orthostatic tachycardia syndrome (POTS) is a severe disorder with significant functional impairment and psychological distress to the patients, warranting development of new and effective treatments. However, the etiology of POTS is poorly understood, and has been largely regarded as a form of dysautonomia. **Hypothesis:** POTS is a complex genetic disease requiring integrative sequencing and translational approaches to identify its genetic underpinning and better understand its pathogenesis with the aim at promoting the development of novel precision-based therapies. **Methods:** There are numerous chronic health conditions in adults associated with the risk of POTS. We studied a unique collection of pediatric POTS subjects including 87 unrelated European cases (61 females and 26 males, 6~21 years old). We adopted exome sequencing to identify pathogenic (P) or likely pathogenic (LP) mutations in these POTS patients with POTS patients. **Results:** Whole exome sequencing of the 87 patients resulted in 8 patients identified each with a heterozygous LP mutation (classified by InterVar) involving a cardiomyopathy gene, including 6 titin (*TTN*, related to dilated cardiomyopathy) mutations (exon188 R26587Q, E26494Q; exon154 I19603V, S19023N, S15470Y; and exon46 K4455R), 1 ryanodine receptor 2 (*RYR2*, related to catecholaminergic polymorphic ventricular tachycardia) mutation (exon21 S756N), and 1 myosin heavy chain 7 (*MYH7*, related to hypertrophic cardiomyopathy) mutation (exon21 R787C). Compared to the frequencies in the Non-Finnish European (gnomAD v3.1) dataset, the POTS LP mutations are significantly enriched in the *TTN* gene (Fisher's exact *P*=1.00E-05), as well as in the three cardiomyopathy genes (Fisher's exact *P*=6.98E-08). In addition to the 3 genes with LP mutations, we also identified rare coding variants with predicted deleterious effects in a number of other genes related to cardiomyopathy in the POTS cases, demonstrating significant *P* value by over-representation analysis (ORA) of gene sets, including some drug target genes. The LP variants of interest were functionally characterized in both cell based assays and *in silico*. **Conclusion:** Our results suggest, for the first time, that POTS may be a clinical manifestation of non-penetrant mutations in cardiomyopathy genes. Further study is warranted to clarify how these mutations contribute to atypical structural or functional changes of the heart in these patients. Several of these mutation profiles involve drug-gene targets and may help promote new therapeutic development for this disabling disorder.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1688 New Histological Approach in Spatial Transcriptomics Implicates Glandular Cell Involvement in Pathophysiology of Sjögren's Disease

Authors:

C. Lessard[1], P. Czarnota[1], R. Wilbrink[2], B. Khatri[1], A. M. Stolarczyk[1], C. Pritchett Frazee[1], C. Li[1], M. C. Marlin[1], K. Wright[3], K. L. Tessneer[1], L. Radfar[4], J. A. James[1], R. H. Scofield[1], I. Adrianto[5], A. Rasmussen[1], J. M. Guthridge[1], A. D. Farris[1]; [1]Oklahoma Med. Res. Fndn., Oklahoma City, OK, [2]Univ. Med. Ctr. Groningen, Groningen, Netherlands, [3]Univ. of Oklahoma Hlth.Sci. Ctr., Oklahoma City, OK, [4]Univ. of Oklahoma Coll. of Dentistry, Oklahoma City, OK, [5]Henry Ford Hlth., Detroit, MI

Background: 10X Visium spatial transcriptomics evaluates gene expression in a 50um tile coordinate of a sectioned tissue, yielding heterogeneous cell sampling. Spatial PCA algorithm used homogenous tissue types with distinct boundaries to identify like tissue regions and determine the cellular context of spatial coordinates [1]. It is less effective at differentiating like tiles from heterogenous tissue types such as the Sjögren's disease (SjD) target tissue, minor salivary gland (MSG). This study introduces a novel algorithm, HistoPCA, that connects pathologically annotated and segmented images with spatial coordinates to group like tiles based on similar cellularity.
Methods: Histopathological images of minor salivary gland biopsies were annotated by tissue type (fibrosis, glandular, inflammatory, fat). Annotations were combined with spatial coordinates using HistoPCA. Image segmentation was used to obtain cell numbers in tiles. Segmentation data was combined with pseudo-color percentages associated with certain tissue types to perform unsupervised KMeans clustering. SjD case-control differential expression (DE) was analyzed using pseudo-bulk gene expression, then DE transcripts were analyzed by Ingenuity Pathway Analysis.
Results: HistoPCA and UMAP with KMeans clustering outperformed Spatial PCA (Adjusted Rand Index (ARI) of 0.52 vs. 0.51), identifying four distinct clusters in MSG. Despite distinct features like inflammation in Cluster 1 (21% inflammatory, 32% fibrosis, 23% glandular tiles) and increased glandular involvement in Cluster 3 (0% inflammatory, 48% fibrosis, 41% glandular tiles), DE and pathway analyses showed similar patterns of dysregulation in SjD cases (n=28) compared to controls (n=16). As a positive control, elevated interferon signatures were observed in all SjD cases and anti-Ro positive SjD cases (n=19). Other immune pathways (T cell receptor, Th1, Th2, and IL-4 signaling, etc.) were also upregulated in SjD cases while CTLA4, IL-10, and IL-12 signaling pathways were downregulated.
Conclusions: HistoPCA successfully grouped like tiles from spatial transcriptomic analysis of heterogeneous MSG. Cluster annotation of HistoPCs, followed by DE and pathway analyses identified four clusters with dysregulated immune pathways in SjD. Dysregulated immune pathways in Cluster 3 (devoid of tiles associated with inflammation) implicate glandular involvement in SjD pathogenesis. Spatially aware technology will provide new insights into the role of different cell/tissue types in SjD pathobiology of the salivary gland.
References:[1] Shang L, et al. Nat Commun. (2022); 13:7203

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1689 New Insights into Genetic Factors Associated with Helicobacter pylori Infection: A Large-Scale GWAS in Japan

Authors:

T. KYOSAKA[1], A. Narita[2], G. Tamiya[1,2]; [1]Tohoku Univ. Sch. of Med. AI and Innovative Med., Sendai, Japan, [2]Tohoku Univ. Tohoku Med. Megabank Organization, Sendai, Japan

Background & Aims

Preventing *Helicobacter pylori* (*H. pylori*) infection is crucial for reducing the risk of gastric cancer. Although susceptibility to *H. pylori* is known to be influenced by genetic factors, large-scale genome-wide association studies (GWAS) have yet to be conducted in Asia, where infection rates are relatively high. In this study, we conducted a large-scale GWAS on a Japanese cohort with the aim of elucidating the genetic background associated with *H. pylori* infection.

Methods

We analysed data from 100,362 individuals from the Tohoku Medical Megabank (TMM) Project database, which included both genomic information and serum anti-*H.pylori* IgG antibody titres. The top 25% of individuals with the highest IgG antibody titres were defined as cases, with the remaining 75% serving as controls, for the purposes of conducting GWAS and fine mapping.

Results

Our GWAS identified several new loci associated with *H. pylori* infection. The most strongly associated variant was rs9273002 (P = 3.32e-27), located downstream of the HLA-DQA1 gene on chromosome 6 at position 6p21.32. Fine mapping revealed significant associations with rs13194769 (P = 1.13e-39), located near HLA-DRB9, and rs1130399 (P = 5.07e-12), situated in the exon region of HLA-DQB1. Alongside these novel loci, significant associations were also observed with variants at loci 8q24.3 (P = 4.94e-8), which have previously been implicated in gastric cancer.

Conclusions

Our large-scale GWAS has revealed new genetic loci associated with *H. pylori* infection. Notably, we found a strong association within the HLA region; these genetic variations may partially explain individual variability in risk for *H. pylori* infection. Future work will involve conducting a cross-ancestry meta-analysis with existing studies on populations of European ancestry to explore the genetic factors associated with *H. pylori* infection in different populations.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1690 New phenotyping approaches are essential to advance genetic discovery for psychiatric disorders.

Authors:

J. Thorp, Z. Gerring, E. Derks; QIMR Berghofer Med. Res. Inst., Brisbane, Australia

Genome-wide association studies have revolutionised the study of genetic factors underlying psychiatric disease and led to the identification of hundreds of robustly associated genetic variants for many disorders. There are, however, several characteristics of psychiatric disorders that complicate genetic studies and the complete understanding of the genetic basis of psychopathology. I will discuss two core themes in psychiatric genetics: pleiotropy and heterogeneity.

Pleiotropy between psychiatric disorders is widespread and pervasive. I will present some of our recent work that has found extensive genetic overlap between disorders at genome-wide, regional, and transcriptomic levels. Furthermore, I will discuss work that has investigated whether the extreme clinical heterogeneity observed in patients is recapitulated at a genetic level, and present recent evidence of genetic heterogeneity in patients with major depressive disorder.

The substantial heterogeneity within and pleiotropy between psychiatric disorders challenges the utility of studying psychopathology as categorical and distinct disorders that are conceptualized in currently used classification systems such as the DSM. There is developing interest in alternative phenotyping / classification systems of psychopathology that overcome limitations of current systems. I will discuss such phenotyping approaches and present our recent research showing that implementation of these systems into genetic studies of psychiatric disorders can offer major benefits and provide novel insight into the genetic basis of psychopathology.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1691 NMR-based analyses identify serum metabolites associated with acute stroke using patients from the MISS and UK biobank

Authors:

D. Sanghera[1], M. Rout[2], A. Vaughan[3], A. Blair[3], S. Stavrakis[3], E. Sidorov[3]; [1]Univ. of Oklahoma Hlth.Sci. Ctr., Edmond, OK, [2]OUHSC, Oklahoma City, OK, [3]Univ. of Oklahoma Hlth.Sci. Ctr., Oklahoma, OK

Stroke is a significant health issue in the United States, and identifying biomarkers for the prevention and functional recovery after an acute stroke remains the highest priority. This study aims to identify circulating metabolite signatures associated with stroke pathophysiology by performing discovery and validation studies. We performed targeted metabolomics profiling of 420 participants of the discovery dataset of Metabolome in an Ischemic Stroke Study (MISS) using high-throughput nuclear magnetic resonance (NMR) spectroscopy. A validation study of significantly altered metabolites was conducted using an independent cohort of 117,988 participants from the UK Biobank, available with metabolomics profiles using the same NMR technology. Our study identified 13 metabolites significantly perturbed during acute stroke. Amino acid phenylalanine was significantly increased, while the circulating levels of glutamine and histidine were significantly decreased in stroke. Serum levels of apolipoprotein A-1, several essential fatty acids, and phosphatidylcholine were reduced, while ketone bodies like 3-hydroxybutyrate and acetoacetate were markedly increased in stroke. Our study demonstrates the potential of high throughput metabolomic technology to unravel metabolic dysregulation in stroke not captured by traditional clinical measurements. Further work will be required to ascertain whether these markers have clinical utility. **Funding:** This work was funded by the College of Medicine Alumni Association (COMAA), the Presbyterian Health Foundation, and the Leinbach Foundation grants, and from Dr. Geoffrey Altshuler Endowment funds from the Children's Health Foundation of the University of Oklahoma Health Sciences Center.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1692 Non-coding regulation of mitochondrial copy number in humans.

Authors:

T. Hall, S. Cannon, G. Hawkes, M. Weedon, K. Patel; Univ. of Exeter, Exeter, United Kingdom

Introduction:
Mitochondrial DNA copy number (mtDNA-CN) alterations are associated with common, complex and rare genetic disorders and can be regulated by nuclear genome variation. Previous studies focused on rare or common nuclear coding variants affecting mtDNA-CN. Using whole-genome-sequence (WGS) data on 199,940 individuals, we aimed to identify rare, non-coding variants' contribution to the nuclear genetic regulation of mtDNA-CN.
Method:
We performed the largest WGS-based analysis for mtDNA-CN to-date using 199,940 individuals from UK Biobank. We used mitoHPC to derive mtDNA-CN and performed association testing, with extensive adjustment for blood cell measurements and technical covariates, on 143,114 individuals passing quality control. We tested 75,311,546 single variants which each had ≥20 carriers in the UK Biobank, and performed 52,749,161 rare variant (MAF <0.1%) genomic aggregate tests split into gene-centric (e.g. proximal-regulatory), non-gene-centric (e.g. intergenic-regulatory), and grouped by measures of conservation, constraint and deleteriousness.
Results:
We identified two aggregated rare coding variants in *JAK2* and *SIRPA* that reduced mtDNA-CN by -0.7 and 0.17 SD at genome wide significance ($p<1x10^{-7}$) respectively. We also identified 19 aggregated proximal rare-noncoding variants units associated with mtDNA-CN (beta 0.08 to -0.6 SD) at genome wide significance. Of these, the proximal variants to *USF2* had the largest effect.
In 12 known genes associated with rare mitochondrial depletion syndrome (lower mtDNA-CN), we identified 9 aggregate non-coding variant units proximal to *TFAM* at Bonferroni corrected significance $p<6x10^{-6}$. The strongest signal was an upstream proximal aggregate unit ($p=3x10^{-6}$) which was driven by a novel, rare, promoter variant associated with an almost 2SD reduction in mtDNA-CN; chr10:58385409C>A ($MAF=3.49x10^{-5}$, beta=-1.9SD [95% CI: -1.3, -2.5], $p<8x10^{-10}$).
Conclusion:
Our WGS based approach identified novel coding and non-coding associations for mitochondrial copy number. Our study highlights the benefit of using WGS to study mitochondrial genetics in humans and opens a next phase of genetic studies to understand the non-coding regulation of mtDNA-CN.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1693 Novel modifiers of age-at-onset for Alzheimer's disease identified in African American datasets as well as meta-analyses across populations.

Authors:

K. Nuytemans[1], S. Guo[1], J. La[2], A. Naj[3], X. Gao[4], Y-J. Li[2], E. Martin[1], Alzheimer Disease Genetics Consortium (ADGC); [1]Univ. of Miami, Miami, FL, [2]Duke Univ., Durham, NC, [3]Univ. of Pennsylvania, Philadelphia, PA, [4]The Ohio State Univ., Columbus, OH

Background: Several studies have attempted to identify genes for age-at-onset (AAO) of Alzheimer's disease (AD) through genome-wide linkage and genome-wide association studies (GWAS). While these efforts thus far have focused predominantly on non-Hispanic white (NHW) populations, we present analyses identifying genetic contributors to AAO in African Americans (AA) as well as across population groups using genotype and phenotype data from the Alzheimer's Disease Genetics Consortium (ADGC). **Methods:** We combined genotype data from 13 AA prospective and case-control ADGC datasets, excluding family-based datasets, performed QC and imputed to the TOPMed R2 reference panel using the Michigan Imputation Server. Data were merged across cohorts using QCTOOL and GTOOL, filtering out variants with MAF≤0.005. We performed two analyses: case-only linear mixed-model regression with AAO as outcome (LMM); and survival analysis using Cox proportional-hazards frailty modeling (COXPHF) with time-to-AD-onset (AAO in cases and censoring of controls at age-at-last-exam). Both models were adjusted for sex, principal components capturing population substructure, and *APOE*-ε4 dosage with a random cohort effect. Second, we performed meta-analyses across AA and previously reported NHW results using MetaSoft2.0.1's Han and Eskin's random effects model. **Results:** We examined 2,262 AA AD cases (69% female) with mean (SD) AAO of 77.7 (8.1) years. The *APOE*-adjusted COXPHF analyses identified seven genome-wide significant regions ($P<5\times10^{-8}$), of which five were driven by African-specific variants (MAF<0.0002 in gnomAD Non-Finish Europeans) and the most significant one is *ABCA7* ($P=3.57\times10^{-10}$), known to be associated with AD risk in AAs. In the LMM analysis, we observed strong novel associations with variants in *CCDC141* on chr2q31.2 ($P_{min}=2.93\times10^{-9}$). *In silico* analyses of LD structure and functional potential suggest *SESTD1*, a gene involved in synaptic regulation, as driving the *CCDC141* locus signal. Meta-analyses combining results from AA dataset and NHW ADGC cohorts (9,220 AD cases/10,350 controls; Li et al. *Alz Dement* 2023) identified *BIN1*, *CR1*, *ABCA7* and *CLSTN2* in the COXPHF analyses as well as loci on chr2q36.3 and chr11p15.4 in the LMM analyses (p<5x10-8 overall, p<0.001 in each dataset). **Conclusion:** In a large ADGC African American dataset, we confirmed AAO associations in the *APOE* region and *ABCA7* and observed evidence for potential novel AD AAO loci including *SESTD1*; many of which might be African-ancestry specific. Meta-analyses across population groups further identified known AD genes as well as two novel loci to be involved in AAO overall.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1694 Novel variants identified in black South African families with vitiligo

Authors:

T. Ngcungcu, B. Nonyane, M. Ramsay; Univ. of the Witwatersrand, Johannesburg, Gauteng, South Africa

Vitiligo, a skin pigmentation disorder characterized by acquired depigmentation and melanocyte deficiency, affects 0.4-2% of the general population across all ethnic groups. Environmental factors play a role in triggering the development of the disorder, but genetic susceptibility is important as family members have an increased risk for developing vitiligo. The occurrence of vitiligo in multiplex families (more than one affected family member) supports the role of genetic variants in the aetiology of vitiligo and the pattern of inheritance suggests that it is polygenic and multifactorial. Published genetic association studies have been done in European and Asians populations, but none in continental African populations. The aim was to identify genetic variants in families of African descent with at least two individuals affected by vitiligo. Three families (two affected members per family and one unaffected family member; n=7) were recruited. Whole exome sequencing was done from libraries prepared using the Ion AmpliSeq™ Exome RDY library 4x2 kit and sequenced using the Ion S5™ sequencer. Three novel variants in the *FLG*, *TSPYL* and *MANF* genes were identified and were present exclusively in the affected participants. These results require validation and family segregation analyses to understand their contribution as large effect variants to the vitiligo phenotype. None of the identified genes have previously been implicated in vitiligo pathogenesis.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1695 Older adults with Alzheimer's disease had a higher burden of infections diagnosed before AD, and it was influenced by sex, race, and *APOE4* status

Authors:

S. Ukraintseva, R. Holmes, H. Duan, D. Wu, A. Yashkin, I. Akushevich, A. Yashin, K. Arbeev; Duke Univ., Durham, NC

Introduction and Method: Infectious diseases may play a role in Alzheimer's Disease (AD) development, though exact mechanism is unclear. Genetic and other risk factors for AD may interact with infections and modulate their relations with AD. Here we compared frequencies of moderate to severe infections in medical history (based on hospitalization ICD10 codes) of UK Biobank participants with and without AD, stratifying these analyses by AD risk factors, such as sex, race, and *APOE4* carrier status. The latter was based on the information about rs429358 and rs7412 polymorphisms. To avoid reverse causation, we only considered infections diagnosed before age 72, and AD diagnosed at ages 73 or later. These age cut-offs were chosen based on the age distributions of AD and infectious diseases in the data, and they allowed us to include majority of cases of both AD and infections in the analysis. We used the following binary variable for infections: *(0) Infection occurred before age 72 vs. (1) no infection before age 72, and both groups survived age 72*. SAS procedure PROC FREQ and Chi-square test were implemented to compare significance between groups. Bonferroni correction (P_{bon}) was also calculated. **Results:** UKB participants who developed AD at ages 73+ had a higher proportion of infectious diseases diagnosed before age 72 compared to those without AD, in the total sample (P_{bon} =1.12E-84, N=42503, AD=64.18%, noAD=42.33%), as well as in females (P_{bon} =1.86E-38, N=20821, AD=60.78%, noAD=40.45%), males (P_{bon} =9.60E-47, N=21682, AD=67.86%, noAD=44.32%), Whites (P_{bon} =3.07E-81, N=41032, AD=64.14%, noAD=42.30%), and Blacks (P_{bon} =0.0121, N=330, AD=74.19%, noAD=39.72%). The difference in the proportion of infections among AD vs noAD was larger and more significant in non-carriers of *APOE4*, compared to carriers: *APOE4* carriers (P_{bon} =7.47E-29, N=8900, AD=61.87%, noAD=41.47%); non-carriers (P_{bon} =3.58E-38, N=24834, AD=68.72%, noAD=42.19%). **Conclusion:** Burden of prior infections was higher in UKB participants who developed AD later in life than in those without AD, in line with the idea that adult infections may contribute to AD development. The higher proportion of infections preceding AD was seen in males, Blacks, and non-carriers of *APOE4*, the latter suggesting that the impact of *APOE4* on AD risk may overweigh that of infections.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1696 Ongoing Natural History study in Phosphomannomutase 2 Congenital Disorder of Glycosylation (PMM2-CDG): Clinical and Basic Investigations

Authors:

T. Moors[1], E. Morava[2], M. Patterson[2], A. Cechova[3], T. Honzik[3], A. C. Edmondson[4], C. Ficicioglu[4], M. G. Serrano[5], R. Barone[6], P. Witters[7], D. Ryman[7], C. Lam[8], M. C. H. Janssen[9], E. Martins[10], D. Quelhas[10], J. Sykut-Cegielska[11], M. Schiff[12], P. De Lonlay[12]; [1]Glycomine, Inc., San Mateo, CA, [2]Mayo Clinic, Rochester, MN, [3]Charles Univ. and Gen. Univ. Hosp., Prague, Czech Republic, [4]Children's Hosp. of Philadelphia, Philadelphia, PA, [5]Hosp. Sant Joan de Deu, Barcelona, Spain, [6]Univ. of Catania, Catania, Italy, [7]Univ. Hosp. Leuven, Leuven, Belgium, [8], Univ. of Washington Sch. of Med., Seattle, WA, [9]Radboud Univ. Med. Ctr., Nijmegen, Netherlands, [10]Centro Hosp.ar Universitário do Porto, Porto, Portugal, [11]Inst. of Mother and Child, Warsaw, Poland, [12]Necker Hosp., Paris, France

Background: Phosphomannomutase 2 congenital disorder of glycosylation (PMM2-CDG) is the most common CDG, yet there are no prospective studies to provide a cohesive natural history. This is a systematic clinical study to capture data across multiple domains to better understand the course of disease. Clinical presentation can be variable, ranging from infants who die in the first weeks of life to mildly affected adults. Apart from neurologic manifestations patients have variable, sometimes severe, immunologic, hematologic, cardiac, pulmonary, gastrointestinal, renal, endocrine, musculoskeletal, developmental and/or ophthalmologic involvement. This global, collaborative study represents the largest and longest set of natural history longitudinal data collected in PMM2-CDG.**Methods**: PMM2-CDG subjects were enrolled based on enzymatic and/or molecular tests. Medical history, physical examination, laboratory testing and imaging studies were performed at baseline and approximately every 6 months, depending on standard of care at the investigator's institution as well as the clinical status of the individual patient. **Results:** 139 subjects enrolled across 11 sites in 9 countries (US & Europe). Mean age at enrollment was 14.7y (range 0.5 to 68.4y), and almost 70% were under 18y. Enrolled subjects had 60 unique genetic variants, which included 6 novel variants not previously reported. Subjects were 15 times more likely to have weight Z-score below -2, reflecting faltering growth across all ages. Laboratory abnormalities included low coagulation parameters (ATIII, Factor XI, Protein C) in most subjects and elevated liver enzymes (2-10x ULN) in children. Nijmegen Progression CDG Rating Scale (NPCRS) data collected after 1 year on study suggests clinical complications arising from coagulation defects occurring in 10% of subjects in the previous 6 months. As of February 2023, 7 (5%) subjects had completed 4 years on study and 54 (39%) 2 years. **Discussion:** This ongoing natural history study provides an observational data set to understand clinical manifestations, morbidity, and disease progression of PMM2-CDG.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1697 Opioid Overdose Death and Gene Dysregulation: A Transcriptome-wide Association Study in Nucleus Accumbens

Authors:

C. Willis[1], B. Quach[1], O. Corradin[2], R. Sallari[3], D. Hancock[1], D. Mash[4], S. Akbarian[5], E. Johnson[1]; [1]RTI Intl., Research Triangle Park, NC, [2]MIT, Cambridge, MA, [3]Axiotl Inc, Cleveland, OH, [4]Nova Southeastern Univ., Fort Lauderdale, FL, [5]Icahn Sch. of Med. at Mount Sinai, New York, NY

The opioid epidemic in the US continues unabated, with 5.6 million people having an opioid use disorder and over 80,000 opioid overdose deaths (OODs) occurring in 2021. Our neurobiological understanding of opioid addiction remains limited. There has been one previous study of differential gene expression associated with opioid overdose death in human nucleus accumbens (NAc), a key brain region for studying addiction. Here, we analyze a new cohort of NAc RNAseq samples, meta-analyze results with the previous study, and compare results with prefrontal cortex, another key brain region for addiction. Bulk RNAseq was performed on NAc tissue collected from 92 decedents (50 OOD cases, 42 controls) of European (64) and African American (28) ancestry. Samples were analyzed separately in two batches then meta-analyzed using the weighted fisher's method. Genes were tested for differential expression by OOD status using negative binomial regression models controlling for plate effects and unsupervised surrogate variables as covariates that accounted for known and unknown covariates. Of 18,867 genes that were meta-analyzed across batches, 125 showed significant differential expression (FDR < 0.05). We replicated 4 genes (*ATG4D*, *SLC10A3*, *SCAMP4*,*SLC19A1*) in a previously published study's results (Seney et al., Biological Psychiatry 2021), after applying Bonferroni correction ($P < 4x10^{-4}$). Meta-analysis of the two independent cohorts resulted in 96 significant genes at FDR < 0.05. These 96 genes were significantly enriched (FDR < 0.05) for gene ontology terms and pathways related to immune function (such as IL2, IL6, and TCR pathways) and calcium ion channel function (such as regulation of calcium ion export across plasma membrane). Only 3 genes' differential expression in NAc overlapped with significant genes in a meta-analysis of bulk RNAseq in prefrontal cortex tissue (*WDFY3, CITED2, ARL4D*). By substantially increasing sample size, this study identifies novel differences in OOD-associated gene expression in human NAc. These results suggest genes that play a role in the immune system and calcium ion channel function which may impact the neurobiology of OOD and indicate that OOD-associated gene dysregulation differs across brain regions.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1699 Overweight as a causal factor for survival in late life: A Mendelian randomization study.

Authors:

H. Duan[1], R. Holmes[1], K. Arbeev[1], O. Bagley[1], D. Wu[1], N. Schupf[2], A. Yashin[1], S. Ukraintseva[1]; [1]Duke Univ., Durham, NC, [2]Columbia Univ., New York, NY

Background: The "obesity paradox" refers to an intriguing association of overweight with better survival at the oldest old ages observed in some studies; however, mechanism is not well understood. Mendelian Randomization (MR) is a common approach to uncovering causality in observational data using SNPs as instrumental variables (IVar). In our previous MR study of the Health and Retirement Study (HRS) data, we implemented a new strategy of selecting IVar, and found significant causal relationship between overweight and survival to ages 85+ in both sexes. Here, we use similar strategy to replicate the HRS finding in the Long Life Family Study (LLFS) data.
Method: To select IVar for MR, we used eight candidate obesity genes (*ADIPOQ, FTO, LEP, LEPR, INSIG2, MC4R, PCSK1, PPARG*), and cross-paired all SNPs in these genes to create new "composite SNP"s with genotypes 0,1,2, based on summed dosage of minor alleles (MA) of original SNPs. For each new composite SNP, '0' corresponded to 0-1; '1' corresponded to 2; and '2' corresponded to 3-4 MA (combined dosage) of the two original SNPs. The composite SNPs were tested to satisfy the assumptions for IVar. F-value>10 criteria was used to ensure IVar strength. This approach yielded considerably more IVar candidates that passed all the assumptions, compared to single-SNP approach. The following binary variables were used as "exposure": (1) 'overweight' BMI 25-30 vs. (0) 'normal' BMI 18.5-25, at ages 75-85; and as "outcome": (1) survived 85+ vs. (0) died before age 85. Covariates included field center, education, sex, smoking status, and first two principal components. The relatedness in LLFS data was addressed using SAS mixed models to calculate statistics. Inverse-Variance Weighted (IVW), and maximum likelihood (ML) methods in R-package MendelianRandomization were applied to test causality. Analyses were stratified by sex.
Result: We found that being overweight at ages 75-85 had significant protective effect on survival at ages 85+ in White males (IVar: 135 "SNP"s, IVW causal estimate(CE)=0.19, p-value=1.02E-15; ML CE=0.19, p-value=3.11E-15), and in total sample (IVar: 47 "SNP"s, IVW CE=0.31, p-value=2.52E-9; ML CE=0.31, p-value=7.98E-9). CEs for females didn’t reach statistical significance.
Conclusion: This MR study of LLFS data confirms our findings in HRS that being overweight in late life causally contributes to longevity in White males. Potentially better resilience of the overweight men to late life stresses may play role in this relationship. The observed sex difference may reflect better resilience of the longest-lived overweight men compared to women, as well as smaller LLFS sample compared to HRS.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1700 Parkinson's disease is associated with an imbalance in circular RNA expression

Authors:

B. Whittle, M. Santibanez-Koref, G. Hudson; Newcastle Univ., Newcastle Upon Tyne, United Kingdom

Neurodegeneration occurring in Parkinson’s disease (PD) can precede clinical diagnosis by decades. Earlier diagnosis of PD, as well as the ability to recognise individuals who will develop non-motor symptoms such as dementia, would facilitate the delivery of life-altering treatments earlier in the disease course. Many biomarkers of PD have been proposed, although, to date, none have been replicated or validated to the extent that they have been deployed in a clinical setting. Recent reports suggest that the effects of early neurodegeneration may be reflected in RNA levels in easily assessable tissues such as blood. However, despite several RNA studies, there is little concordance or agreement - in part hindered by cohort sizes, the heterogeneous nature of PD and the variability in analysis methodologies. We present the largest quantification of whole blood linear and circular RNAs (circRNA) in early-stage PD patients, using RNA sequencing data from two cohorts (PPMI = 259 PD, 161 Controls; ICICLE-PD = 48 PD, 48 Controls). We identified and replicated an increase in *TMEM252* and *LMNB1* gene expression in PD. We identified novel differences in the expression of circRNAs from *ESYT2*, *BMS1P1*, and *CCDC9*, in addition to replicating trends of some previously reported circRNAs. CircRNA expression showed a specific global reduction in PD patients, complemented by an increase in *RNASEL* expression. Overall, using circular RNA as a diagnostic biomarker in PD did not show any clear improvement over linear RNA. However, integrating the expression of the linear and circular transcriptome identified potential PD-related changes in the global expression of circRNAs. Interestingly, our data suggest that genes involved in mounting an immune response, specifically through the regulation of RNA subtypes, may play an important role in the development and pathoetiology of PD.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1701 Pathway-based genetic risk score analyses identify biological pathways linking together hypertension and dementia-related cognitive impairment traits

Authors:

O. Meirelles[1], C. Cabrera[2], H. Warren[3], D. Li[1], P. Munroe[4], L. Launer[1]; [1]Natl. Inst. on Aging, Baltimore, MD, [2]Queen Mary Univ. of London, London, United Kingdom, [3]William Harvey Res. Inst., Barts and The London Sch. of Med., QMUL, London, United Kingdom, [4]Queen Mary Univ. of London, London, United Kingdom

Cerebral small vessel disease is highly prevalent and contributes to cognitive impairment and dementia in the elderly. Elevated blood pressure (BP) increases risk for cerebrovascular damage, but the related mechanisms are uncertain. Published studies show associations between BP polygenic risk scores (PRS) and brain-MRI (magnetic resonance imaging) endophenotypes of dementia. However, analyses of full genome-wide BP-PRS cannot reveal which biological pathways are involved. Using novel methodological approaches, we constructed genetic risk scores (GRS) based on individual BP-associated pathways ("BP-path-GRS"). We selected 59 KEGG pathways most enriched for BP-genes from GWAS-based pathway analysis. Per pathway we listed all *k* genes, then per gene we extracted the variant most significantly associated with BP from BP-GWAS. Following pairwise-LD filtering, these *k* top variants were combined to construct three BP-path-GRS variables per pathway, using BP-trait-specific (systolic-BP, diastolic-BP, Pulse Pressure) weights in the GRS from BP-GWAS. Weights were adjusted according to the number of independent variants per gene, determined by Principal Component Analysis. We tested each of the 177 BP-path-GRS for association with three brain-MRI phenotypes [grey (GM), white (WM) matter, WM hyperintensity (WMH) volume]. Analyses were performed within n=37,599 UK-Biobank individuals with MRI data available. 27 of the 59 BP-pathways were significantly associated with at least one brain-MRI trait. Five pathways were robustly associated with WMH for all three BP-path-GRS. Two of these pathways remain jointly independently associated after multivariate stepwise analyses: "Cytomegalovirus" and "Endocrine/Calcium-regulating" pathways. These results suggest new underlying biological pathways linking hypertension and dementia, which could indicate potential novel drug targets.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1702 Personality traits are consistently associated with blood mitochondrial DNA copy number estimated from genome sequences in two genetic cohort studies.

Authors:

R. Oppong[1], A. Terracciano[2], M. Picard[3], Y. Qian[1], T. J. Butler[1], D. Zweibaum[1], T. Tian[1], T. Tanaka[1], A. Z. Moore[1], E. Simonsick[1], M. Gorospe[1], S. M. Resnick[1], F. Cucca[4], S. W. Scholz[5], B. J. Traynor[6], D. Schlessinger[1], L. Ferrucci[1], J. Ding[1]; [1]Natl. Inst. on Aging, BALTIMORE, MD, [2]Dept. of Geriatrics, Florida State Univ., Tallahassee, FL, [3]Dept. of Neurology, Columbia Univ., New York, NY, [4]IRGB CNR, Monserrato, Italy, [5]NIH, Bethesda, MD, [6]Natl. Inst Aging, Bethesda, MD

Levels of mitochondrial DNA copy number (mtDNAcn) in tissues and blood can be altered in conditions like diabetes and major depression and may play a role in aging and longevity. However, little is known about the association between mtDNAcn and personality traits, which are linked to emotional states, metabolic health, and longevity. This study aimed to test the hypothesis that blood mtDNAcn is related to personality traits and mediates the association between personality and mortality risk. We assessed the big five personality domains and facets using the Revised NEO Personality Inventory (NEO-PI-R), assessed depressive symptoms using the Center for Epidemiologic Studies Depression Scale (CES-D), estimated mtDNAcn levels from whole-genome sequencing, and tracked mortality in participants from the Baltimore Longitudinal Study of Aging (722 participants with complete data, mean age 75 (48 - 100) years, 48% women). Results were replicated in the SardiNIA Project (587 participants with complete data, mean age 57 (15 - 96) years, 62% women). We found that mtDNAcn was negatively associated with the Neuroticism domain and its facets and positively associated with facets from the other four domains (Extraversion, Openness, Agreeableness, and Conscientiousness). These associations were consistent across both cohorts and robust to adjustment for potential confounders. Additionally, higher depressive symptoms (CES-D) were associated with lower mtDNAcn levels, consistent with the Neuroticism finding. Finally, our analyses showed that mtDNAcn mediated the association between personality and mortality risk. To our knowledge, this study is the first to demonstrate a replicable association between blood mtDNAcn levels and personality traits, revealing a novel link between mitochondrial biology and emotional traits and states relevant to human health. Furthermore, we found that mtDNAcn mediates the association between personality and mortality risk, suggesting that mtDNAcn may be a biomarker of the biological process that underlies the association between personality and mortality risk. Currently, we are extending our study to the psychosocial traits using UK Biobank data and the preliminary analyses show that negative psychosocial traits are associated with lower mtDNAcn, and positive psychosocial traits are associated with higher mtDNAcn. These initial findings are in line with our results from the BLSA and SardiNIA cohorts.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1703 Personalizing body mass index thresholds for obesity using polygenic scores.

Authors:

I. Shim[1], M. Kim[1,2,3], B. Kim[1], P. Natarajan[2,3], R. Do[4], W-Y. Park[5], A. V. Khera[6], P. T. Ellinor[2,3], A. C. Fahed[2,3], H-H. Won[1,5]; [1]Samsung Advanced Inst. for Hlth.Sci. & Technology (SAIHST), Sungkyunkwan Univ., Seoul, Korea, Republic of, [2]Massachusetts Gen. Hosp. and Harvard Med. Sch., Boston, MA, [3]Broad Inst. of Harvard and MIT, Cambridge, MA, [4]Icahn Sch. of Med. at Mount Sinai, New York, NY, [5]Samsung Genome Inst., Samsung Med. Ctr., Seoul, Korea, Republic of, [6]Verve Therapeutics, Boston, MA

Background

The global obesity crisis continues to be a key driver of type 2 diabetes (DM-2) and cardiovascular diseases (CVD). Current guidelines use a 30 kg/m^2 body mass index (BMI) threshold to define obesity and decide on behavioral interventions, anti-obesity medications, and sometimes weight reduction surgery. Since BMI is a highly heritable trait, we sought to determine if obesity thresholds for management should be personalized based on genetic predisposition.

Methods

We quantified genetic susceptibility to obesity in 335,835 UK Biobank participants of European ancestry using a polygenic score (PGS) for BMI. We defined two obesity outcomes, DM-2 and a composite CVD outcome consisting of 7 obesity-related cardiovascular morbidities (coronary artery disease, heart failure, atrial fibrillation, pulmonary embolism, venous thromboembolism, hypertension, and aortic valve stenosis). We quantified the risk associated with the two outcomes among a population with obesity (BMI ≥30 kg/m^2) compared to a normal weight reference group (BMI 18.5-24.9 kg/m^2). We then identified BMI thresholds for equivalent risks across different genetic risk categories. Logistic regression models were adjusted for age, sex, education level, socioeconomic status, smoking, genotyping array, principal components of ancestry, and obesity-related lifestyle factors.

Results

Among 335,835 participants (mean age 57.37±7.99 years; 53.6% female; mean BMI 27.45±4.71 kg/m^2), individuals with obesity as defined by BMI ≥30 kg/m^2 had 8-fold (OR 7.99, 95% CI 7.52-8.50) increased risk of DM-2 and 4-fold (OR 3.68, 95% CI 3.58-3.78) increased risk of composite CVD outcome, compared to a group with normal weight. However, genetic risk-adjusted BMI thresholds varied from 28.53 kg/m^2 (top 5% PGS) to 30.27 kg/m^2 (bottom 5% PGS) for DM-2 risk, and from 28.91 kg/m^2 (top 5% PGS) to 31.23 kg/m^2 (bottom 5% PGS) for the composite CVD risk. BMI threshold showed a dose-responsive decrease with PGS increase. With the genetic risk-adjusted BMI threshold, 6.9% of participants with conventionally defined obesity were recalibrated, while the net population with obesity remained stable.

Conclusions

Incorporating genetic risk into BMI thresholds can enrich personalized risk assessment and better inform management strategies. We found that nearly 7% of individuals with obesity were recalibrated after considering genetic risk.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1704 Phenome-wide analyses using clinical single nucleotide polymorphism data in pediatricepilepsy.

Authors:

S. Ganesan[1,2,3,4], S. Parthasarathy[1,2,3], K. Zhu[2], M. Owens[2], C. Penney[2,5], P. Vaidiswaran[1,2,3], J. M. Miller[2], E. T. Dechene[2,6], L. K. Conlin[6,7], I. Helbig[1,2,3,8]; [1]Div. of Neurology, Children's Hosp. of Philadelphia, Philadelphia, PA, [2]Dept. of BioMed. and Hlth. Informatics (DBHi), Children's Hosp. of Philadelphia, Philadelphia, PA, [3]The Epilepsy NeuroGenetics Initiative (ENGIN), Children's Hosp. of Philadelphia, Philadelphia, PA, [4]Sch. of BioMed. Engineering, Sci. and Hlth.Systems, Drexel Univ., Philadelphia, PA, [5]Dept. of Data Sci. and Biostatistics Unit, Children's Hosp. of Philadelphia, Philadelphia, PA, [6]Div. of Genomic Diagnostics, Children's Hosp. of Philadelphia, Philadelphia, PA, [7]Pathology and Lab. Med., Univ. of Pennsylvania Perelman Sch. of Med., Philadelphia, PA, [8]Dept. of Neurology, Univ. of Pennsylvania Perelman Sch. of Med., Philadelphia, PA

Single Nucleotide Polymorphism (SNP) arrays have been used in the clinical diagnostic context to identify disease-causing copy number variations for decades. SNP data collected in large scale by various research studies and biobanks have empowered Genome Wide Association Studies (GWAS) to identify common variants associated with a particular phenotype of interest, and phenome-wide association studies (PheWAS) have further emerged to relate variants across the phenotypic landscape. However, diagnostic genotyping represents a data source with relatively untapped potential to identify such associations across the phenome, especially given the wealth of clinical information readily accessible in the electronic medical record (EMR). Here, we leverage the unique opportunity afforded by secondary research use of clinical SNP arrays of 11,860 individuals with EMR data from the Children's Hospital of Philadelphia to detect epilepsy-associated variants and perform PheWAS. We identified 2,301 cases with an epilepsy-related ICD10 code (G40, G41, R56 or P90). GWAS analysis revealed 49 genomic loci that achieved genome-wide significance, representing 28 genes. Of these, 4 genes have known epilepsy associations from the GWAS Catalog. Furthermore, 9 genes with genome-wide significant loci also have known rare variant associations with Mendelian epilepsy or neurodevelopmental phenotypes, suggesting possible convergence of common and rare variant mechanisms in these genes. Lastly, PheWAS across 11,312 unique ICD10 diagnoses mapped to 1,547 phecodes identified 335 phenome-wide significant SNP-phenotype associations including atypical development with variants in *PRDM5* and *MAD2L1*. Overall, our analyses indicate the relationship between common variants and phenotypes in individuals with epilepsy phenotype and entail for further evaluation. These associations reflect the potential for discovery from merging already available clinical and diagnostic genotyping datasets with corresponding deep EMR phenotyping.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1705 Phenome-wide association study on polygenic risk score of obesity in Korean

Authors:

S. Yang, M. Yuk, K. Oh, Y. Kwak, J. Lee, J. Youn, N. Song; Chungbuk Natl. Univ., Cheongju, Korea, Republic of

Background Obesity is a leading cause of non-communicable diseases such as diabetes, hypertension, hyperlipidemia, metabolic syndrome, cancer, and cardiovascular diseases. Prediction and prevention of obesity are important in terms of the increased prevalence of obesity and the social costs. In this study, we developed polygenic risk score (PRS) for obesity and conducted a phenome-wide association study (PheWAS) to investigate the trait-related consequences of obesity.

Methods This study used epidemiological genetic data from 58,690 participants of the Korean Genome and Epidemiology Study_Health EXAminee (KoGES_HEXA) study. We considered 92 phenotypes of potential trait-related consequences obtained by survey-based questionnaires and health examination. The genetic data were generated by the Korea Biobank Array and imputed using Korean Reference Genome. To develop PRS of obesity, we used 10 single-nucleotide polymorphisms (SNPs), including rs506589, rs11675198, rs3755804, rs792361, rs9348441, rs2076308, rs6265, rs8050136, rs6567160, and rs35560038, associated with obesity from the genome-wide association study (GWAS, $P<5.00\times10^{-8}$). The PheWAS is analyzed by a logistic or linear regression model for binary or continuous traits, respectively, with adjustment for sex and age.

Results The mean of PRS was 0.333 (SD=0.09) in obesity group and 0.319 (SD=0.09) in non-obesity group with significant difference ($P<5.16\times10^{-64}$). Each 0.1 increase in the obesity-PRS was associated with a 1.18 higher risk of obesity. In PheWAS, the eight phenotypes had an association with obesity-PRS; The most associated phenotype was hypertension (OR=1.60, $P=6.51\times10^{-5}$) and followed by arthritis (OR=1.86, $P=1.61\times10^{-4}$), intestinal polyp (OR=1.71, P=6.28×10-3), family history of hypertension(OR=1.30, P=0.007), breast cancer (OR=0.23, P=0.008), Hemoglobin(Beta=0.12, P=0.03), and family history of fatty liver (OR=1.47, P=0.05). Although not all phenome-wide associations were statistically significant, most diseases (61%) tended to show a positive direction with obesity-PRS.

Conclusion This study found that obesity-PRS was associated with increased risk of various diseases, especially hypertension supporting the potential causal relationship between obesity and chronic diseases. Obesity-related genetic data could attribute to identifying individuals at risk prevent obesity-related diseases.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1706 Phenome-wide genetic associations of educational attainment with mental and behavioral disorders

Authors:

A. Ayer, R. Wang, N. Banerjee, A. J. Averitt, A. Moscati, N. Lin, DiscovEHR-RGC, Mayo Clinic Project Generation, Regeneron Genetics Center, M. Cantor, A. Baras, E. Stahl, G. Coppola, V. M. Rajagopal; Regeneron Genetics Ctr., Tarrytown, NY

Background Many mental and behavioral disorders have strong genetic relationships with educational attainment (EA), and causality may be bidirectional and mediated via socioeconomic factors. Understanding these relationships can help interpret gene discoveries made by large-scale rare variant studies for such phenotypes. Here, we use common variants to investigate causal associations of EA with mental and behavioral disorders and bring insight into related rare variant associations. **Methods** We used the latest EA GWAS results to calculate a polygenic score (EA-PGS) for 350,475 individuals of European ancestry from five electronic health record-based cohorts. We then performed a phenome-wide association study using 83 mental and behavioral disorders based on ICD10 codes (Chapter V: F00-F99). Next, we used Mendelian Randomization (MR) to refine results and interpret direction of causality of significant associations. EA-PGS was calculated using LDpred, and PGS associations were computed using logistic regression in unrelated Europeans adjusted for appropriate covariates. Common variant associations were calculated using REGENIE and summary statistics were analyzed by MR, using the 'TwoSampleMR' R package. Statistical significance was determined using an FDR of 0.05. **Results** We identified 24 phenotypes significantly ($p < 5.7 \times 10^{-3}$) associated with EA-PGS, predominantly belonging to three groups: substance use, neurotic & stress-related, and mood disorders. MR results supported these findings and elucidated direction of causality. We found evidence for bidirectional causal associations between EA and all three disorder groups, but effects were relatively stronger in the direction of EA to disorders than disorders to EA (which could be partly due to a lack of strong disorder-specific instruments for the disorders). Substance use disorders, particularly tobacco use, showed the strongest associations. One standard deviation (SD) decrease in EA was associated with a 2.16 odds ratio of having a tobacco use disorder (95% CI = 1.90-2.45; $p = 3.7 \times 10^{-32}$), while one SD increase in genetic risk for tobacco use was associated with 0.14 SD decrease in EA ($\beta = 0.14$; SE = 0.013; $p = 1.6 \times 10^{-28}$). **Discussion** Our findings highlight strong genetic links between substance use disorders and EA, which is likely primarily mediated by socioeconomic factors. The results have helped interpret our recent findings on the links between rare variants in neurodevelopmental genes (which have strong negative effects on cognitive function and thereby on educational and socioeconomic outcomes) and tobacco use, which we believe is mediated by EA and related socioeconomic factors.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1707 Phenotype-driven filtering of whole-exome data from English-speaking individuals with specific language impairment (SLI) exhibiting precise grammar impairment reveals novel rare variants associated with SLI.

Authors:

M. Raza[1], E. M. Andres[2], K. K. Earnest[1], H. Xuan[1], C. Zhong[1], M. L. Rice[1]; [1]The Univ. of Kansas, Lawrence, KS, [2]Univ. of Missouri, Columbia, MO

Individuals with specific language impairment (SLI) are delayed in language acquisition and show persistent deficits in language despite average non-verbal intelligence and otherwise typical development, providing a test case for the specificity of language. There are two main SLI accounts; one focused on grammar acquisition and another implicating more domain-general mechanisms related to working memory. The current study focuses on the grammar acquisition account. In English-speaking children with SLI (ages 3;0-8;11 years), grammar impairment can be measured through performance on a grammar assessment, the Test of Early Grammatical Impairment (TEGI). The TEGI is highly sensitive and specific, consistently shows high heritability estimates (up to .92) and was used in only one prior molecular genetic study. Twin and family aggregation studies consistently indicate genetic factors are involved in language and SLI, but precise genetic mechanisms have not been revealed. Therefore, we set out to identify rare variants associated with TEGI performance. We performed whole exome sequencing (WES) in eight families with SLI (n = 74 total; n = 34 with TEGI score) and follow-up Sanger sequencing of identified co-segregating variants in additional unrelated probands who completed the TEGI (n = 146). We prioritized rare exonic variants shared by members of at least two families who performed below average on the TEGI under two filtering workflows: 1) novel and 2) previously reported candidate genes. We observed candidate variants on six new genes (*PDHA2*, *PCDHB3*, *FURIN*, *NOL6*, *IQGAP3*, and *BAHCC1*), and two genes previously reported for overall language ability (*GLI3* and *FLNB*). In at least one unrelated proband, we observed the same variants as those identified in the WES family members on *PDHA2*, *PCDHB3*, *NOL6*, and *IQGAP3*. Support from other studies suggests *PCDHB3*, a protocadherin gene, and *NOL6*, critical for ribosome synthesis, are possible important targets of future SLI investigation. Identifying the same variants associated with TEGI performance in multiple families emphasizes the utility of precise phenotyping and family-based genetic study.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1708 Plasma proteomic biomarkers of aortic stenosis: A Mendelian randomization study.

Authors:

T-Y. Yang[1,2,3], A. M. Small[4,5,6], L. Dufresne[1], G. Peloso[7,8], P. Natarajan[9,6], J. C. Engert[1,3,10], G. Thanassoulis[1,10], Therapeutic targets for AoRtic stenosis using GEneTics (TARGET) Consortium; [1]Preventive and Genomic Cardiology, McGill Univ. Hlth.Ctr. and Res. Inst., Montreal, QC, Canada, [2]Ctr. for Genomic Med., Kyoto Univ., Kyoto, Japan, [3]Dept. of Human Genetics, McGill Univ., Montreal, QC, Canada, [4]Div. of Cardiovascular Med., Dept. of Med., Brigham and Women's Hosp., Harvard Med. Sch., Boston, MA, [5]Dept. of Cardiology, Boston VA Hlth.care System, West Roxbury, MA, [6]Cardiovascular Res. Ctr. and Ctr. for Genomic Med., Massachusetts Gen. Hosp., Boston, MA, [7]Massachusetts Veterans Epidemiology Res. and Information Ctr. (MAVERIC), VA, Boston Hlth.care System, Boston, MA, [8]Dept. of Biostatistics, Boston Univ. Sch. of Publ. Hlth., Boston, MA, [9]Program in Med. and Population Genetics and the Cardiovascular Disease Initiative, Broad Inst. of MIT and Harvard, Cambridge, MA, [10]Div. of Experimental Med., McGill Univ., Montreal, QC, Canada

Introduction: Aortic stenosis (AS) is the most common valvular heart disease with a prevalence of 10%-15% among individuals over 75 years. Currently, aortic valve replacement is the only available treatment. Understanding the genetic etiology of AS could help identify potential therapeutic targets. Large-scale plasma protein measurements can be integrated with genomic data to define protein quantitative trait loci (pQTL). Mendelian randomization (MR) of pQTLs with AS could identify actionable, causal protein biomarkers for the disease. **Methods:** We employed a two-sample MR approach to identify causal contributors to AS among plasma proteins. Publicly available pQTL results from the Atherosclerosis Risk in Communities (ARIC) study were analyzed with our recently published meta-analysis of ten AS GWAS (Meta-AS). Results were replicated using GWAS summary data from the Million Veteran Program (MVP). **Results:** Using 2,004 cis-pQTLs identified in ARIC Europeans, we generated 1,238 MR results with the inverse variance weighted method using multiple SNPs as instrumental variables. We found that the most significant proteins were ITPA (OR 1.05, 95% CI 1.03-1.08; $p = 3.79 \times 10^{-5}$; FDR < 0.05), ACE (OR 1.04, 95% CI 1.02-1.06; $p = 1.73 \times 10^{-4}$), and PCSK9 (OR 1.20, 95% CI 1.09-1.32; $p = 1.88 \times 10^{-4}$). Additionally, the results of the weighted median analysis confirmed the significance of these three exposures (ITPA $p = 1.49 \times 10^{-3}$; ACE $p = 6.04 \times 10^{-3}$; PCSK9 $p = 8.59 \times 10^{-4}$). The egger intercepts were also found to be insignificant. In the MVP replication, all three of these proteins were significant: ITPA (OR 1.03, 95% CI 1.01-1.06; $p = 1.04 \times 10^{-3}$), ACE (OR 1.03, 95% CI 1.01-1.05; $p = 1.55 \times 10^{-3}$), and PCSK9 (OR 1.19, 95% CI 1.03-1.38; $p = 1.65 \times 10^{-2}$). These three exposures were also tested in MVP African Americans using ARIC cis-pQTLs generated in African Americans, but no significant results were obtained. **Conclusions:** In the largest MR study to date of pQTLs in AS, we identified three causal proteins for AS - ACE, PCSK9, and IPTA, all of which replicated in an independent sample. These results suggested three proteins that could mediate the pathogenesis of AS and represent potentially actionable targets. Notably, ACE and PCSK9 are already recognized as viable drug targets. Additional research is warranted to determine whether these pharmacotherapies might be repurposed for AS.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1709 Plasma proteomics and insulin sensitivity: evidence from observational, intervention and genetic studies

Authors:

P. Kho[1,2], D. Zanetti[3], N. Wary[1], F. Abbasi[1,4], J. W. Knowles[1,4,5,6], D. J. Panyard[7,2], K. T. Watson[8], L. Stell[9,2], L. C. Lazzeroni[10], S. Gustafsson[11], L. Lind[11], J. R. Petrie[12], T. L. Assimes[13,2,4,5,14]; [1]Dept. of Med., Div. of Cardiovascular Med., Stanford Univ. Sch. of Med., Stanford, CA, [2]VA Palo Alto Hlth.Care System, Palo Alto, CA, [3]Inst. of genetic and bioMed. research (IRGB) / Natl. Res. Council (CNR), Monserrato, Italy, [4]Stanford Diabetes Res. Ctr., Stanford Univ. Sch. of Med., Stanford, CA, [5]Stanford Cardiovascular Inst., Stanford Univ. Sch. of Med., Stanford, CA, [6]Stanford Prevention Res. Ctr., Stanford Univ. Sch. of Med., Stanford, CA, [7]Dept. of Genetics, Stanford Univ. Sch. of Med., Stanford, CA, [8]Dept. of Psychiatry, Stanford Univ. Sch. of Med., Stanford, CA, [9]Dept. of BioMed. Data Sci., Stanford Univ. Sch. of Med., Stanford, CA, [10]Dept. of Psychiatry and Behavioral Sci., Stanford Univ., Stanford, CA, [11]Dept. of Med. Sci., Uppsala Univ., Uppsala, Sweden, [12]Sch. of Hlth.and Wellbeing, Coll. of Med. Vet. and Life Sci., Univ. of Glasgow, Glasgow, United Kingdom, [13]Dept. of Med., Div. of Cardiovascular Med., Stanford Univ. Sch. of Med.,, Stanford, CA, [14]Dept. of Epidemiology and Population Hlth., Stanford Univ. Sch. of Med., Stanford, CA

Insulin resistance, a primary risk factor for the development of type 2 diabetes and coronary artery disease, has significant health implications. In two large European cohorts, we have previously shown that plasma proteins measured using the proximity extension assay (PEA) substantially increase the variance explained of a gold standard direct measure of insulin sensitivity (IS) beyond that explained by routinely acquired clinical risk factors. However, it remains unclear whether the same proteins can track IS in non-European populations, whether they can track short term changes in IS, and which associations we observed are causal.

To address these knowledge gaps, we measured 1471 plasma proteins with PEA in a diverse set of participants who had previously undergone one or more insulin suppression tests at Stanford University including 655 White, 122 South Asian, 95 East Asian, 86 Hispanic, and 54 Black volunteers. A total of 810 proteins associated with baseline IS in European subjects at FDR < 0.05, of which 395 passed Bonferroni correction. Associations observed correlated strongly with those observed in non-White populations (r=0.68 to 0.83, all $P<4.3\times10^{-113}$). Next, we integrated our Stanford data with the two European studies (RISC n=1024, and ULSAM n=889) through meta-analysis which yielded 255 candidate plasma proteins at FDR < 0.05, of which 75 remained significant even after adjusting for BMI. In a subset of participants who underwent short term interventions that substantially improved IS through the use of thiazolidinediones drugs (N=53) or weight loss (N=66), we measured the same 255 proteins a second time after the intervention and observed a very high concordance between the observed and the expected change of the protein levels based on the direction of association observed at baseline (79.6% concordance with chi^2 $p=6.7x10^{-16}$ before BMI adjustment, and 73.3% with chi^2 $p=1.3x10^{-8}$ after adjustment). Lastly, we conducted Mendelian randomization (MR) analyses leveraging recently identified genetic instruments by the UK Biobank Pharma Proteomics Project and identified fifteen candidate causal proteins at a nominal P<0.05. Seven (AIFM1, IL10RB, LBP, LEP, MEGF9, PTPRN2 and QDPR) showed persistent associations after MR sensitivity analyses. Our follow up study reinforces the utility of plasma proteins in the non-invasive assessment of IS beyond established clinical risk markers. The same proteins are likely to be informative across all populations and to be able to accurately track short term changes in IS. Lastly, we shed light on the molecular mechanisms underlying insulin resistance and identify several novel candidate therapeutic targets.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1710 Plasticity of human microglia and their interaction with the brain immune system in AD

Authors:

D. Lee[1], C. Porras[1], P. Weiler[2], R. Kosoy[1], M. Pjanic[1], J. Bendl[1], P. Fnu[1], C. Spencer[1], K. Therrien[1], D. Mathur[1], S. P. Kleopoulos[1], Z. Shao[1], M. Alvia[1], C. Casey[1], A. Hong[1], S. Argyriou[1], K. G. Beaumont[1], R. Sebra[1], P. Auluck[3], S. Marenco[3], C. P. Kellner[1], D. Bennett[4], G. Voloudakis[1], F. J. Theis[2], V. Haroutunian[1], G. E. Hoffman[1], J. F. Fullard[1], P. Roussos[1]; [1]Icahn Sch. of Med. at Mount Sinai, New York, NY, [2]Helmholtz Ctr. Munich, Munich, Germany, [3]Natl. Inst. of Mental Hlth., Bethesda, MD, [4]Rush Univ. Med. Ctr., Chicago, IL

Myeloid-origin human brain immune cells (HBICs), including microglia and perivascular macrophages (PVMs), have been identified as playing a central role in the onset and progression of Alzheimer's disease (AD). However, the field lacks a unified taxonomy describing HBIC subtypes due to their heterogeneity and the limited scale of previous studies involving human brain tissue. In this study, we provide a detailed consensus on HBICs using multi-modal single-cell assays from diverse populations across the lifespan with varying degrees of AD pathology. Establishing a unified taxonomy of human microglial states could help describe the progression of AD in terms of changes in microglial states. To achieve this, we generated two largest independent single-cell datasets to date. The first dataset provides a comprehensive, in-depth view of microglial transcriptomes using scRNA-seq profiles of live microglia purified from an autopsy of 137 unique donors. We identify 13 distinct phenotypes of human microglia and show how they react and adapt to environmental cues. We validate and further characterize the microglial taxonomy using the second dataset, snRNA-seq profiles collected from 1,470 unique donors, which provides a much greater breadth of coverage and span of a lifetime. We study the population-level variation in microglial heterogeneity using these complementary datasets. We observe state-specific expression of surface proteins from validated microglial subtypes using CITE-seq. We describe microglial states implicated at different stages of AD pathology. Based on our findings, we propose a two-stage disease mechanism for pathogenic progression (Amyloid and Tau) and cognitive impairments (dementia). We discover AD-risk candidate genes PTPRG, DPYD, and IL15, which are implicated in both the early and late stages of AD. Our resource and approach provide unprecedented systems-level insights into transcriptional dysregulation of HBICs and its implications in AD pathogenesis.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1711 Pleiotropic relationship between Alzheimer's disease and eleven immune diseases.

Authors:

N. Enduru, B. Fernandes, Y. Dai, Z. Zhao; The Univ. of Texas Hlth.Sci. Ctr. at Houston, Houston, TX

Background: Alzheimer's disease (AD) is a polygenic disease with an emerging global public health challenge. Recent evidence suggests that the pathogenesis of AD is not confined to neurons but instead underscores involvement of neuroinflammation in driving the disease. In this study, we aimed to investigate the shared genetic architecture between a plethora of immune diseases (ID) and AD using published genome-wide association studies (GWAS) and then identify genes related to both immune diseases and AD. **Methods**: We collected GWAS summary statistics data for AD and 11 featured IDs (allergic rhinitis, asthma, atopic dermatitis, celiac disease, Crohn's disease, hypothyroidism, primary sclerosing cholangitis, rheumatoid arthritis, systemic lupus erythematosus, ulcerative colitis, and vitiligo). We performed conjunctional/conditional false discovery rate (cFDR) analysis through pleioFDR to identify the shared pleiotropic relationship between AD and each ID. Additionally, local genetic correlation (LAVA) was conducted to assess genetic overlap at the pleiotropic loci between AD and each ID identified through conjFDR analysis. **Results**: We identified 70 unique shared loci from 71 SNPs between AD and ID with a conjFDR $_{AD\&ID} < 0.05$ across the 11 trait pairs. These loci were associated with 43 distinct protein-coding genes and 16 non-protein coding genes. Notably, two lead SNPs (rs12790721 and rs602662, nearest genes LINC01059 and FUT2, respectively, with CADD score > 12.37) exhibited high deleteriousness and seven lead SNPs (rs4233366, rs739954, rs2074404, rs10748526, rs3764310, rs4275659 and rs4958435, rs17229044, and rs1052553, nearest genes ADAMTS4, HBEGF, WNT3, TSPAN14, DHODH, ABCB9 and TNIP1, CLEC16A, and MAPT, respectively) showed potential regulatory functionality based on RegulomeDB scores. LAVA revealed 76 shared loci between AD and ID. Among them, 45 loci had significant heritability for both diseases in pair, 10 of which had significant correlation ($p < 0.05$). Twelve loci had 95% confidence intervals covering maximum variance value 1, suggesting complete shared genetic signal between the phenotypes. The FUMA for pathway analysis revealed several pathways associated with AD and IDs, including cytokine production, activation regulation of immune response, neuron projection, protein lipid complex, protein ligase and lipid bindings. **Conclusion**: There is a shared genetic overlap between AD and immune diseases, including shared molecular pathways.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1712 Pleiotropy evaluation across four abdominal hernia subtypes detects novel loci and shared genetic basis.

Authors:

H. Choquet[1], R. Mostaedi[2], J. Yin[1], T. Hoffmann[3], N. Ahituv[4], E. Jorgenson[5], M. Kvale[6]; [1]Kaiser Permanente - Div. of Res., Oakland, CA, [2]Kaiser Permanente - East Bay Richmond Med. Ctr., Richmond, CA, [3]Univ. of California, San Francisco (UCSF), San Francisco, CA, [4]UCSF, San Francisco, CA, [5]Regeneron Genetics Ctr., Tarrytown, NY, [6]Inst. for Human Genetics, San Francisco, CA

Purpose: Abdominal hernias can lead to serious medical morbidity, including bowel incarceration and strangulation, and emergency hernia repair surgery is associated with a substantial risk of morbidity and mortality. Array-based heritability estimates found a stronger contribution of genetic risk factors for umbilical hernia (16.0%) compared to inguinal (12%), ventral (7.0%), or femoral (6.0%), suggesting different etiology according to the hernia subtype. To date, more than 80 genetic loci have been associated with different forms of hernia, and several loci were associated with more than one hernia subtype. However, the overlap between the underlying genetic mechanisms remained largely unknown. **Methods:** We conducted a multiethnic genome-wide association (GWA) meta-analysis of each hernia subtype (inguinal, femoral, umbilical, ventral), combining results from the Kaiser Permanente Genetic Epidemiology Research on Adult Health and Aging (GERA) and the UK Biobank cohorts. Individual GWA analyses of each hernia subtype were performed using logistic regression adjusted for age, sex, and genetic ancestry principal components. Following our GWA analyses, we investigated genetic correlation between the four hernia subtypes using LD score regression. We also examined shared effects (pleiotropy) across hernia subtypes for each identified hernia-associated locus. **Results:** The total numbers of genome-wide significant loci ($P<5x10^{-8}$) associated with each hernia subtype (whether pleiotropic or not) from these GWA analyses were 63 (inguinal hernia), 26 (umbilical hernia), 14 (femoral hernia), and 11 (ventral hernia). Among those 98 unique loci, 25 were novel (i.e., have not been previously reported to be associated with any subtype of hernia). The four abdominal hernia subtypes were genetically correlated with one another ($0.34 \leq r_g < 0.76$; *P*-value<0.05). Patterns of pleiotropy across hernia subtypes were highly diverse, with only three loci (*LYPLAL1-SLC30A10*, *EFEMP1*, and *WT1-AS*) significantly associated with all four hernia subtypes; one locus (*CHRDL1*) significantly associated with three hernia subtypes; and five loci (*CWC27-ADAMTS6*, *AIG1*, *SRPX*, *RBPMS*, and *COL8A1*) significantly associated with two hernia subtypes. **Conclusions:** This large and multiethnic genetic study of abdominal hernias identified many novel genetic loci and subtype-specific genetic effects, which could help better understand the biological mechanisms through which hernias develop.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1713 Polygenic and transcriptional risk scores identify chronic obstructive pulmonary disease subtypes

Authors:

M. Moll[1], J. Hecker[2], J. Platig[3], K. Pratte[4], R-s. Wang[1], D. Hill[5], I. Konigsberg[6], J. Chiles[7], C. Hersh[8], P. Castaldi[9], K. Glass[1], J. Dy[1], D. Sin[10], R. Tal-Singer[1], M. Mouded[11], S. Rennard[12], G. Andersen[13], G. Kinney[6], R. Bowler[14], J. Curtis[15], M-L. Mcdonald[16], E. Silverman[17], M. Cho[18], B. Hobbs[19]; [1]Brigham & Women s Hosp., Boston, MA, [2]Harvard T.H. Chan Sch. of Publ. Hlth., Boston, Ma, Germany, [3]Univ. of Virginia, Charlottesville, VA, [4]Dept. of Biostatistics, Natl. Jewish Hlth., Denver, CO, [5]Northeastern Univ., Boston, MA, [6]Univ. of Colorado, Denver, CO, [7]Univ. of Alabama Birmingham, Birmingham, AL, [8]Brigham and Women's Hosp., Boston, MA, [9]Brigham and Women s Hosp., Boston, MA, [10]Univ. of British Columbia, Vancouver, BC, Canada, [11]Novartis, Cambridge, MA, [12]Univ. of Nebraska Med. Ctr., Omaha, NE, [13]Univ. of Melbourne, Melbourne, Australia, [14]Natl. Jewish Hlth., Denver, CO, [15]Univ. of Michigan, Ann Arbor, MI, [16]Univ. of Alabama of Birmingham, Birmingham, AL, [17]Brigham & Women's Hosp., Boston, MA, [18]Brigham and Women s Hosp., Duxbury, MA, [19]Brigham and Women s Hosp., Needham Heights, MA

Rationale: Chronic obstructive pulmonary disease (COPD) is heterogeneous. Polygenic and other Omics-based risk scores can identify high risk COPD subgroups, but the clinical and biological characteristics of high-risk groups are unclear. **Objectives:** Define COPD subgroups using previously-published genetic risk (polygenic risk score, PRS) and blood gene expression (transcriptional risk score, TRS) and assess differences in disease severities, activities, and risk for progression. **Methods:** By maximizing differences in protein biomarkers, we created two PRS and three TRS groups in a training set [n=1,374 non-Hispanic white (NHW) COPDGene participants], which we characterized using STRING protein-protein interaction network analyses. We applied Enrichr drug repurposing analyses and multivariable regressions to test group associations with COPD-related outcomes, adjusting for potential confounders. **Measurements and Main Results:** We examined these six omics-defined groups in non-overlapping test sets (n=1,133 NHW COPDGene, n=299 African American (AA) COPDGene, n=468 ECLIPSE). Two groups differed substantially, defining ‘high disease activity’ and ‘severe disease risk’ subtypes. ‘High disease activity’ participants had low PRS/high TRS, accelerated prospective FEV_1 decline (COPDGene: -51 mL/year; ECLIPSE: -40 mL/year), and altered extracellular matrix, MMP12, TGF-β, and IL-16 biology. ‘Severe disease risk’ participants had high PRS/high TRS, lower spirometry, greater emphysema and airway thickness, and changes in STAT5, IL-20, and TNF signaling. Both subtypes had lower mean BMI than other subtypes, with altered growth hormone, leptin, and prolactin signaling. ‘High disease activity’ individuals were enriched in gene sets perturbed by treatment with fibrates, calcium channel blockers, and antithyroid medications; “severe disease risk” individuals were enriched in gene sets perturbed by treatment with 5-lipoxygenase, carbonic anhydrase, and angiotensin-converting enzyme inhibitors. **Conclusions:** PRS and TRS identified subtypes with distinct clinical and biological characteristics. Polygenic risk scores combined with other Omics can be used to dissect complex disease heterogeneity and identify subtype-specific therapeutic strategies.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1714 Polygenic risk advances overall risk stratification for Parkinson's disease in combination with substantia nigra hyperechogenicity.

Authors:

A. Metspalu[1], M. Kals[1], T. Toomsoo[2,3], M. Vaht[1], T. Nikopensius[1], M. Teder-Laving[1], A. Reigo[1]; [1]Inst. of Genomics, Univ. of Tartu, Tartu, Estonia, [2]Confido Med. Ctr., Estonia, Tallinn, Estonia, [3]Sch. of Natural Sci. and Hlth., Tallinn Univ., Tallinn, Estonia

Parkinson's disease (PD) is the second largest neurodegenerative disease in the world. Both environmental and genetic factors play an important role in a complex multifactorial etiology of PD. The objective of this study was to determine whether the polygenic risk for PD was associated with an increased likelihood of PD in combination with hyperechogenicity of the substantia nigra (SN). We performed a pilot study from individuals of European ancestry with no evidence of PD and age of >55 years in Estonian Biobank.

Polygenic risk score (PRS) was calculated for everyone who met the inclusion criteria (n=41,042) and individuals with extreme PRS values for PD (the top and bottom 1%) were selected for the further analysis. From November 2021 to June 2022, all study participants were interviewed and underwent neurological assessments (including Sniffin' Sticks 12 items test for hyposmia, the Movement Disorder Society diagnostic for de novo PD, and transcranial sonography (TCS) for the area of echogenicity).

Four hundred and forty-seven individuals were invited to the study, 225 participated (response rate was 50.3%), of whom 107 (51 females) were at high genetic risk and 97 (45 females) low risk. Participants stratified by SN hyperechogenic area of SN⩾0.22 vs SN<0.22 were similar for comorbidities and family history of PD except for migraine with aura, which was more commonly present in the hyperechogenicity group (16% (9/56) vs 3% (3/148); p=0.001). The SN groups differed for polygenic risk, de novo PD and results of Sniffin' sticks test (all p<0.001). All 12 de novo PD patients had SN⩾0.22. A multivariable logistic model adjusted for age and sex showed a significant association between better Sniffin' sticks test result and a reduced risk of hyperechogenicity (adjusted odds (aOR) was 0.58, 95% CI 0.46-0.95; p<0.0001). The model explained 11% of the variance (area under the curve (AUC) was 0.71) in SN outcome. The addition of the PRS risk group significantly increased the variance explained to 21% (AUC=0.81). The aOR of PD for the high polygenic risk group was more than 6 times higher (aOR=6.35, 95% CI 2.85-14.13; p<0.0001). Interaction between Sniffin' sticks test and PRS risk groups was not significantly associated with SN outcome (p=0.78).

These finding show that the risk for PD is genetically influenced. An individual's PRS could be clinically actionable and adds predictive value for the development of PD, if used with other non-genetic predictors (lifestyle, comorbidities, echogenicity in SN). We showed for the first time that prodromal markers such as TCS and hyposmia can strengthen the effect of PRS in early detection of PD, which needs to be investigated prospectively.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1715 Polygenic risk of idiopathic pulmonary fibrosis and COVID-19 severity

Authors:

C. Flores[1,2,3], B. Guillen-Guio[4,5], I. Marcelino-Rodriguez[6,2], J. M. Lorenzo-Salazar[1], O. C. Leavy[4,5], R. J. Allen[4,5], Scourge Cohort Group, J. A. Riancho[7,8,9], A. Rojas[10], P. Lapunzina[11,12,13], A. Carracedo[14,12,15,16,17], L. V. Wain[4,5]; [1]Genomics Div., Inst. Tecnologico y de Energias Renovables, Granadilla de Abona, Spain, [2]Res. Unit, Hosp. Univ.rio Nuestra Señora de Candelaria, Santa Cruz de Tenerife, Spain, [3]Centro de Investigacion Biomedica en Red de Enfermedades Respiratorias (CIBERES), Inst. de Salud Carlos III, Madrid, Spain, [4]Dept. of Population Hlth.Sci., Univ. of Leicester, Leicester, United Kingdom, [5]NIHR Leicester BioMed. Res. Ctr., Leicester, United Kingdom, [6]Publ. Hlth.and Preventive Med. Area, Univ. de La Laguna, San Cristobal de La Laguna, Spain, [7]Servicio de Med. Interna, Hosp. Univ.rio Marqués de Valdecilla, Cantabria, Spain, [8]IDIVAL, Cantabria, Spain, [9]Departamento de Med. y Psiquiatria, Univ. de Cantabria, Cantabria, Spain, [10]Tecnologico de Monterrey, Escuela de Med. y Ciencias de la Salud, Monterrey, Mexico, [11]Inst. de Genetica Medica y Molecular (INGEMM), Hosp. Univ.rio La Paz-IDIPAZ, Madrid, Spain, [12]Centro de Investigación Biomédica en Red de Enfermedades Raras (CIBERER), Inst. de Salud Carlos III, Madrid, Spain, [13]ERN-ITHACA-European Reference Network, Madrid, Spain, [14]Grupo de Med. Xeomica, CIMUS, Univ.e de Santiago de Compostela, Santiago de Compostela, Spain, [15]Inst. de Investigación Sanitaria de Santiago (IDIS), Santiago de Compostela, Spain, [16]Centro Singular de Investigacion en Med. Molecular y Enfermedades Cronicas (CIMUS), Univ.e de Santiago de Compostela, Santiago de Compostela, Spain, [17]Fundación Publ.a Galega de Med. Xenomica, Centro Natl. de Genotipado, Sistema Galego de Saude (SERGAS), Santiago de Compostela, Spain

Rationale: Patients with severe coronavirus disease 2019 (COVID-19) can develop residual lung abnormalities consistent with lung fibrosis. Existing studies assessing the genetic overlap of idiopathic pulmonary fibrosis (IPF) and COVID-19 have primarily concentrated on previously identified IPF genome-wide significant risk variants, rather than on combined whole-genome polygenic risk. **Objective:** We hypothesized that the use of IPF whole-genome polygenic risk scores (PRSs) will assist in further characterizing the shared genetic component with COVID-19. Here we derived PRS models from IPF genome-wide association study (GWAS) results and studied their association with COVID-19 severity. **Methods:** We used results from the largest meta-GWAS of clinically defined IPF risk (base dataset; n=24,589) and individual-level imputed data from the SCOURGE study of patients with COVID-19 (target dataset; n=15,024). We calculated IPF PRSs using PRSice-2 and assessed their association with COVID-19 hospitalization, severe illness, and critical illness. We also evaluated the association stratifying by age and sex. Results were validated using an independent PRS method. Enrichment analyses and pathway-specific PRSs were performed to study biological pathways associated with COVID-19 severity. **Results:** IPF PRSs were significantly associated with COVID-19 hospitalization and severe illness. The strongest association was found in patients aged 60 years or younger, especially among younger males (OR=1.16; 95%CI=1.08-1.25; $p=6.39x10^{-5}$). A pathway enrichment analysis of the variants included in the best model fit and subsequent pathway-specific PRS analyses supported the link of Cadherin and Integrin signaling pathways with COVID-19 severity when stratified by age and sex. **Conclusion:** Our results suggest that there is a genome-wide genetic overlap between IPF and severe COVID-19 that is dependent on age and sex and adds further support that the pathogenesis of both IPF and severe COVID-19 share underlying biological mechanisms. This could imply that individuals with high IPF genetic risk are at an overall increased risk of developing post-severe COVID-19 lung sequelae. **Funding:** Wellcome Trust 221680/Z/20/Z; GlaxoSmithKline Asthma + Lung UK Chair in Respiratory Research (C17-1); Instituto de Salud Carlos III (COV20_00622, PI20/00876); European Union (ERDF) 'A way of making Europe'; Fundación Amancio Ortega, Banco de Santander; Cabildo Insular de Tenerife (CGIEU0000219140 and 'Apuestas científicas del ITER para colaborar en la lucha contra la COVID-19'); Fundación Canaria Instituto de Investigación Sanitaria de Canarias (PIFIISC20/57).

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1716 Polygenic risk score performance in cancer prediction and association with other cancers and cardiometabolic disease

Authors:

O. Clark[1], H. Desai[2], M. Vujkovic[3], C. Teerlink[4,5], P. Alba[4,5], K. Min Lee[4], J. Lynch[4,5], S. L. DuVall[4], VA Million Veteran Program, K-M. Chang[6,7], K. Maxwell[8,7,9]; [1]Div. of Internal Med., Pennsylvania Hosp., Univ. of Pennsylvania, Philadelphia, PA, [2]Div. of Hematology/Oncology, Dept. of Med., Perelman Sch. of Med., Univ. of Pennsylvania, Philadelphia, PA, [3]Div. of Translational Med. and Human Genetics, Dept. of Med., Perelman Sch. of Med., Univ. of Pennsylvania, Philadelphia, PA, [4]Dept. of VA Informatics and Computing Infrastructure, Dept. of VA Salt Lake City Hlth.Care System, Salt Lake City, UT, [5]Div. of Epidemiology, Dept. of Internal Med., Univ. of Utah Sch. of Med., Salt Lake City, UT, [6]Dept. of Med., Perelman Sch. of Med., Univ. of Pennnsylvania, Philadelphia, PA, [7]Corporal Michael J. Crescenz VA Med. Ctr., Philadelphia, PA, [8]Div. of Hematology/Oncology, Dept. of Med., Perelman Sch. of Med., Philadelphia, PA, [9]Dept. of Genetics, Perelman Sch. of Med., Univ. of Pennnsylvania, Philadelphia, PA

Background: Genome-Wide Association Studies (GWAS) have identified single nucleotide polymorphisms (SNPs) that are associated at low effect size with many cancers. Polygenic risk scores (PRS) have been shown to improve risk prediction in individuals of European genetic ancestry (EUR); however, predictive ability in patients of African genetic ancestry (AFR) is less clear. Furthermore, whether cancer PRS demonstrate pleiotropic effects on other cancers or diseases is unknown. **Methods:** We construct PRS from GWAS significant SNPs identified by the largest published GWAS for 13 cancers (bladder, breast, colorectal, esophageal, glioma, lung, melanoma, oral cavity, pancreatic, prostate, renal, testicular, and thyroid) in a large racially diverse biobank, The Million Veterans Program (MVP). We perform a phenome-wide association study (PheWAS) to determine associations between cancer-specific PRS and other phenotypes with emphasis on cross-cancer pleiotropy and cardiometabolic phenotypes. **Results:** All PRS were significantly associated with their respective cancers in EUR individuals (OR 1.04 to 1.70). In AFR individuals, PRS were significantly associated with their respective cancers for bladder, breast, colorectal, lung, pancreatic, prostate, testicular and thyroid (OR 1.05 to 1.48). Addition of the PRS component to basic models showed >1% increase in discriminatory accuracy (ΔAUC) for 7 of 13 cancers in EUR individuals, but only 1 of 13 cancers in AFR individuals. Cancer-cancer pleiotropy in EUR showed association of lung cancer PRS with Non-Hodgkin Lymphoma (OR=1.08, p<0.001), with other nominal associations. In AFR, bladder cancer PRS was associated with esophageal cancer (OR=1.32, p=0.03), breast with esophageal cancer (OR=1.30, p=0.04), ovarian with thyroid cancer (OR=1.34, p=0.03), testicular with non-Hodkin lymphoma (OR=1.16, p=0.01), and thyroid with kidney cancer (OR=1.10, p=0.04). Regarding cardiac phenotypes, colorectal PRS was associated with hypertrophic obstructive cardiomyopathy (OR=1.18, p=0.02), esophageal cancer with heart failure (OR=1.10, p<0.01), glioma with acute pericarditis (OR=1.20, p<0.01), and lung cancer with Granulomatosis with Polyangiitis (OR=1.51, p<0.001).**Conclusions:** Models including age, principal components, gender and cancer-specific PRS are strongly associated with their respective cancers in EUR individuals, and have less strong associations in AFR individuals, owing almost entirely to the non-PRS components of the model. While rare cross-cancer and cardiometabolic pleiotropic effects of cancer PRS exist, most cancer PRS show tissue and disease-specific genetic associations.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1717 Polygenic risk scores for major depressive disorder provide insights into medication use prediction in EHR-linked biobank.

Authors:

S. Lapinska[1], V. Venkateswaran[1], K. Boulier[1], Y. Ding[1], A. Pimplaskar[2], L. O. Loohuis[1], B. Pasaniuc[1]; [1]UCLA, Los Angeles, CA, [2]Univ. of California, Los Angeles, Los Angeles, CA

Major depressive disorder (MDD) is the most common psychiatric illness. Studies have shown that MDD has a complex etiology with possible population differences in predisposing genetic factors. Investigating the relationship between ancestry and MDD could provide insight into how individuals may respond to certain treatments. We use a publicly available polygenic risk score (PGS) trained on European individuals from the UK Biobank data for major depressive disorder to investigate the clinical utility of PGS in medication use prediction within the diverse UCLA ATLAS biobank (N &gt 40,000).First, we found that the standardized MDD-PGS is associated with the phecode for MDD across genetically inferred ancestry (GIA) groups including European American (EUR) (OR: 1.259, CI: [1.215, 1.306]), Hispanic Latin American (AMR) (OR: 1.117, CI: [1.036, 1.204]), African American (AFR) (OR: 1.126, CI: [0.971, 1.304]), and East Asian American (EAS) (OR: 1.425, CI: [1.264, 1.606]). Next, we identified binary medication-use phenotypes based on treatment options for individuals diagnosed with MDD using de-identified patient information within the EHR-linked ATLAS biobank. Medication-use was defined into three categories: long term use of the most recent medication class, treatment resistant, and efficacious to one medication class. For all analyses, we restricted to individuals diagnosed with MDD who have no other psychiatric comorbidities like bipolar, schizophrenia, substance use, and alcohol use disorders. Using a GIA- stratified logistic regression model adjusted for age, age2, sex, and first 5 principal components, we found MDD-PGS significantly associated with long term serotonin use in MDD diagnosed individuals with AFR ancestry (OR: 0.590, CI: [0.370, 0.941], p-value: 0.027). This suggests that higher genetic risk for MDD correlates with decrease in serotonin usage in individuals of AFR ancestries. In addition, MDD-PGS significantly associated with long term antipsychotic use in MDD diagnosed individuals with EAS ancestry (OR: 1.738, CI: [1.015, 2.978], p-value: 0.044) suggesting that these individuals have an increase in antipsychotic usage with higher genetic risk for MDD.Overall our study demonstrates the potential use of PGS to determine effective treatments for individuals diagnosed with major depressive disorder across ancestries. Further work to understand the relationship between polygenic risk scores, medication use, and ancestry is needed, but our current findings show a promising clinical utility of PGS.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1718 Polygenic risk scores predict blood pressure traits across the lifespan

Authors:

K. Øvretveit[1], E. M. L. Ingeström[1], M. Spitieris[1], V. Tragante[2], K. H. Wade[3], L. Thomas[1], B. Wolford[1], U. Wisløff[1], D. Gudbjartsson[2], H. Holm[2], K. Stefansson[2], B. Brumpton[1], K. Hveem[1]; [1]Norwegian Univ. of Sci. and Technology, Trondheim, Norway, [2]deCODE genetics, Reykjavik, Iceland, [3]Univ. of Bristol, Bristol, United Kingdom

Background: Hypertension is a major modifiable cause of morbidity and mortality that affects over 1 billion people worldwide. Hypertension-related disease risk occurs at high blood pressure (BP) levels and well down into the normal range. The genetic contribution to complex phenotypes such as BP traits is quantifiable with polygenic risk scores (PRSs), but the performance of such risk constructs can differ based on factors such as derivation method, trait polygenicity, and the size and characteristics of the base, training, and validation data.
Aims: Develop multiple PRSs for BP traits incorporating a range of <1,000 to >1,000,000 genetic variants, compare their strength of association with various BP traits across the lifespan, and test the predictive performance of the top performing PRSs on hypertension-related morbidity and mortality.
Methods: We included participants from the Trøndelag Health Study (HUNT; n = 86,569), deCODE genetics (n = 81,117), and the Avon Longitudinal Study of Parents and Children (ALSPAC) study (n = 6,807). Four methods were used to construct PRSs for systolic BP (SBP), diastolic BP (DBP), and pulse pressure (PP): a weighted PRS restricted to genome-wide significant variants, pruning and thresholding (P+T), and the two Bayesian approaches (LDpred and PRS-CS). 87 scores across the three BP traits were trained in an independent cohort (deCODE). The best performing scores were applied in cohorts of children (ALSPAC) and adults (HUNT).
Results: Bayesian methods incorporating >1,000,000 genetic variants improved phenotypic association by 44% to 80% compared to PRSs restricting to variants below the genome-wide significance threshold. These PRSs were associated with BP levels across the lifespan and hypertension-related disease risk. Differences between those with low (1st decile), average (2nd to 9th decile), and high (10th decile) PRS appear to emerge in the first 3-5 years of life and be maintained throughout the life course. In adulthood, those with a high SBP PRS had increased risk for cardiovascular and chronic kidney disease, whereas the risk was reduced among those with low a SBP PRS.
Conclusions: Given the importance of exposure time in disease pathogenesis and the differences in BP between genetic risk levels throughout life, PRSs may help identify high-risk individuals prior to hypertension onset, thereby improving prevention and management of this public health challenge.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1719 Polygenic risk scores: Prediction of type 2 diabetes mellitus in multiplex families from the multiethnic GENNID study.

Authors:

L. Simon, J. Y. Wan, B. Wu, K. L. Edwards; Univ. of California, Irvine, Irvine, CA

Background: Type 2 diabetes mellitus (T2DM) is a major global health problem. Prevalence differs by ethnicity in the US, with Blacks and Hispanics having higher rates compared to Asians and Whites. While T2DM is a complex, polygenic disease with a strong genetic component, T2DM-associated common variants identified through GWAS have only explained a small proportion of T2DM heritability. Composite genetic risk measures such as polygenic risk scores (PRS) may enhance T2DM risk prediction. Most existing T2DM PRS are derived from European-dominant source populations and may not perform as well in diverse populations. We selected ethnic-specific PRS (adjusted for age, sex and BMI) constructed from multiethnic populations by Polfus et al (2021) and evaluated how these ethnic-specific PRS performed in our multiethnic GENNID families enriched for T2DM.
Methods: The GENNID study target population consists of 1651 subjects in 340 T2DM multiplex families from African American (AA), European American (EA), Japanese American (JA) and Mexican American (MA) families. PLINK and R were utilized for data QC of the GENNID dataset and PRS analyses. PRS Beta weights were derived from Polfus et al for all 582 genome-wide significant variants and were then used to generate PRS scores for each ethnic group. Using GLMMs with a family random effect, we assessed T2DM PRS prediction by also computing the AUC of ROC curves for each GENNID ethnicity and evaluated the impact of adjusting these models for age, sex and BMI.
Results: Of the 582 SNPs included in the ethnic-specific PRS constructed by Polfus et al, 537, 562, 419, and 549 overlapped in GENNID AA, EA, JA, and MA ethnic groups, respectively. T2DM PRS risk prediction models adjusted for age, sex and BMI in GENNID yielded AUCs of 89.3% (95% CI: 84.4-94.3%) for AA, 61.3% (55-67.7%) for EA, 65.7% (54.5-76.8%) for JA, and 60.6% (54.4-66.8%) for MA. Using only computed PRS (without modeling covariate adjustments) resulted in poor AUCs with 95% CIs (within 50%) when applied to each GENNID ethnic group. Compared to those reported by Polfus et al, all AUCs were lower in GENNID except for the adjusted AA model.
Conclusions: We applied an existing multiethnic T2DM PRS to an independent multiethnic sample of families enriched for T2DM to quantify T2DM risk prediction. As our unadjusted PRS models yielded very poor AUCs compared to our adjusted PRS models, it appears necessary to adjust for appropriate covariates to optimize the generalizability of T2DM prediction using PRS. Future research aimed at understanding how to enhance the generalizability and effectiveness of PRS across multiethnic populations is still needed to optimize T2DM genetic risk prediction.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1720 Polygenic Scores for DZ Twinning in Mothers of Naturally and ART Conceived Twins and Singletons

Authors:

D. Boomsma[1], N. Hubers[1], C. Page[2], H. Mbarek[3], C. Lambalk[4], L. Ligthart[5], R. Pool[1], J- . Hottenga[1], S. Håberg[6], E. Ehli[7], G. Willemsen[5], J. Harris[8], N. Martin[9], J. van Dongen[10]; [1]VU Univ. Amsterdam, Amsterdam, Netherlands, [2]4The Ctr. for Fertility and Hlth., Norwegian Inst. of Publ. Hlth., Oslo, Norway, [3]Qatar Genome, Doha, Qatar, [4]AUMC, Amsterdam, Netherlands, [5]VU Univ., Amsterdam, Netherlands, [6]Ctr. for Fertility and Hlth., Norwegian Inst. of Publ. Hlth., Oslo, Norway, [7]Avera Inst. for Human Genetics, Sioux Falls, SD, [8]Norwegian Inst Pub Hlth, Silver Spring, MD, [9]Queensland Inst Med Res, Brisbane, Australia, [10]VU Amsterdam, Amsterdam, Netherlands

Natural dizygotic (DZ) twinning results from a double ovulation and has a genetic component. Endocrinology studies in mothers of DZ twins observed higher levels of follicle stimulating hormone (FSH) and *FSHB* was, together with *SMAD3*, one of the first loci identified as significant in genome-wide association studies (GWAS) of having DZ twins. Several additional loci such as *FSHR*, and *GNRH1* have recently been found that confirm the role of genes involved with female endocrinology in fertility. A polygenic score (PGS) for spontaneous DZ twinning explained ~1.5% of the phenotypic variance when comparing mothers of naturally conceived DZ twins to controls. Here we compare the PGS in mothers of naturally conceived DZ twins with other twin mothers in the Netherlands twin register (NTR) and in the Norwegian Mother, Father and Child Cohort Study (MoBa). In the NTR, we contrast the PGSs in mothers who conceived DZ or monozygotic (MZ) twins naturally and those who conceived DZ or MZ twins through artificial reproductive techniques (ART) and test differences with logistic regression. Mothers of naturally conceived DZ twins differed in age, body composition, parity and smoking behavior compared to the other groups. The PGSs, both corrected and uncorrected for these demographic and lifestyle traits, were higher in mothers of naturally conceived DZ twins (N=1423) compared to all other twin mothers (N=2263) or compared to fathers of twins (N=2158). For ART mothers whose fertility treatment involved hormones without IVF (N=187) the PGS was significantly lower than for the other three groups (DZ-IVF and MZ mothers). We are currently replicating and extending these analyses in mothers of twins and mothers of singletons from MoBa (N = ~ 95.000 mothers). Our study shows that mothers of spontaneously conceived DZ twins are genetically different and that a PGS for DZ twinning may also be associated with infertility treatment.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1721 Pooled and sex-specific European-ancestry GWAS meta-analyses of restless legs syndrome identify 164 risk loci, enabling advanced risk prediction, causal inference exploration, and treatment target identification

Authors:

B. Schormair[1,2], C. Zhao[2], S. Bell[3], M. Didriksen[4], M. S. Nawaz[5], P. Nandakumar[6], International EU-RLS-GENE Consortium, INTERVAL study, The Danish Blood Donor Study, The D.E.S.I.R. study group, 23andMe Research Team, D. B. Rye[7], G. Rouleau[8], K. Berger[9], H. Stefansson[10], H. Ullum[11], K. Stefansson[10], D. Hinds[6], E. Di Angelantonio[12], K. Oexle[2], J. Winkelmann[2]; [1]Klinikum rechts der Isar der TU München, Inst. of Human Genetics, Munich, Germany, [2]Helmholtz Munich, Inst. of Neurogenomics, Munich, Germany, [3]Univ. of Cambridge, Dept. of Clinical NeuroSci.s, Cambridge, United Kingdom, [4]Dept. Clinical Immunology, Copenhagen Univ. Hosp., RigsHosp.et, Copenhagen, Denmark, [5]deCODE Genetics/Amgen, Reykjavik, Iceland, [6]23andMe, Sunnyvale, CA, [7]Dept. of Neurology and Program in Sleep, Emory Univ., Atlanta, GA, [8]Montreal Neurological Inst.-Hosp., Montreal, QC, Canada, [9]Inst. of Epidemiology and Social Med., Univ. of Münster, Münster, Germany, [10]deCODE genetics, Reykjavik, Iceland, [11]Statens Serum Inst., Copenhagen, Denmark, [12]Univ. of Cambridge, Natl. Inst. for Hlth.and Care Res. Blood and Transplant Res. Unit in Donor Hlth.and Behaviour, Cambridge, United Kingdom

Restless legs syndrome (RLS) is a prevalent chronic sensorimotor disorder, affecting up to 10% of the elderly population in Europe and North America. It impacts severely on mental and physical health by disrupting sleep and daytime functioning and being frequently comorbid with metabolic, cardiovascular, and psychiatric diseases. Delayed diagnosis, sub-optimal therapy and the chronic nature of the disease further worsen the situation of affected individuals. Progress towards better health care for RLS patients is impeded by the limited knowledge of the underlying pathology and interrelationships with other disorders.

To advance disease prediction and find novel entry points for treatment, we performed a meta-analysis of genome-wide association studies (GWAS) in 1,589,569 individuals of European ancestry using N-GWAMA (116,403 cases and 1,473,166 controls). Our analyses also included the first sex-stratified GWAS and first genetic investigation of the X chromosome in RLS. Functional interpretation of the GWAS results was done by gene annotation and prioritization (using DEPICT, FUMA, MAGMA, MetaXcan), pathway and cell type enrichment analyses (DEPICT, MAGMA, CELLEX, CELLECT), as well as extensive genetic correlation (LDSC) and Mendelian randomization analyses (LHC-MR, IVW-MR). Finally, we evaluated the performance of different models integrating genetic and non-genetic effects in RLS risk prediction, using traditional polygenic risk scores as well as machine learning approaches (PRSice; GLM, DNN and RF models in H2O AutoML and PyTorch).

The pooled meta-analysis of both sexes increased the number of known risk loci to now 164 loci comprising 196 independent lead SNPs, with the first sex-specific GWAS on RLS revealing largely overlapping genetic predispositions of the sexes (r_g=0.96). Annotation of genes in risk loci by mapping against the druggable genome identified 13 potential candidates targeted by existing compounds for expanding treatment options, e.g. *GRIA1* and *GRIA4* encoding glutamate receptor subunits. Mendelian randomization suggested RLS as a causal risk factor of diabetes and hypertension. Among the tested risk prediction models, machine-learning approaches combining genetic and environmental information performed best (AUC=0.82-0.91). Our study identified targets for drug development, prioritized potential causal relationships between RLS and relevant comorbidities and risk factors for follow-up, and provided evidence that gene-environment interaction is likely highly relevant for RLS risk prediction.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1722 Population genetics analysis of *CLDN16* and *CLDN19* of familial hypomagnesemia with hypercalciuria and nephrocalcinosis.

Authors:

I. Patel[1], K. Lovrenert[1], M. Baum[2], F. Hildebrandt[3], D. Bodner[4], F. Schumacher[5], C-H. Wu[6]; [1]Case Western Reserve Univ., Cleveland, OH, [2]Boston Children's Hosp., Boston, MA, [3]Boston Children's Hosp, Boston, MA, [4]Univ. Hosp., Cleveland, OH, [5]Case Western Reserve Univ, Cleveland, OH, [6]Univ. Hosp./Case Western Reserve Univ, Cleveland Heights, OH

Familial hypomagnesemia with hypercalciuria and nephrocalcinosis (FHHNC) is an autosomal recessive progressive renal disorder characterized by excessive urinary calcium and magnesium excretion. FHHNC results in progressive loss of kidney function, typically presenting during early childhood. Current therapeutic options consist of oral magnesium supplementation and thiazide diuretics. Neither treatment has significant effects and about 50% of FHHNC cases result in the need for renal replacement therapy as early as 20 years of age. Efforts to uncover the genetic etiologies of FHHNC have resulted in the identification of two causal genes: *CLDN16* (Claudin 16, also known as Paracellin-1), located on chromosome 3q28, and *CLDN19* (Claudin 19), located on chromosome 1p34.2. Proteins encoded by *CLDN16* and *CLDN19* are integral membrane proteins, a component of tight junction strands, and primarily expressed on the Loop of Henle. As reports investigating the genetic epidemiology are relatively scarce, this investigation works to estimate the genetic prevalence of FHHNC through the *CLDN16* and *CLDN19* genes and compare it to the clinical prevalence.

The 1000 Genomes Project (1KG) was utilized to identify genetic variants in the general population of 2,504 individuals. The Human Gene Mutation Database (HGMD) was used to identify pathogenic variants of *CLDN16* and *CLDN19*. Overlapping variants procured from both databases were evaluated. The pathogenic allele frequency, carrier rate, and affected rate were calculated and estimated based on the Hardy-Weinberg equilibrium. 1KG contained 625 and 140 variants for *CLDN16* and *CLDN19*, respectively. HGMD contained 69 and 21 unique pathogenic variants of *CLDN16* and *CLDN19* respectively. The intersection between HGMD and 1KG resulted in no pathogenic variants of *CLDN16* and two pathogenic variants of *CLDN19* across three heterozygous carriers (pathogenic variant allele frequency of 0.06%). The expected carrier rates were 0 for *CLDN16*, and 1 in 835 for *CLDN19*, and for the cumulative rate combining both genes. The expected affected rates were 0 for *CLDN16*, 1 in 696,668 for *CLDN19*, and for the cumulative rate combining both genes.

FHHNC is a rare disorder and the population prevalence is unknown. Only ~100 patients with a *CLDN16* causative variant ~70 patients with a *CLDN19* causative variant have been described in the literature. This investigation found no *CLDN16* related carriers and three *CLDN19* related carriers. The genetics findings are not consistent with the current clinical observations, which requires further investigation.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1723 † Population-scale analysis of the trinucleotide repeat expansion in the Huntingtin gene (HTT) from 854,251 human exomes

Authors:

V. Pounraja[1], S. O'Keeffe[1], X. Bai[1], F. Grau[2], S. Wolf[1], R. Sundaramoorthy[2], M. D. Kessler[1], N. Parikshak[1], W. Salerno[1], G. Coppola[1], DiscovEHR Collaboration, Penn Medicine BioBank, Mayo Clinic Project Generation, S. Gelfman[1]; [1]Regeneron Genetics Ctr., Tarrytown, NY, [2]Regeneron Pharmaceuticals, Tarrytown, NY

Background: A pathological trinucleotide CAG repeat expansion (TRE) in the Huntingtin gene (HTT) causes Huntington's disease (HD) with variable risk (36-39 lower risk, ≥40 higher risk). However, population scale characterization of normal and expanded repeat length is lacking. Here, we leveraged whole-exome sequencing (WES) data to call HTT repeat expansions in the UK Biobank and five additional large EHR cohorts. Using these calls, we estimated prerisk/risk prevalence in various ancestries, and validated the results both experimentally and through association analyses.
Methods: We used gangSTR to call repeat expansions across 854,251 samples, including 762,895 samples of European ancestry (89%), 41,130 samples of African ancestry (5%), and 31,784 samples of South Asian ancestry (4%). To estimate the accuracy of this calling approach, we experimentally validated 27 HD cases and 27 randomly chosen controls using a PCR-based approach (Asuragen AmplideX PCR kit). We assessed various sequence derived metrics and applied strict quality control criteria to the full cohort, followed by comparison of the distributions of repeat lengths across cohorts and ancestries. Last, we binarized repeat length based on either prerisk/pathogenic (≥36 & ≥40) or percentile (Top 1%/0.1%/0.01%) cut-offs and associated them with case-control status using 49 HD cases and 558,736 controls.
Results: The sensitivity and specificity of the predictions estimated based on PCR validation was 74% and 100%, respectively. Four HD cases predicted with non-pathogenic expansions were validated to carry normal repeat length. The prevalence of pre-risk (0.23%; n=1,916) and pathogenic expansions (0.053%; n=437) were higher than the prevalence of HD in general population (0.003-0.006%), yet consistent with literature. This excess of carriers with expanded repeats could be attributed to false positives or patients at risk of developing HD. We also found the prevalence of prerisk expansions to be higher in Europeans compared to south Asians (0.24% vs. 0.16%; Fisher's exact p=7.4e-3). A meta-analysis showed that the top 0.1% repeat length of 38 associated most strongly with HD (OR=213.4; p=8.9e-37), compared with the risk threshold cutoff of ≥40 that had the largest effect size (OR=331.4, p=1.9e-32).
Conclusion: This is the largest population-scale characterization of HTT repeat expansions called using whole exome sequencing data. Our results demonstrate that normal and pathogenic repeat expansions can be identified with high specificity and adequate sensitivity and help refine population prevalence and penetrance estimates.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1724 Potentialmodificationsof *ABCG2 rs4148155* and *SLC22A12 rs75786299* on gout and nephrolithiasis: A hospital-based, case-control study.

Authors:

C-H. Lin, C-T. Lin, Y-M. Chen; Taichung Veterans Gen. Hosp., Taichung, Taiwan

Objective: A strong connection between gout and specific genetic variations, namely ABCG2 rs4148155 and SLC22A12 rs75786299, has been established. Gout, a condition characterized by excessive levels of uric acid, is a significant risk factor for the formation of urate stones. However, the precise interplay between ABCG2 rs4148155 and SLC22A12 rs75786299 in relation to gout and nephrolithiasis remains to be explored. Consequently, the objective of this study was to examine the correlation between these genetic variants and the occurrence of gout and nephrolithiasis.Methods: This case-control study utilized data from the Taiwan Precise Medicine Initiative (TPMI) and involved 35,280 adults. The genotypes *ABCG2 rs4148155* and *SLC22A12 rs75786299* were examined, and multivariable logistic regression was employed to analyze the adjusted odds ratios (ORs) of gout and nephrolithiasis Results: The study revealed that the rs4148155 G allele had a frequency of 33%, while the rs75786299 A allele had a frequency of 2.4%. Both the ABCG2 rs4148155 (GG and AG) and SLC22A12 rs75786299 (GA) polymorphisms showed significant associations with gout and nephrolithiasis. The risk of gout was 1.60 (1.51-1.69) for individuals carrying a single copy of the rs4148155 G allele and 3.22 (2.66-3.90) for individuals carrying both the rs4148155 G and rs75786299 A alleles. Similarly, the risk of nephrolithiasis was 1.09 (1.02-1.17, p=0.01) for single carriers of the rs4148155 G allele and 1.40 (1.09-1.81, p=0.009) for individuals carrying both the rs4148155 G and rs75786299 A alleles. Conclusion: This study supports an additive model for the development of gout and nephrolithiasis. Carriers of both the rs4148155 G and rs75786299 A alleles face an increased risk of developing these conditions. The findings underscore the importance of implementing precision healthcare strategies to address gout and nephrolithiasis in individuals carrying the rs4148155 G and rs75786299 A alleles.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1725 Precision risk profiling of Type 1 Diabetes individuals in the INNODIA, Danish and UK Biobank cohorts, using HLA imputation and haplotype analysis.

Authors:

S. Kaur[1], C. Brorsson[2], T. S. Ahluwalia[1], C. H. Johansen[2], S. Brunak[2], F. Pociot[1]; [1]Steno Diabetes Ctr. Copenhagen, Herlev, Denmark, [2]Univ. of Copenhagen, Copenhagen, Denmark

Background and aims: Genome wide association studies (GWAS) in Type 1 Diabetes (T1D) have identified ~152 susceptibility loci so far. One of the major risk locus in T1D is the major histocompatibility complex (MHC) region on chr6 encoding the human leucocyte antigen (HLA) proteins. Class II HLA DR and DQ loci contribute to an overall genetic risk of 30-50%. The MHC region is highly polymorphic with more than 6,000 classical HLA alleles and amino-acid polymorphisms. The accurate genotyping of this region is possible but highly expensive. In this study, we imputed the classical alleles and amino acid polymorphisms at class I and II HLA loci in the newly diagnosed type 1 diabetes (T1D) individuals from INNODIA and Danish Biobank cohorts to identify high-risk HLA haplotypes.

Materials and methods: The HLA imputation was performed on 695 INNODIA samples (301 T1D and 394 unaffected family members) originally genotyped on the Affymetrix UK Biobank Axiom Array using SNP2HLA and the T1D Genetics Consortium (T1DGC) reference panel. The T1DGC reference panel contains 5,868 SNPs (genotyped with Illumina Immunochip) and 4-digit classical HLA types for HLA-A, -B, -C, -DPA1, -DPB1, -DQA1, -DQB1 and -DRB1 for 5,225 unrelated individuals of European descent. The Danish Biobank samples consisting of 2,853 newly diagnosed T1D and 1,983 healthy control individuals genotyped on ImmunoChip were also imputed for HLA alleles and amino-acid polymorphisms. From the UK Biobank, genotyping data for 3,397 T1D individuals and 7476 healthy controls (Caucasians) were extracted for validation studies. The genotype concordance and imputation accuracy for the INNODIA samples were calculated for SNP2HLA imputed alleles and Axiom HLA Imputation.

Results and Conclusions: Using the T1DGC reference panel, we imputed 8,961 makers (1,390 haplotypes) for the 695 individuals (301 cases, 394 unaffected family members) genotyped in the INNODIA cohort. A total of 8,667 markers achieved high imputation accuracy (r^2 dosage>0.8) that included 348 two and four- digit HLA alleles and 384 amino acid positions (1,130 residues). Previous studies have shown that risk conferring residues at three positions (DQB1_aa57_alanine, DRB1_aa71_lysine and DRB1_aa13_histidine) tags the well-established DRB1*03 and DRB1*04 haplotypes known to confer the strongest risk. Our preliminary results showed strongest association for DQB1_aa57(AD|AV|AS), DRB1_aa13(HY|HF|HG) and DRB1_aa11 (VG|VL|VD) haplotypes based on a sliding window approach of 3-SNP haplotypes in both INNODIA and Danish Biobank cohorts. Further analyses are ongoing to validate the findings in the UK Biobank cohort.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1726 Predicting future obesity with the polygenic score in a Japanese working-age cohort

Authors:

K. Tanaka[1], N. Itokawa[1], T. Takahashi[2], M. Koido[1], N. Asanoma[3], N. Shiomi[4], A. Fujino[4], Y. Kamatani[1], M. Seyama[5], T. Suzuki[2], Y. Murakami[2,6]; [1]Lab. of Complex Trait Genomics, Dept. of Computational Biology and Med. Sci., Graduate Sch. of Frontier Sci., The Univ. of Tokyo, Tokyo, Japan, [2]Project Div. of Genomic Med. and Disease Prevention, The Inst. of Med. Sci., The Univ. of Tokyo, Tokyo, Japan, [3]NTT Smart Data Sci. Ctr. and NTT Computer and Data Sci. Lab., Nippon Telegraph and Telephone Corp., Tokyo, Japan, [4]Bio-Med. Informatics Res. Ctr. and NTT Basic Res. Lab., Nippon Telegraph and Telephone Corp., Tokyo, Japan, [5]Med. Business Planning Office, Nippon Telegraph and Telephone Corp., Tokyo, Japan, [6]Div. of Molecular Pathology, The Inst. of Med. Sci., The Univ. of Tokyo, Tokyo, Japan

Obesity is a global health threat and is associated with age-related changes. Body mass index (BMI) is the most widely used measure of obesity. However, the genetic effects on BMI changes with age during adulthood remain less understood. We evaluated BMI changes using polygenic score (PGS) in a working-age cohort comprised of employees of Nippon Telegraph and Telephone Corporation (NTT) group companies who underwent annual health examinations. This longitudinal cohort consisted of up to 36,389 participants with whole genome genotypes, including data from 1990 to 2022 (range in age: 17 - 67 years). The data of all participants was collected with written informed consent. After stringent quality control, genotype data were imputed using the 1000 Genomes Project reference panel (phase3v5, N = 2504). We constructed PGS using PRS-CSx for 1,015,557 SNPs from GWAS of BioBank Japan (n=158,284) and that of Tohoku Medical Megabank (n=47,070) summary statistics. The association of BMI-PGS and severe obesity (BMI>35) was evaluated by Kaplan-Meier curves and a multivariate Cox proportional hazards regression analyses. Among individuals in the top decile of BMI-PGS, 7.2% developed severe obesity, compared with only 2.4 % of those in the remaining samples (log-rank test $p=1.75x10^{-101}$). As a result of Cox proportional hazards regression adjusted for baseline BMI, the effect of PGS was hazard ratio of 1.22 per 1-SD ([95%CI 1.14-1.32], $p=7.75x10^{-7}$).We also estimated a baseline and linear change over time of individual BMI trajectory using the longitudinal linear mixed effects model, where the baseline and change were modeled as random intercept and slope, respectively. The estimated random slope was in the range of [-0.96-1.61]. Including fixed effects, 76% of individuals had a positive slope and a tendency to gain weight. Further, we evaluated the effect of BMI-PGS on the estimated random effects using linear regression adjusted for sex, age, and genotype principal components (PCs). This analysis identified the significant effect for the baseline (beta per 1-SD of PGS increase 0.30 [95%CI 0.29-0.31], $p=3.40x10^{-718}$) as expected, as well as the linear change (beta per 1-SD of PGS 0.079 [95%CI 0.068-0.089], $p=1.22x10^{-49}$), which further shed light on the genetic architecture of age-related BMI changes.In summary, we have demonstrated that BMI-PGS not only strongly affects baseline BMI but also affects BMI changes. These results suggest that BMI-PGS can predict future obesity independently of baseline BMI.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1727 Predicting mechanisms of action at genetic loci associated with discordant effects on type 2 diabetes and abdominal fat accumulation.

Authors:

Y. Aberra[1], L. Ma[2], J. L. Bjorkegren[2], M. Civelek[1]; [1]Univ. of Virginia, Charlottesville, VA, [2]Icahn Sch. of Med. at Mount Sinai, New York, NY

Metabolic syndrome (MetSyn) is a cluster of dysregulated metabolic conditions that occur together to increase the risk for cardiometabolic disorders such as type 2 diabetes (T2D). One key condition associated with MetSyn, abdominal obesity, is measured by computing the ratio of waist-to-hip circumference adjusted for the body-mass index (WHRadjBMI). WHRadjBMI and T2D are complex traits with genetic and environmental components, which has enabled genome-wide association studies (GWAS) to identify hundreds of loci associated with both. Statistical genetics analyses of these GWAS have predicted that WHRadjBMI is a strong causal risk factor of T2D and that these traits share genetic architecture at many loci. To date, no variants have been described that are simultaneously associated with protection from T2D but with increased abdominal obesity. Here, we used colocalization analysis to identify genetic variants with a shared association for T2D and abdominal obesity. This analysis revealed the presence of five loci associated with discordant effects on T2D and abdominal obesity. The alleles of the lead genetic variants in these loci that were protective against T2D were also associated with increased abdominal obesity. We further used publicly available expression, epigenomic, and genetic regulatory data to predict the effector genes (eGenes) and functional tissues at the 2p21, 5q21.1, and 19q13.11 loci. We also computed the correlation between the subcutaneous adipose tissue (SAT) expression of predicted effector genes (eGenes) with metabolic phenotypes and adipogenesis. We proposed a model to resolve the discordant effects at the 5q21.1 locus. We find that eGenes gypsy retrotransposon integrase 1 (*GIN1*), diphosphoinositol pentakisphosphate kinase 2 (*PPIP5K2*), and peptidylglycine alpha-amidating monooxygenase (*PAM*) represent the likely causal eGenes at the 5q21.1 locus. Taken together, these results are the first to describe a potential mechanism through which a genetic variant can confer increased abdominal obesity but protection from T2D risk. Understanding precisely how and which genetic variants confer increased risk for MetSyn will develop the basic science needed to design novel therapeutics for metabolic syndrome.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1728 Predicting Neurodegenerative Disease Risk using Feature Propagation on SNP Network with Brain Imaging Endophenotypes

Authors:

D-g. Lee, S-H. JUNG, L. Shen, D. Kim; Univ. of Pennsylvania, Philadelphia, PA

Background: The structural characteristics of the brain, specifically the decrease of individual gray (or white) matter volumes, provide valuable insights into brain function and cognitive decline, including the development of neurodegenerative diseases (NDDs). In addition, genetic factors can play a significant role in changes in brain volumes, influencing biological activities and interacting in complex ways. In this study, we aim to investigate the relationship between genetic factors, structural brain volume, and the risk of NDDs, such as dementia. Method: The proposed method utilizes a SNP network to capture interactions between genetic variants based on linkage disequilibrium scores. The network represents the complex relationships among genetic factors. Feature propagation, a technique that refines genetic variant information based on graph-based semi-supervised learning (GSSL), is then employed to predict NDD risk scores. Meanwhile, brain imaging endophenotypes, following the concept of transcriptome-wide association studies (TWAS), are used to enhance prediction power and identify significant regions affecting specific diseases. Moreover, a generative artificial intelligence model is applied to produce pseudo-imaging endophenotypes for samples with only genetic information. Result: The proposed method was applied to data including (i) both genetic and brain imaging information from two datasets: UK Biobank (UKBB), Alzheimer's Disease Neuroimaging Initiative (ADNI), and (ii) only genetic information from Alzheimer's Disease Sequencing Project (ADSP) dataset. The results were cross-validated from the two datasets (UKBB and ADNI), and a generative model was applied to ADSP data. The final prediction results were compared with other models, including the polygenic risk score-continuous shrinkage (PRS-CS) model. Compared to the other methods, the proposed method demonstrated superior prediction performance and revealed genetic variants and brain regions that impact NDDs. Conclusion: By integrating genetic information and brain imaging data, this study achieved improved disease risk prediction and a deeper understanding of the phased mechanism of NDDs associated with genetic variants and brain lesions. Furthermore, the insights gained into the relationship between genetics, brain structure, and diseases are expected to contribute to advancing our knowledge of disease progression and inform future research and clinical strategies for the prevention and management of these conditions.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1729 Prediction of Alzheimer's Disease Biomarkers Using Genomics and Cerebrospinal Fluid Metabolomics

Authors:

M. Francia[1], N. Ramesh[1], T. Boltz[1], L. M. Reus[1], M. Bot[1], W. F. van der Flier[2], P. Visser[2], S. van der Lee[2], B. M. Tijms[2], C. E. Teunissen[3], L. Olde Loohuis[1], R. Ophoff[1]; [1]Univ. of California Los Angeles, Los Angeles, CA, [2]Vrije Univ.it Amsterdam, Amsterdam, Netherlands, [3]Amsterdam UMC, Amsterdam, Netherlands

Metabolite measurements from the cerebrospinal fluid (CSF) have been employed as diagnostic tools for neurodegenerative disorders, such as Alzheimer's disease (AD). AD's pathogenesis is identified by lowered measured levels of CSF amyloid Beta 1-42 (AB), and increased levels of phosphorylated tau (pTau) and total tau (tTau). Due to their direct contact with the brain, metabolites from the CSF are relatively close to these phenotypes and could be used to study biological pathways and signatures associated with neurological disorders. Furthermore, CSF metabolites could help elucidate the effect of the genetic architecture of these disorders, as genome wide association studies (GWAS) have shown that metabolic features can have a heritable component. While the *APOE* e4 allele has been implicated as the main AD risk locus, other sites of the genome have been found to be associated with AD, indicating a complex, polygenic architecture. Therefore, CSF metabolites are suitable to identify biological pathways impacted by both AD genetic factors and its hallmark biomarkers. For this purpose, we collected more than 5,000 CSF metabolite measurements from AD and mild cognitive impairment patients, and healthy individuals from the Amsterdam Dementia cohort (ADC), as well as imputed genotype data. Using the most recent AD GWAS to calculated a polygenic risk score (PRS) for AD that adds the effect of the *APOE* alleles as a weighted sum, we found a modest prediction of pTau and tTau by the PRS in linear models after correcting for age and sex, however this effect was mainly driven by *APOE* e4 allele counts, and is smaller in comparison to the large effect size that is seen for AB. No CSF metabolites were found to be significantly predicted by the PRS or *APOE* e4 allele counts. We then used an elastic net model to predict AD biomarkers levels using the measured CSF metabolites. This approach identified 18 and 20 metabolites that predicted CSF pTau and CSF tTau levels respectively, among them Cholesterol, Isothreonic acid, and Nicotinoylglycine. Within these CSF metabolites, we did not find any that predicted CSF AB levels. These results indicate that these biomarkers show independent, but biologically correlated aspects of AD pathogenesis. AB reflects the influence of the *APOE* e4 allele while pTau and tTau are associated with ongoing cellular processes. This is consistent with prior reports that sporadic AD pathogenesis could be driven by a dual-cascade. Further analysis on a separate cohort of healthy individuals, in which the same CSF metabolites were measured, is currently ongoing to replicate these results.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1730 Prediction of future coronary artery disease risk in children from the Growing Up in Singapore Towards Healthy Outcomes birth cohort

Authors:

F. Ng[1,2,3], L. Ling[4,5], J. Choo[6], D. Wang[7,8,9], M. Kumar[8,9], L. Ang[8,9], R. Dorajoo[10,11], M. Chan[4,5], A. Low[4,5], W-P. Koh[12,8], D. Ooi[1], Y. Lee[1,8,2], X. Chang[13,1,2], C-K. Heng[1,2,3]; [1]Dept. of Paediatrics, Yong Loo Lin Sch. of Med., Natl. Univ. of Singapore, Singapore, Singapore, [2]Khoo Teck Puat – Natl. Univ. Children's Med. Inst., Natl. Univ. Hlth.System, Singapore, Singapore, [3]Cardiovascular Disease Translational Res. Programme, Yong Loo Lin Sch. of Med., Natl. Univ. of Singapore, Singapore, Singapore, [4]Dept. of Med., Yong Loo Lin Sch. of Med., Natl. Univ. of Singapore, Singapore, Singapore, [5]Dept. of Cardiology, Natl. Univ. Heart Ctr., Natl. Univ. Hlth.System, Singapore, Singapore, [6]Cardiology Service, KK Women's and Children's Hosp., Singapore, Singapore, [7]Natl. Heart and Lung Inst., Imperial Coll. London, London, United Kingdom, [8]Singapore Inst. for Clinical Sci., Agency for Sci. Technology and Res., Singapore, Singapore, [9]Bioinformatics Inst., Agency for Sci. Technology and Res., Singapore, Singapore, [10]Genome Inst. of Singapore, Agency for Sci., Technology and Res., Singapore, Singapore, [11]Hlth.Systems and Services Res., Duke-NUS Med. Sch. Singapore, Singapore, Singapore, [12]Hlth.y Longevity Translational Res. Programme, Yong Loo Lin Sch. of Med., Natl. Univ. of Singapore, Singapore, Singapore, [13]Dept. of Infectious Diseases, The Univ. of Melbourne, The Peter Doherty Inst. for Infection and Immunity, Melbourne, Australia

Studies have shown that atherosclerosis begins during childhood and interventions for coronary artery disease (CAD) should commence early to reduce the incidence of CAD in adulthood. To identify individuals at high CAD risk for interventions, risk assessments would have to be conducted. During childhood, the threshold levels of CAD risk factors that are well established in adults, such as those for plasma lipid levels, are not applicable for children. One way to predict CAD risk in children would be to use polygenic risk scores (PRS) that are constructed from single-nucleotide polymorphisms associated with CAD. In this study, we used various PRS of CAD risk in Chinese adult populations that were reported in two previous studies to determine if they could predict measures of early cardiovascular health at ages 6, 8 and 10 in children of the Growing Up in Singapore Towards healthy Outcomes (GUSTO), a multi-ethnic birth cohort from Singapore. One of the PRS used was consolidated from CAD and eight CAD related traits PRS and thus referred to as metaPRS. However, since the GUSTO children are still young, their eventual CAD outcomes are unknown. To overcome this, we constructed a CAD Risk Index (CRI) to estimate their risk of CAD using gender, body mass index, plasma high-density lipoprotein levels, plasma low-density lipoproteins levels, triglyceride glucose index and carotid-femoral pulse wave velocity. The CRI is used as an outcome variable for evaluating the performance of the various PRS. We found the metaPRS to be significantly associated with CRI in Chinese at ages 6 (beta 0.304, SE 0.140, p 0.031), 8 (beta 0.327, SE 0.126, p 0.010) and 10 (beta 0.347, SE 0.149, p 0.021) and in pooled ethnic groups at ages 8 (beta 0.203, SE 0.093, p 0.029) and 10 (beta 0.251, SE 0.110, p 0.023) but not at age 6 (p 0.068). We also found childhood smoking exposure to be significantly associated with the CRI at age 6 (beta 0.031, SE 0.009, p 5.80×10^{-4}). We have validated the metaPRS and found it to be significantly associated with CAD in the Chinese adult population in Singapore. We concluded that PRS is a useful tool for predicting future CAD risk in children.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1731 Predictive modeling of alcohol consumption using blood-based biomarkers: A machine learning approach

Authors:

M. Hassan[1], A. Gentry[1], R. Peterson[2,1], B. Webb[3,1]; [1]Virginia Commonwealth Univ., Richmond, VA, [2]SUNY Downstate Hlth.Sci. Univ., Brooklyn, NY, [3]RTI Intl., Research Triangle Park, NC

Methods for predicting alcohol consumption are considered important tools for identifying individuals with a higher risk of Alcohol Use Disorder (AUD). Developing quick, accurate, and reliable tools for predicting alcohol use is important for research and translational applications. While self-report and structured diagnostic interviews serve these goals, biological-based indices may provide a complementary approach.Towards this goal, we applied machine learning (ML) methods to predict Alcohol Consumption (AC) and other outcomes from blood-based biomarkers. This research has been conducted using the UK Biobank Resource (ID30782). Five different phenotypic outcomes were predicted using 303 objective markers including 51 blood biochemistry assays, and 249 Nuclear Magnetic Resonance (NMR) metabolites, as well as covariates known to influence these outcomes including age, sex, and statin use. Outcomes included alcoholic Drinks per Week (DPW), height, Body Mass Index (BMI), Body Fat Percentage (BF%), and predicted Major Depression Symptoms (MDDsx). Four machine learning models were evaluated including LASSO, Ridge regression, Gradient Boosting Machines (GBM), and MBOOST. Performance was assessed across ML models using adjusted r-squared, Mean Absolute Error (MAE), and LDSC-based genetic correlation between observed and predicted measures in independent samples. Results show that LASSO achieved better performance compared to Ridge regression when solely trained on age and sex as features. The r-squared results for height, BF%, BMI, AC, and MDDsx are 0.53, 0.48, 0.01, 0.07, and 0.07 respectively. When adding blood-based measures, the results are improved for BF%, BMI, and DPW to 0.71, 0.46, and 0.33, respectively, but there is no significant change for height (0.55) or MDDsx (0.002). We also observed significant genetic correlation between observed and predicted AC. Compared to traditional methods for AC prediction, ML methods provide explainable models that are able to identify blood-based biomarkers and other features that are highly connected to AC. In addition, combination of ML and blood-based biomarkers presents an opportunity to predict patterns of alcohol consumption. A potential application of these predictions is in ever larger genetic studies where self-report or structured interviews related to AC may be absent but biomarkers may be available. Importantly, prediction of AC from blood-based measures represents a clinical opportunity to identify individuals for whom additional screening using structured interviews may be warranted.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1732 Pregestational Diabetes Induces GATA4 Expression and Activity in Neonatal Heart and Associated with Cardiac Fibrosis

Authors:

S. Mattapally; Mattapally Technologies, Birmingham, AL

Diabetes, cardiovascular disease, and congenital heart defects all run in families and are caused by a mix of hereditary and environmental factors. Most diseases are family, which shows a genetic predisposition, although environmental, stochastic, and epigenetic factors are also important. Enhancers are genomic sequences that are essential for controlling the levels of tissue-specific gene expression. As a result of chromosomal rearrangements, genetic diversity within enhancers, or epigenetic manipulation, an increasing number of disorders are connected to poor enhancer activity. Chakravarti group reported PAX3, is not expressed at this embryonic stage, there has been no discernible transcriptional shift in the human fetal gut development program controlled by RET Enhancers TFs, GATA2, RARB, and NKX2-5. The goal of the current investigation was to learn more about the cardiac consequences of pregestational diabetes. Here, we investigated if diabetes a pathological alteration in the newborn heart and creates an adverse environment for embryonic and fetoplacental development. Two groups of female Sprague Dawley rats (weighing 200-250g each) were created. (8 people per group). While Type 2 diabetes (Dia) was developed by nicotinamide and then a single intraperitoneal injection of 50 mg/kg streptozotocin (STZ), control (Con) rats were fed a normal diet. Following the diagnosis of diabetes, a trio breeding (two Con/Dia females and one Con male) was started to produce postnatal neonatal pups that were two days old from both a pregestationally diabetic and non-diabetic mother. In order to conduct molecular biology, biochemical, and histopathological research, newborn hearts were harvested and kept at -80^{o}C. In the newborn heart of the Dia group, we saw a significant (p 0.05) rise in lipid peroxidation and a significant (p 0.05) fall in endogenous antioxidants like SOD and GSH in a con group. Only neonates delivered to pregestationally diabetic mothers (Dia) were shown to have a substantial cardiac fibrosis in addition to oxidative stress as compared to controls. GATA4, NKX3-2, and SOX2 transcription factor activity in the neonatal heart of the Dia group increased more than twofold. We performed real-time PCR for mRNA quantification and western blot for protein expression to determine if enhanced transcription factor activity is linked to increased expression of mRNA, protein, or both. In neonatal diabetic hearts, we performed western blot for the genes GATA4, NKX2.5, SOX2, RB, and FOXO3A. Upcoming genomic technologies will be used in our research to uncover connections.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1733 Prenatal nutritional environment is associated with late-life cognition in the Health and Retirement Study, a natural experiment

Authors:

E. Vasiljevic[1], C. Engelman[2]; [1]Univ. of Wisconsin - Madison, Madison, WI, [2]Univ. of Wisconsin Madison, Madison, WI

INTRODUCTION: Early-life exposures are important to several late-life health outcomes. We sought to study the effect of an *in utero* nutritional environment and its interaction with Alzheimer's disease (AD) genetic risk on late-life cognitive function.
METHODS: We used a natural experiment created by the pellagra epidemic, a nutritional disease caused by a vitamin B_3 deficiency, in the early 20^{th} century, to evaluate the association between the *in utero* nutritional environment and late-life cognitive function in a subsample of Health and Retirement Study participants with data for all measures (N = 17,554). Pellagra mortality rates for a person's state of birth and gestation year were uses as a proxy for the prenatal nutritional conditions. Using mixed effects regression models, we tested the association between the *in utero* pellagra mortality rate and a 27-point global cognition score. We also stratified analyses by race and gender, and we evaluated whether the *in utero* exposure could modify the AD polygenic score's (PGS) effect on global cognition.
RESULTS: *In utero* pellagra mortality rates were significantly associated with late-life global cognition ($\beta = -0.025$, 95% confidence interval: −0.035 to −0.015). Within race and gender strata, the effects of *in utero* pellagra mortality rates on global cognition were no longer significant in all but two groups (Black females and females with other non-White racial backgrounds). The interaction between the *in utero* exposure and the AD PGS was significant. The genetic effect on cognition amplified with increasing (progressively worse) *in utero* exposure levels.
DISCUSSION: These associations imply that effects of the prenatal nutritional environment persist throughout the life course to affect late-life cognitive function. Moreover, the results imply that prenatal exposures can magnify genetic effects on late-life cognitive function. Dementia research would benefit from expanding research efforts to include more investigation of periods across the life course.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1734 Prevalence of overweight and obesity in individuals with Urea Cycle Disorders: results from a multicenter study of the Urea Cycle Disorders Consortium

Authors:

N. Ho[1,2], P. Lupo[1], K. Simpson[3], B. Rawls-Castillo[2], J. Barber[3], S. Ali[1,2], S. Elsbecker[4], N. Ah Mew[3], Members of the Urea Cycle Disorders Consortium, S. Nagamani[1,2], L. Burrage[1,2]; [1]Baylor Coll. of Med., Houston, TX, [2]Texas Children's Hosp., Houston, TX, [3]Children's Natl. Hosp., Washington, DC, [4]Univ. of Minnesota Masonic Children's Hosp., Minneapolis, MN

Urea cycle disorders (UCDs) are inborn errors of metabolism where dietary therapy comprises restricting protein in diet, supplying adequate calories, and avoiding fasting. Currently, it is not known whether such dietary therapy increases the risk for overweight and obesity in UCDs. Using data from "Longitudinal Study of Urea Cycle Disorders" (LS-UCD), an ongoing, multicenter, natural history study of UCDs conducted by the Urea Cycle Disorder Consortium, we calculated the overall prevalence of overweight and obesity in individuals with UCDs and compared the prevalence with that of the U.S. general population (NHANES dataset). For LS-UCD participants, data from individuals ages ≥2 years were included. Participants with incomplete data and those enrolled after liver transplantation were excluded. Height, weight, and clinical covariates, i.e., age, type of UCD, symptomatic status, history of hyperammonemia, and use of nitrogen-scavenging medications were analyzed from the participant's most recent study visit. Each participant's BMI (kg/m^2) was calculated and categorized using the CDC definitions for underweight, normal weight, overweight, or obese. Z-tests of two proportions were used to compare the proportion of overweight/obese in UCDs vs the U.S. general population; and logistic regression models were used to find associations between clinical factors associated with overweight/obesity in individuals with UCDs. The study analyzed data from 644 individuals (321 children; 323 adults). We found that overweight/obesity were as prevalent in the pediatric population of the LS-UCD (32%), as in the general U.S. pediatric population (35%, p=0.31). However, the prevalence of obesity was lower in adults in LS-UCD (31%), as compared to the U.S. general population (42%, p<0.001). In children, symptomatic status was associated with an odds ratio of 2.1 (CI 0.7-5.9) for obesity though this result was not statistically significant (p=0.06). In adults, age was associated positively with obesity/overweight. Odds ratios for obesity were 4.4 (CI 2.2-8.7) for 40-59 years, and 3.4 (CI 1.1-10.6) for >60 years, compared to the 20-39-year-old reference group. Given these findings, further assessment of longitudinal data is needed to follow the pediatric cohort over time and determine whether the prevalence of obesity remains similar to the general population as these children reach adulthood or whether there is a lower prevalence as they reach adulthood, similar to the adult cohort in our study. Routine nutrition assessments and long-term follow-ups may give insight into the lifelong effects of dietary therapy in UCDs.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1735 Prioritization of causal genes expressed in liver associated with nonalcoholic fatty liver disease using a Mendelian randomization approach

Authors:

E. Gobeil; CRIUCPQ-ULAVAL, Quebec, QC, Canada

Introduction: There is currently no specific preventive or curative pharmacological treatment for nonalcoholic fatty liver disease (NAFLD), a chronic disease that affects one in four adults worldwide. The genetic architecture of NAFLD needs to be precisely characterized to prioritize potential therapeutic targets and develop effective therapies. **Objective**: To identify liver-expressed genes significantly associated with NAFLD by combining genome-wide association study (GWAS) summary statistics, Mendelian randomization (MR), and genetic colocalization. **Methods**: We performed genetic colocalization and MR analyses on 75 independent NAFLD genetic loci identified by the Million Veteran Program (90,408 cases and 128,817 controls) using GWAS summary statistics from a meta-analysis of 16,532 NAFLD cases and 1,240,188 controls. Liver expression of these genes was assessed using bulk RNA sequencing in 246 participants undergoing bariatric surgery. **Results**: Genetic colocalization was confirmed at 7 loci (*AKNA, CWF19L1, C2orf16 (GCKR), EPHA2, HLA-G, SLC2A2/GLUT2 and TRIB1*). MR analyses revealed that hepatic expression of the *TRIB1, GCKR,* and *ZNF56 (TM6SF2)* genes was positively associated with NAFLD. **Conclusion**: Genetic colocalization analyses enabled us to identify which of the genes significantly associated with NAFLD are expressed in the liver, sharing a common variant. These genes could represent therapeutic targets of interest for the development of liver-targeted therapies to prevent or treat NAFLD.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1736 Profiling adaptive immune receptor repertoire (AIRR) in Graves' disease patients with antithyroid drug (ATD)-induced agranulocytosis.

Authors:

Y. Chen[1], S-K. Lai[2], I-H. Chiu[3], P-L. Chen[1,2,4,5,6]; [1]Graduate Inst. of Med. Genomics and Proteomics, Coll. of Med., Natl. Taiwan Univ., Taipei, Taiwan, [2]Genome and Systems Biology Degree Program, Academia Sinica and Natl. Taiwan Univ., Taipei, Taiwan, [3]epartment of Med. Res., Natl. Taiwan Univ. Hosp., Taipei, Taiwan, [4]Dept. of Med. Genetics, Natl. Taiwan Univ. Hosp., Taipei, Taiwan, [5]Div. of Endocrinology and Metabolism, Dept. of Internal Med., Natl. Taiwan Univ. Hosp., Taipei, Taiwan, [6]Graduate Inst. of Clinical Med., Coll. of Med., Natl. Taiwan Univ., Taipei, Taiwan

Graves' Disease (GD) is the primary cause of hyperthyroidism, with a prevalence ranging from 1.0% to 1.6% in the general population. Antithyroid drug-induced agranulocytosis (TiA) is a concerning adverse effect that occurs in 0.1% to 0.37% of GD patients during their treatment. Previous studies have suggested an association between TiA susceptibility and genetic variations, particularly in the human leukocyte antigen (HLA) genes. While genetic studies on GD and TiA have provided valuable insights, many aspects of the underlying pathophysiological and immunological mechanisms remain elusive. We hypothesize that the adaptive immune receptor repertoire (AIRR), including B lymphocyte receptors (BCRs, also known as immunoglobulins or antibodies) and/or T lymphocyte receptors (TCRs, also known as TRs), may contribute significantly to the missing genetic factors in TiA. To investigate the pathophysiological and immunological mechanisms of AIRR and TiA, we have recruited more than 64 TiA patients and collected serial samples from several individuals. We employed two single-cell RNA sequencing (scRNA-seq) platforms, BD Rhapsody and 10X Genomics, to assess the gene usage repertoire of V(D)J in both TCR and BCR. Subsequently, we identified the amino acid sequences of the complementarity-determining region 3 (CDR3) in major responsive clones from TiA patients at the acute and recovery stages (with an interval of at least six months between the two stages). Herein, we present our preliminary findings of six samples from two TiA patients. We observed distinct profiles of BCR clonotypes between the acute and recovery phases within the same patient, indicating a potential role of B cells in the pathogenesis of TiA. In contrast, no significant differences were observed in TCR clonotype profiles. Furthermore, there are no differences in major clonotype of TCR or BCR between TiA patients either in acute or recovery phases. The consistency of the V(D)J gene usage repertoire results across both platforms was confirmed by the samples of one patient. Based on these findings, we propose a digenic hypothesis regarding the risk of GD and/or TiA. The interaction between AIRR and antigens presented by HLA molecules is crucial for immune recognition and response. Certain HLA types, such as HLA-B38:02 and HLA-DRB1*08:03 genetic variations, have been strongly associated with TiA susceptibility. Therefore, we plan to further expand the cohort and analyze the correlation between AIRR and TiA by collecting data from patients with the same HLA type.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1737 Profiling the inflammatory bowel disease subtypes using genetics, serum biomarkers, and smoking information

Authors:

R. Liu[1,2], D. Li[3], T. Haritunians[3], Y. Ruan[1], M. Daly[2,1], H. Huang[2,1], D. McGovern[3]; [1]Broad Inst. of MIT and Harvard, Cambridge, MA, [2]Analytic and Translational Genetics Unit, Massachusetts Gen. Hosp., Boston, MA, [3]Widjaja Inflammatory Bowel Res. Inst., Cedars-Sinai Med. Ctr., Los Angeles, CA

Crohn's disease (CD) and Ulcerative colitis (UC) are two etiologically related yet distinctive subtypes of inflammatory bowel diseases (IBD). Due to the heterogeneous presentation of IBD, differentiating CD from UC can be challenging in up to 20% of patients using conventional clinical approaches in a subset of patients. Diagnostic uncertainty and subsequent misclassification in IBD are associated with higher rates of complicated disease and relapse. We employed a novel molecular-based IBD subtypes prediction model aggregating genetics, serum biomarkers, and smoking information for the first time to investigate the contribution of these factors to IBD subtypes.

We performed the genetic association analysis on the International IBD Genetics Consortium (IIBDGC) Immunochip samples (European ancestry) using phenotypes coded as CD (15,987) v.s. UC (12,613). We identified 36 loci associated with IBD subtypes, 25 of which overlap with known IBD genetic risk loci. NOD2 and MHC made the largest contribution to the variation of IBD subtypes, which explained 7.5% and 3% of variance, respectively. The polygenic risk score (PRS) model training using IIBDGC cohort and tested in an independent CEDARS cohort explained 19% of the variation in IBD subtypes.

We assessed the contribution of serum biomarkers (ASCA, ANCA, CBir1, OmpC, and I2) and smoking status in the variance of subtypes of IBD. The serology-only model and smoking-only model explained 40% and 1% of variance, respectively. Among serum biomarkers, CBir1 makes the largest contribution explaining 14% of variance. Conditional analysis revealed that the genetic, serology and smoking information each makes independent contributions to distinguish IBD subtypes despite the fact that their contributions are modestly overlapping. The combined model, with genetics, serology and smoking information, explained 46% of variance. We also found colorectal CD (L2) is significantly more challenging to be distinguished from UC than small bowel CD (L1) (17% and 46% of variance, respectively), indicating the different molecular basis underlying CD patients with different disease locations.

As the largest study incorporating genetics data with serum and environmental stimulations (smoking), our study revealed their complementing contributions to IBD disease subtypes and locations with statistical power, suggesting a model combining the molecular and environmental information may complement current diagnostic strategies and help classify patients based on biologic state.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1738 Protective Rare Variants for Cognitive Preservation in the Mid-Western Amish

Authors:

Y. Song[1], W. Wang[1], K. Miskimen[1], R. Laux[1], S. Fuzzell[1], S. Hochstetler[1], L. Main[1], M. Osterman[1], P. Wang[1], Y. Liu[1], M. Prough[2], D. Dorfsman[2], A. Fernandez[2], S. Slifer[2], L. Adams[2], L. Caywood[2], J. Clouse[2], S. Herington[2], A. Lynn[1], J. Vance[2], M. Cuccaro[2], P. Ogrocki[1], A. Lerner[1], M. Pericak-Vance[2], W. Scott[2], J. Haines[1]; [1]Case Western Reserve Univ., Cleveland, OH, [2]Univ. of Miami, Miami, FL

A shift in focus from risk to resilience for Alzheimer's disease (AD) encourages efforts to uncover novel AD biological mechanisms. Rare variants identified through whole genome sequencing (WGS) may represent an important and understudied component of complex trait genetics. While population-based studies are powered to discover associations with common genetic variants, founder populations are better powered for discovery of previously unknown rare alleles that have risen to higher frequency due to genetic drift. We examined WGS data from a Midwestern Amish population in a genome-wide search for rare coding variants shared only among cognitively unimpaired (CU) individuals, determined by consensus review of medical history and neuropsychological testing. Rare variants (allele frequency <0.05 across all samples) with a minor allele count (MAC)>10 in the CU group, and 0 in the cognitive impairment (CI) group were annotated to determine likely loss of function. Variants shared by CU individuals at least 80 years old at the most recent exam were further investigated for family relationship and allele frequency comparison with the general population. After extensive QC, 1,048 individuals were used to identify rare variants (n=11,854,817). Of these individuals, 634 were CU (mean age=81.53±6.10, 60% female) and 184 were CI (mean age=84.77±6.60, 62% female). 51,616 variants with MAC>10 were found only in CU, including 316 unique coding variants (288 *missense*, 4 *inframe_insertion*, 6 *inframe_deletion*, 11 *frameshift*, 4 *stop_gained*, 3 *splice_donor*) and 180 synonymous variants. Among them, 7 missense and 9 synonymous variants had MAC>20 and the rest had MAC of 11-20. Out of 7 missense and the rest 28 coding variants, 3 variants were shared by age ≥ 80 CU; one *frameshift* variant in MEGF6 and another in GPN2 on chromosome 1 and one *inframe_deletion* variant in NOP2 on chromosome 12. The allele frequencies of the 3 variants were (3.0×10^{-5}, 1.5×10^{-5}, 0.0124) in TOPMed compared to (0.0057, 0.0086, 0.0057) in our data, respectively. Each variant was shared within at least one CU sibship of size 2 to 8 and all variant sharing CU individuals were related within one big family including up to the 4th generation ancestors. Additionally, several untranslated region or nonsense-mediated decay transcript variants existed within 26 known AD gene regions, among which 8 of them are reported to harbor protective variants for AD. We identified numerous rare coding variants potentially impacting cognitive preservation, providing a rich resource for further investigation of genes that may influence cognitive function in older individuals.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1739 Proteome-derived unsupervised learning identifies biologically distinctive subtypes of Parkinson's disease.

Authors:

A. Doostparast Torshizi, S. Li; Janssen, Spring House, PA

Background: Parkinson's disease (PD) shows a high degree of heterogeneity. To develop safe and effective therapies, characterizing the right set of patients is of paramount importance. In this study, we have used high-throughput proteomics data from idiopathic PD patients as well as detailed cross-sectional and longitudinal clinical and pathological measurements to subtype patients using network modeling and unsupervised machine learning (ML).
Methods: We used the data generated by the Parkinson's Progression Markers Initiative (PPMI) with a total of 504 idiopathic patients. The CSF protein data contains 5,193 analytes corresponding to 4,198 unique proteins. Network modeling was applied to identify the set of features for unsupervised ML-based subtyping. Consensus-clustering based on two approaches, agglomerative hierarchical clustering and k-means, was applied. The identified subtypes were then followed up based on the following datasets: whole-genome sequencing data (WGS), Unified Parkinson's Disease Rating Scale (UPDRS), Montreal Cognitive Assessment Scale (MOCA) scores, and pathological measurements including alpha-synuclein (a-syn), total tau (tTau), and phospho-tau (pTau).
Results: Network modeling led to 1,387 features. Two subtypes S1 (n=215) and S2 (n=190) were identified. Baseline and 10-year longitudinal analysis measurements of a-syn, tTau, and pTau revealed significant increase in S1. Longitudinal analysis of UPDRS-I scores to assess intellectual impairment identified late-stage divergence between the two subtypes. Longitudinal evaluation of motor symptoms identified significant mid-stage increase in S1 which was complemented with cognition scores. Differential expression (DE) analysis between each subtype versus healthy controls revealed only 5 DE proteins in S1 while 603 proteins were DE in S2. Enrichment analysis of known PD-associated genetic variants in the identified subtypes revealed distinct genetic profiles between 2 subtypes.
Conclusion: Precision medicine-based therapeutic target discovery requires data-driven subtyping of patients. Using ML along with a wide array of molecular, phenotypic, and pathological measurements we identified two subtypes of idiopathic PD patients which paves the way to inform a more impactful target discovery.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1740 Proteomic Changes in the Development of Type 2 Diabetes

Authors:

H. Highland[1], P. Sharma[2], M. Graff[3], E. Frankel[4], W. Zhu[4], H-H. Chen[5], L. E. Petty[4], R. Roshani[4], M. Lee[6], E. R. Gamazon[7], J. B. McCormick[8], S. Fisher-Hoch[6], J. E. Below[9], K. E. North[10]; [1]Univ North Carolina at Chapel Hill, Chapel Hill, NC, [2]Univ. of North Carolina, Chapel Hill, NC, [3]Univ North Carolina, Chapel Hill, NC, [4]Vanderbilt Univ., Nashville, TN, [5]Academia Sinica, Taipei, Taiwan, [6]Univ. of Texas Sch. of Publ. Hlth., Brownsville, TX, [7]VUMC Clare Hall, Univ. of Cambridge, Nashville, TN, [8]UTH SPH Brownsville, Brownsville, TX, [9]Vanderbilt Univ Med Ctr., Nashville, TN, [10]Univ North Carolina, Wake Forest, NC

Background: The population at the Texas-Mexico border faces a rapidly increasing prevalence of type 2 diabetes (T2D) at rates higher than the general US population. This population has limited access to health care and medical treatments. While T2D is known to be a complex multifactorial disease, only part of the biological changes underlying disease progression are understood. The Cameron County Hispanic Cohort (CCHC) has followed Mexican American residents of South Texas over the last 20 years with stored biological samples.
Methods: Proteomics were measured from the plasma of 261 people at 2 timepoints on the O-link platform. Using 2010 ADA criteria, individuals were classified as having diabetes or not. Approximately 100 people developed T2D between visits. Using a linear mixed-effect model including T2D, age, sex, BMI, and PEER factors, we tested for changes in plasma protein levels with the development of T2D.
Results: Among 2921 proteins considered, 48 proteins displayed a significant change with incidence of T2D, using a stringent Bonferroni correction of ($p < 1.7x10\text{-}5$). Significant proteins included those involved in known T2D pathologies and glucose homeostasis, e.g. people with T2D have significantly lower levels of LPL, PTPRS, ISM1, PPY, and CCDC80. Adipogenesis and lipid metabolism proteins were also represented including APOA4 and LPL. Proteins involved in vascular changes, angiogenesis, and circulatory disorders were also altered, including increased AOC3, DPP6, and ADA2 and decreased TGFBR3, ISM1, TYMP, and CD248. Many associated proteins were related to extracellular structure and bone maintenance including MXRA8, PRCP, COL15A1, and DMP1.
Conclusions: Changes in the human proteome that occur in conjunction with the development of type 2 diabetes may show both mechanisms through which glycemic dysregulation occur and changes caused by this regulation. Vascular, microvascular, and endothelial changes are markers of many diabetes complications. Improved understanding of the proteomic changes highlights potential points of therapeutic intervention to decrease T2D complications.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1741 Pseudotime analysis of post-mortem brain RNA profiles identify FMOD as a key gene in Alzheimer's disease

Authors:

I. Piras, M. Naymik, M. Huentelman; Translational Genomics Res. Inst., Phoenix, AZ

Pseudotime (PT) methods are machine learning-based algorithms that enable extracting latent temporal information from cross-sectional studies. We applied PT analysis to Alzheimer's disease (AD) and control samples at various clinical and neuropathological stages. The analysis was conducted using post-mortem RNA-sequencing bulk data from four brain regions: Brodmann areas (BA) 10, 22, 36, and 44 (Mount Sinai dataset; #syn7391833). Using the phenoPath method, we generated PT trajectories, modeling the brain region as a covariate.PT trajectories significantly correlated with clinical and neuropathological variables (adj-p range: 2.5^{-13} to 0.040), except with Clinical Dementia Rating and Braak stage in BM10 and BM44, respectively (adj-p $\geq$ 0.461). Through this analysis, we identified 2,364 PT-associated genes (PTGs) across the four brain regions ($| r | \geq 0.6$ and BH < 0.05). Notably, differential expression analysis (#syn2580853) did not detect 60.9% of these PTGs. We generated multi-brain co-expression networks using the Multiscale Embedded Gene Expression Network analysis (MEGENA) method, revealing 752 modules significantly correlated with PT (adj-p < 0.05). Subsequent key-driver analysis of these associated modules, conducted leveraging a brain-specific Bayesian regulatory network, unveiled identified 1,323 significant key drivers linked to 298 unique genes (adj-p < 0.05). Interestingly, *FMOD* (Fibromoduline), emerged as the top key-driver in all the four brain regions (adj-p range: 3.4^{-86} to 3.5^{-45}). In all four brain regions, *FMOD* was consistently included in modules positively correlated with PT, enriched for extracellular matrix and collagen Gene Ontology functional classes, and pericyte genes. *FMOD* also showed a significantly positive correlation with PT in BA10, BA22 and BA36.This study primarily highlights the significance of *FMOD*, the gene coding for Fibromodulin. This small interstitial proteoglycan plays an important role in the organization of collagen. Across of four regions analyzed, *FMOD* emerged consistently as the top key driver, potentially involved in collagen and extracellular matrix alterations observed in AD (PMC8430252; PMC4931855; PMC6031366). Furthermore, our application of PT algorithm confirmed the capability of PT algorithm to order the samples according to disease severity purely based on molecular profiling, without providing previous knowledge. This approach also facilitated the identification of genes that were not detectable through differential expression analysis, thereby uncovering previously hidden associations between molecular and phenotypic characteristics.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1742 PyGAD machine learning algorithm improve COVID-19 prediction accuracy

Authors:

C. Fallerini[1,2], G. Martelloni[1,2], **K. Zguro**[1,2], G. Brunelli[1,2], S. Minetto[1,2], M. Lista[1,2], D. Maffeo[1,2], E. Pasquinelli[1,2], G. Rollo[1,2], S. Basso[1,2], E. Antolini[1,2], M. Rozza[1,2], A. Rina[1,2], S. Croci[1,2], A. Degl'Innocenti[1,2], M. Baldassarri[1,2], M. Bruttini[1,2], I. Meloni[1,2], GEN-COVID Multicenter Study, S. Furini[3], A. Renieri[1,2,4], S. Olmi[5]; [1]Med Biotech Hub and Competence Ctr., Dept. of Med. Biotechnologies, Univ. of Siena, Siena, Italy, [2]Med. Genetics, Univ. of Siena, Siena, Italy, [3]Dept. of electrical, electronic, and information engineering "Guglielmo Marconi", Univ. of Bologna, Bologna, Italy, [4]Genetica Medica, Azienda Ospedaliero-Univ.ria Senese, Siena, Italy, [5]CNR-Consiglio Nazionale delle Ricerche - Istituto dei Sistemi Complessi, Sesto Fiorentino, Italy

COVID-19 pandemia represented a worldwide challenge in the past 3 years. Even if the emergency is ending, prediction on susceptibility to severe disease is still a major medical need for better treatment and health policy actions. We have developed a novel predictive model called post-Mendelian because it is putting together the impact of rare variants, known to have impact in Mendelian disorders, with common polymorphisms in coding regions, likely impacting on the function, building on IPGS (Integrated PolyGenic Score). The biological value of the model was then validated by carrying functional analysis of a series of specific genes (features). Using LASSO logistic regression we were able to obtain a performance at 70-75%. We now present an improvement of the model using PyGAD machine learning algorithm based on the previously extracted features. We use the genetic algorithm PyGAD to fit the coefficients of the frequency of the variants, in the formula of the IPGS, to find the best correlation with the severity of the disease for every patient. Compared to the previous model, we maintain the linearity and the division into 4 pieces, each of them describing variants with different frequencies. The new model is reaching performances at 80-88%. The main change is related to avoiding fixed rules on the relationship between ultra rare, rare, low frequency and common variants and to fit the model: on a first analysis we observe that the best solution of the fit lives on a minimum of the parameters space which is degenerate, this means that, a priori, it is impossible to establish if a class of variants is more important that another one. The increased flexibility of the model may better represent the complex biological scenario and heterogeneity among patients, suggesting that we look for patterns that involve more variants with different frequencies. Further analysis will be necessary in order to validate the model in different populations with the final goal of translation in clinical practice for helping patients and public health decision makers.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1743 Quantifying individual-level mosaic chromosomal alteration fitness: genetic determinants and clinical consequences

Authors:

Y. Pershad[1], T. Mack[1], H. Poisner[1], Y. Jakubek[2], NHLBI TOPMed Consortium, R. Mathias[3], A. Reiner[4], P. Scheet[5], P. Auer[6], S. Jaiswal[7], J. Weinstock[8], A. Bick[1]; [1]Vanderbilt Univ., Nashville, TN, [2]Univ. of Kentucky, Lexington, KY, [3]Johns Hopkins Univ, Baltimore, MD, [4]Univ of Washington, Seattle, WA, [5]MD Anderson Cancer Ctr., Houston, TX, [6]Med. Coll. of Wisconsin, Milwaukee, WI, [7]Stanford Univ., Stanford, CA, [8]Johns Hopkins Univ., Baltimore, MD

Clonal hematopoiesis (CH) is characterized by a clonally expanded population of hematopoietic stem cells. CH can be caused by single nucleotide mutations in myeloid cancer driver genes, termed CHIP or larger structural rearrangements called mosaic chromosomal alterations (mCAs). Factors modulating variations in mCA clonal fitness are poorly understood.

To address this, we extended a recently developed method called passenger-approximated clonal expansion rate (PACER) to quantify mCA fitness from a single blood sample. In this study, we apply PACER to 6,381 individuals from the NHLBI TOPMed cohort with mCAs involving gain, loss, and copy-neutral loss of heterozygosity.

To validate PACER, we compared our fitness estimates to an alternative method that infers mCA fitness using variant allele frequency probability distributions from the UK Biobank. PACER estimates, aggregated by mCA location and type, exhibit a high correlation ($R2 = 0.49$) with this orthogonal approach.

Unlike prior methods, PACER enables quantifying mCA clone expansion rate at the individual level, facilitating the identification of germline risk loci associated with clonal expansion and quantifying phenotypic consequences of mCA expansion rate.

We leveraged PACER to identify germline mutations that alter mCA fitness. We conducted a genome-wide association study of mCA PACER and identified a single locus (TCL1A) that reached genome-wide significance. Our PACER analyses also nominate NRIP1, a loci previously associated with mCA prevalence at genome-wide significance, as acting through modulating clonal fitness.

We sought to understand whether individuals with mCAs expanding at a faster rate had adverse clinical outcomes. Within the subset of mCAs that are known to cause lymphoid malignancies, we identified that increased mCA expansion rates were associated with higher lymphocyte counts ($p = 0.019$). Future work will evaluate whether individuals with increased mCA expansion rate also are at highest risk of progression to overt hematologic malignancy.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1744 Quantifying the extent of pleiotropy using rare variant association data in 394,841 human exomes

Authors:

W. Lu[1], B. Neale[2], L. O'Connor[1], K. Karczewski[3]; [1]Broad Inst., Cambridge, MA, [2]Massachusetts Gen. Hosp., Boston, MA, [3]Broad Inst., Medford, MA

Pleiotropy, the property of a single genetic component acting on unrelated phenotypes, is pervasive in the human genome. Although previous studies have discovered widespread pleiotropy for common variants, a systematic analysis of the pleiotropic effect of rare coding variants is yet to be conducted. Utilizing Genebass, a large-scale rare variant association study on UKB WES data, we evaluate the impact of rare coding variants across diverse phenotypic domains, which quantifies the extent of exome-wide pleiotropy and provides additional context to GWAS loci. The data encompass over 8M single-variant tests and 75K gene-based burden tests across 4,529 phenotypes from 394,841 WES samples. Across 239 curated uncorrelated endpoint phenotypes, 313 (0.5%) genes are associated with multiple phenotypes, which occurs at a higher rate than expected by chance (permutation p < 1e-100; mean of permuted genes = 0.07%). Among 37 gene categories previously known to have functional relevance and/or a role in disease, we find a depletion of pleiotropic genes among groups with vital biological functions and an enrichment of pleiotropic genes within disease-related groups. We then identify hundreds of genes showing pleiotropy across 6 independent phenotypic domains. Particularly, pLoF variants in two developmental disorder related genes GIGYF1 and KDM5B are associated with the most phenotypic domains (4). For instance, pLoF variants in KDM5B are associated with phenotypes including atrial fibrillation and flutter, several blood biomarkers and physical measures, and fluid intelligence score, suggesting potential insights into the mechanisms of developmental disorders.

Pleiotropy can manifest as an allelic series, where variants with different functional impacts may exhibit heterogeneous phenotypic effects. We observe this phenomenon in the gene-based results of 31 genes. For instance, pLoF variants in ATM are only associated with 7 cancer diagnoses, such as breast and pancreas cancer, while its missense variants are only associated with 3 blood biomarkers. We develop a statistical test to explore allelic heterogeneity within genes. For a pleiotropic gene, we test whether its association with two traits stems from the same set of variants with highly correlated effects, or different variants. For example, we identify evidence for a heterogeneous effect of a subset of missense variants in the ALB gene (encoding albumin) on albumin and calcium levels, suggesting that their effects on calcium are not mediated by albumin level itself. In this way we demonstrate shared biological pathways among disease groups, providing insight into the genetic basis of human disease.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1745 Quercetin ameliorates the autoimmune, lipodystrophic and neurodegenerative phenotype observed in CLEC16A KO mice.

Authors:

R. Pandey[1], B. Strenkowski[2], C. N. Thaise[1], M. Bakay[2], H. Hakonarson[1]; [1]Children's Hosp. of Philadelphia, Philadelphia, PA, [2]Children's Hosp. Of Philadelphia, Philadelphia, PA

CLEC16A is implicated in multiple autoimmune diseases. We generated an inducible whole-body knockout (KO) *CLEC16A* mice to address the role of CLEC16A loss of function. Published data from our KO mice suggests a derailed functional link between CLEC16A and impaired mitophagy and autophagy, two critical cellular processes involved in maintaining cellular homeostasis and preventing the accumulation of damaged organelles and proteins. While Quercetin, a natural flavonoid with antioxidant and anti-inflammatory properties, has been shown to modulate various cellular pathways, the effects of quercetin on mitophagy and autophagy in the context of CLEC16A deficiency remains unknown. In this study, we utilized *CLEC16A* KO mice as a model to investigate the impact of quercetin on mitophagy and autophagy to determine potential rescue of the observed defect and prevent the onset of the autoimmune, inflammatory, and neurodegenerative phenotypes. **Methods:** *CLEC16A* knockdown in adult mice leads to dysregulated mitophagy promoting robust autoimmune inflammatory response with severe neurological symptoms. We first confirmed the deficiency of CLEC16A in our mouse model using molecular and immunoblot analysis. Subsequently, we administered quercetin to *CLEC16A* KO mice and evaluated its effects on mitophagy and autophagy markers and phenotype rescue. **Results:** Our results demonstrate that CLEC16A deficiency resulted in impaired mitophagy and autophagy in various tissues. Importantly, treatment with quercetin significantly enhanced mitophagy and autophagy in these tissues. Quercetin administration resulted in upregulation of key mitophagy markers, including PINK1 and Parkin, as well as autophagy markers, including LC3-II with downregulation of P62. Overall Quercetin improved the survival of KO mice by modulating ER stress possibly by facilitating an increase in the number of autophagosomes and autolysosomes in quercetin treated *CLEC16A* KO mice compared to untreated controls. **Conclusion:** Overall, our findings suggest that quercetin supplementation promotes mitophagy and autophagy in *CLEC16A* KO mice, providing new insights into potential therapeutic strategies for autoimmune diseases associated with CLEC16A dysfunction. Further investigations are warranted to elucidate the precise molecular mechanisms underlying the beneficial effects of quercetin on these cellular processes and to evaluate its therapeutic potential in patients with multiple autoimmune diseases attributed to variants in CLEC16A by reversing the dysregulated mitophagy/autophagy processes introduced by these variants.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1746 Rapid Diagnosis and Longitudinal Monitoring of Bacteremia in Pediatric Septic Arthritis and Acute Hematogenous Osteomyelitis Cases Using Longread Sequencing

Authors:

P. Raj[1], T. Relangi[1], N. Tareen[2], R. Sridharan[1], B. Zhang[1], C. Arana[1], N. V. Oers[1], L. Copley[3], L. Hooper[1]; [1]UT Southwestern Med Ctr, Dallas, TX, [2]Children's Med. Ctr., Dallas, TX, [3]Children's Hosp., Dallas, TX

Existing gold-standard method for clinical diagnosis of bacterial infection is culture. But large blood specimen requirement and longer turnaround time often makes this method unsuitable for young children. We optimized a rapid metagenomic sequencing workflow that requires 1ml specimen (blood and joint fluid) to rapidly characterize pathogens using longread sequencing technology. We analyzed 50 specimens from healthy Children and 25 specimens from children diagnosed with Septic Arthritis, AHO and 22q11.2 deletion syndrome. Full length 16S rRNA gene sequencing and whole genome analysis identified presence of several pathogenic and commensal species including *Staphylococcus aureus, Streptococcus pyogenes, Haemophilus parainfluenzae, Veillonella parvula, Streptococcus oralis* and *Lactobacillus animalis*. Healthy blood showed the presence of *Streptococcus* DNA signature that was associated with elevated level of several pro-inflammatory cytokines in serum, such as INFg, IL1a, IL1b, IL17a and MIF. *Streptococcus pyogenes* species in Septic arthritis cases was strongly associated with inflammatory immune responses i.e. IL6, IL8. Our pipeline took 8 h from DNA extraction to taxonomic classification using EPI2ME pipeline. Sequencing results were 100% concordant with culture results in case of synovial fluid specimens. Blood and saliva specimens showed DNA signatures of commensal organisms which were validated using deep Illumina sequencing. Reference based genome assembly analysis was carried out in subset and variants were detected. Data analysis pipeline characterized pathogens at species level taxonomy. Host cytokine analysis showed activation of proinflammatory cytokines consistent with active bacteremia. We developed a metagenomics workflow that can rapidly characterize bacterial pathogens in blood and synovial fluid to guide clinical diagnosis and antibiotic therapy in children.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1747 Rare coding variants in *PREB* associate with sex-specific effects on LDL cholesterol.

Authors:

R. Hoffing[1], A. Dubey[1], A. Deaton[1], A. Xiao[1], K. Buckowing[1], L. Ward[1], A. Holleman[1], L. Krohn[2], P. LoGerfo[1], M. Plekan[1], S. Akle Serrano[1], P. Nioi[1]; [1]Alnylam Pharmaceuticals, Cambridge, MA, [2]Alnylam Pharmaceuticals, Montreal, QC, Canada

Sex-specific analyses can uncover new genome-phenome associations for complex traits that would otherwise be hidden. An exome-wide association study of LDL cholesterol was conducted separately for 187,061 men and 219,861 women in the UK Biobank. Aggregated rare predicted loss of function (pLOF) and damaging missense variants (REVEL $\geq$ 0.77) in *PREB* were found to associate with decreased LDL cholesterol in men ($p = 7.3x10^{-7}$, effect = 0.47 SD decrease), but not women (p = 0.46, effect = 0.06 SD increase). Silencing of *PREB* in HepB3 cells resulted in differential expression of known LDL-regulators, including *GPAM* ($p = 2.9x10^{-30}$, log2FC = -1.31), *HMGCR* ($p = 2.7x10^{-23}$, log2FC = -1.16), and *ACAT2* ($p = 8.8x10^{-21}$, log2FC = -1.12). A gene set enrichment analysis revealed significant downregulation of the cholesterol biosynthesis pathway ($p = 1.4x10^{-8}$, enrichment score = -0.86) upon *PREB* silencing. Overall, *PREB* shows genetic and functional evidence for a role in regulating LDL and could be a potential therapeutic target for hypercholesterolemia.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1748 Rare de novo deleterious variants increase risk of neural tube defects.

Authors:

Y. Lei[1], P. Lupo[1], R. Finnell[2]; [1]Baylor Coll. of Med., Houston, TX, [2]Baylor Coll. Med., Houston, TX

It is well established that Neural Tube Defects (NTDs) have complex environmental and genetic etiologies. However, the contribution of *de novo* mutations to the etiology of NTDs has not been well studied. We performed whole exome sequencing in 125 NTD trios which were mainly collected from Dell Children's Medical Center at Austin and Texas Children's Hospital at Houston. We identified 170 *de novo* variants which included 40 synonymous variants, 76 non-deleterious variants, 32 deleterious missense variants and 22 loss of function (LoF) variants. Eleven genes implicated in NTD mouse models including *EP300*, *PAX3*, *SHROOM3*, *SMURF2* and *TBXT* were detected in *de novo* deleterious variants class. The DenovolyzerR package was used for the statistical analysis. Compared with expected numbers calculated based on 1,078 autism spectrum disorder trios and 6,503 individuals from exome sequencing project, we observed a significant enrichment of *de novo* deleterious missense variants (1.42 folds, p=0.036) and *de novo* LoF variants (1.78 folds, p=0.008). Recurrent *de novo* deleterious variants were identified in *TBXT* gene, and a few gene families including *SATB* family (*SATB1* and *SATB2*) and *MED13* family (*MED13* and *MED13L*). Overall number of *de novo* variants in NTDs was also significantly high than expected (1.2 folds, p=0.012). There was no significant difference in number of *de nov*o synonymous variants and predicted to be tolerant missense variants compared to expected. Our study indicated that *de novo* deleterious variants contribute to risk of NTDs, while synonymous and tolerant missense variants did not increase NTD risks.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1749 Rare genetic variation in gonadotropin signaling identified in women with PCOS.

Authors:

G. Stokes[1], R. Bauer[1], C. Parker[1], R. Legro[2], M. Urbanek[3]; [1]Northwestern Univ., Chicago, IL, [2]Pennsylvania State Univ., Hershey, PA, [3]Northwestern Univ, Oak Park, IL

Background: Polycystic ovary syndrome (PCOS) is a complex endocrinopathy affecting 6-20% of reproductive age women. PCOS is defined by oligo/anovulation, hyperandrogenism, and abnormal ovarian morphology. Women with PCOS have altered frequency and amplitude of gonadotropin releasing hormone (GnRH) secretion, which causes an imbalance of gonadotropin hormones: luteinizing hormone (LH) and follicle stimulating hormone (FSH). An imbalance of LH:FSH ratio results in increased androgen production and oligo/anovulation—two cardinal characteristics of PCOS. Twin and family studies show PCOS to be highly heritable, and genome-wide association studies (GWAS) implicate several loci near genes involved in gonadotropin secretion. However, PCOS is a complex and polygenic disease, and current risk loci account for <10% of PCOS heritability with no known causal variants. Therefore, our goal is to determine whether genetic variation in gonadotropin genes contributes to PCOS pathogenesis by sequencing the gonadotropin genes in a PCOS cohort.
Objective: Our objective is to determine whether genetic variation in gonadotropin genes contributes to PCOS pathogenesis and to elucidate the molecular mechanisms of gonadotropin signaling dysfunction in PCOS.
Design/Methods: We performed targeted sequencing of gonadotropin pathway genes—*FSHR, LHCGR, FSHB* and *LHB*—in two cohorts of women with PCOS. We applied the American College of Medical Genetics (ACMG) criteria to each variant to determine a pathogenicity score. Variants were assigned a score based on population data, computational data, and functional data.
Results: We identified 14 rare missense variants (MAF < 0.01) in *FSHR* and *LHCGR* in a cohort of 673 women with PCOS and 123 reproductively healthy women. Additionally, we identified 8 rare missense variants in *FSHB* and *LHB* in a cohort of 602 women with PCOS and 165 reproductively healthy women. Based on ACMG criteria, there is strong existing evidence to classify 5 variants as likely pathogenic, 1 as pathogenic, and 5 as benign. 11 variants were identified as variants of unknown significance (VUS).
Future directions: PCOS GWAS loci have been found near gonadotropin genes, but no studies have identified pathogenic variants in gonadotropin genes or performed functional evaluations of gonadotropin gene variants. This study identifies and will functionally evaluate gonadotropin gene variants in women with PCOS. Identifying and understanding genetic contributions to complex endocrinopathies like PCOS can increase the ability to find causal variants and improve treatment efficacy.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1750 † Rare missense variants in protein intrinsically disordered regions: impact on condensates and common diseases

Authors:

R. Tian, C. Zhang, I. Klein; Dewpoint Therapeutics, Boston, MA

Biomolecular condensates are membrane-less compartments that concentrate a subset of biomolecules through phase separation from the surrounding cellular milieu and control numerous biological processes. Growing evidence supports that aberrant condensates serve as central nodes of dysregulation to drive the pathogenesis of various diseases, such as cancer, cardiomyopathy, neurodegeneration, and viral infections.
Intrinsically disordered regions (IDRs) within proteins contribute to condensate formation. Disease mutations located in IDRs perturb normal phase separation, leading to dysregulation of biological processes and pathogenesis. Here, we explored the role of genetic variation in IDRs in common diseases using whole-exome sequencing data in the UK Biobank. We examined 16 ICD10-defined diseases with >=1% case prevalence in individuals of European genetic ancestries, including autoimmune, cardiometabolic, eye, respiratory and neurological diseases.
We used 'metapredict' to predict 18,985 protein IDRs based on their amino acid sequence on the canonical transcript and annotated missense variants by degree of deleterious effects (likely, possibly and unlikely). To identify risk genes, we aggregated ultra-rare missense variants with each group of deleterious effects into a burden score and tested the association between the risk of a common disease and the gene-wide burden across the genome. We found that 1/3 of missense variants are located in IDRs across 80.9% of the protein-coding genes, which may play a role in some common diseases. We identified several IDR-wide significant (P<2.5e-6) disease-gene associations. Our initial findings include *ZNF366* in inflammatory bowel disease, *TGIF2* in non-alcoholic steatohepatitis risk and *CEP290* in heart failure risk. Both ZNF366 and TGIF2 are transcription factors (TFs) while CEP290 contains 13 coiled-coil domains. Given current knowledge on the role of TFs and coiled-coil domains in condensates, our findings suggest that *ZNF366*, *TGIF2* and *CEP290* may be involved in condensates that serve as central nodes of dysregulation in common diseases, and that mutations within these genes affect disease risk through disrupting the normal function of these condensates.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1751 Rare pathogenic missense variants in *LMNA* identified in women with polycystic ovary syndrome (PCOS)

Authors:

R. Bauer[1], L. Gorsic[1,2], R. Legro[3], M. G. Hayes[1], M. Urbanek[1]; [1]Northwestern Univ., Chicago, IL, [2]Mayo Clinic, Rochester, MN, [3]Penn State Hershey Med. Ctr, Hershey, PA

Background: Polycystic ovary syndrome (PCOS) is the most common form of anovulatory infertility among reproductive age women. In addition to experiencing reproductive symptoms, women with PCOS are at elevated risk of developing obesity, insulin resistance (IR), and type 2 diabetes. Familial partial lipodystrophy type 2 (FPLD2) is a disorder of lipid storage and insulin resistance caused by dominant missense alleles in the gene encoding the intermediate filaments lamin A/C (*LMNA*). Women with this Mendelian disorder of IR also experience symptoms of PCOS such as amenorrhea and hyperandrogenism. We therefore hypothesize that genetic variation in *LMNA* also contributes to PCOS. In other words, we hypothesize that PCOS falls into the phenotypic spectrum of disorders caused by variation in *LMNA.* **Objective:** We aim to identify and evaluate variation in *LMNA* that underlies PCOS pathogenesis in two cohorts of PCOS patients. **Methods:** We sequenced the *LMNA* gene in two cohorts, totaling 845 PCOS patients and 157 reproductively healthy controls. We comprehensively screened *LMNA* for genetic variation that is likely to alter the lamin A/C proteins, including missense, nonsense, splicing or frameshift variants. To determine which of these variants are pathogenic, we evaluated them according to criteria outlined by the American College of Medical Genetics (ACMG). **Results:** In the first cohort, we identified 7 missense variants in 8 cases out of 602, and no variants in 125 reproductively healthy controls ($X2=3.1$, $p=0.081$, $OR>1.78$ with study controls; $X2=46.8$, $p<1x10^{-8}$, $OR=8.5$ with gnomAD non-Finnish European population controls). 6 of these variants are pathogenic and 1 is likely pathogenic. When assessed individually, 6 of the 7 variants are significantly enriched in our cohort when compared to gnomAD non-Finnish European population controls ($OR > 5.0$). In the replication cohort, we identified 4 variants in 6 cases out of 243, and 1 out of 32 reproductively healthy controls ($X2=0.05$, $p=0.83$, $OR = 0.79$ with study controls; $X2=34.2$, $p<0.00001$, $OR= 7.73$ with population controls). Of the 4 variants in cases, 3 are pathogenic and 1 is likely pathogenic. The variant identified in a control is a variant of unknown significance. All *LMNA* variants in both cohorts are likely damaging as predicted by 3 computational methods. Additionally, 6 of the 11 variants in PCOS cases have previously been identified in individuals with lipodystrophy. **Conclusion:** Together with previous identification of *LMNA* variants in women with PCOS by our lab (Urbanek et al. 2009 *JCEM*) and others (Crespo et al. 2022 *JES*), this work further establishes pathogenic variation in *LMNA* as a disease mechanism for PCOS.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1752 Rare Recurrent Copy Number Variations in Children with Neurodevelopmental Disorders

Authors:

M. Khan[1], J. Glessner[1,2], X. Chang[1], Y. Liu[1], F. Otieno[1], M. Lemma[1], F. Mentch[1], C. Kao[1], M. March[1,2], J. Connolly[1], H. Hakonarson[1,2]; [1]Ctr. for Applied Genomics, Children's Hosp. of Philadelphia, Philadelphia, PA, [2]Perelman Sch. of Med., Univ. of Pennsylvania, Philadelphia, PA

Neurodevelopmental disorders (NDDs) such as attention deficit hyperactivity disorder (ADHD) and autism spectrum disorder (ASD) are illustrative examples of clinical cases with diagnosis impeded by current lack of definitive corroborating genetics and genomics. Both NDDs have been shown to share similar biological etiologies as well as genetic pleiotropy. ADHD and ASD have complex genetic associations implicated by rare recurrent copy number variations (CNVs). Platforms aimed at investigating genetic-based associations, such as high-density microarray technologies, have been successful in elucidating some of the underlying disease biology associated with NDDs and related psychiatric diseases. Previous studies investigating NDDs have uncovered CNVs associated with genes within shared candidate genomic networks, such as metabotropic glutamate receptor (mGluR) protein-protein interactions. To examine shared biological pathways across two of the most common NDDs, we investigated CNVs across 15,689 individuals with ADHD (n = 7,920), ASD (n = 4,318), or both (n = 3,416), as well as 19,993 controls. Cases and controls were matched by genotype array (i.e., Illumina array versions). Three case-control association studies each calculated and compared the observed vs. expected frequency of CNVs across individual genes, loci, pathways, and gene networks. Quality control measures of confidence in CNV-calling, prior to association analyses, included visual inspection of genotype and hybridization intensity. To extend our previous observations implicating a key role of the mGluR network in both ADHD and autism, we exhaustively queried patients with ASD and/or ADHD for CNVs associated with the 273 genomic regions of interest within the mGluR gene network (genes with one or two degrees protein-protein interaction with mGluR 1-8 genes). We uncovered CNTN4 deletions enriched in NDD cases (P = 3.22E-26, OR = 2.49). Additionally, we found PRLHR deletions in 40 ADHD cases and 12 controls (P = 5.26E-13, OR = 8.45). We also discovered clinically diagnostic relevant 22q11.2 duplications and 16p11.2 duplications in 23 ADHD + ASD cases and 9 controls (P = 4.08E-13, OR = 15.05) and 22q11.2 duplications in 34 ADHD + ASD cases and 51 controls (P = 9.21E-9, OR = 3.93); those control samples were not with previous 22qDS diagnosis in their EHR records. Taken together, we have uncovered multiple disease causing CNVs in both ASD and ADHD cases. Our latest work and direction includes genome-wide gene interaction network analysis in subjects with multiple associated psychiatric comorbidities.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1753 Rare variant analysis of 30,648 heart failure cases using exome sequences from multi-ancestry population-based biobanks and clinical trials

Authors:

A. Nag[1], G. Alamgir[1], Q. Wang[2], A. Walentinsson[3], A. Genomics Initiative[1], J. Berumen-Campos[4], P. Kuri-Morales[4], R. Tapia-Conyer[4], J. Alegre-Díaz[4], J. Emberson[5], J. Torres[5], R. Collins[5], D. Saleheen[6], Q-D. Wang[1], K. Hansson[1], S. Petrovski[1], D. Paul[1], K. Smith[1]; [1]AstraZeneca, Cambridge, United Kingdom, [2]AstraZeneca, Waltham, MD, [3]AstraZeneca, Gothenburg, Sweden, [4]Natl. Autonomous Univ. of Mexico, Mexico City, Mexico, [5]Univ. of Oxford, Oxford, United Kingdom, [6]Columbia Univ., USA & CNCD, Pakistan, New York, NY

Background: Heart failure (HF) affects ~1-2% of adult populations and has a large unmet therapeutic need. Identifying genes harbouring rare variants that associate with HF risk may provide novel therapeutic targets. Methods: We ascertained HF cases (~31% non-Europeans) from three global population-based biobanks (UK Biobank, Mexico City Prospective Study, and Pakistan Genome Resource) and six AstraZeneca HF clinical trials with exome sequences, using a combination of ICD-10 codes and self-reported diagnoses. Exome sequences from biobank participants without cardiovascular diseases and participants in non-cardiovascular clinical trials were used as controls. We tested the aggregate effect of rare non-synonymous variants within a gene by performing collapsing analyses - under 10 different genetic models - per ancestral group across four analysis strata (i.e., three for biobanks and one for clinical trials). We used Fisher's exact test to compare the frequency of qualifying variants between HF cases and controls. For each gene, we then meta-analysed association statistics across ancestral groups and strata [cases=30,648, controls=258,060]. We also performed separate collapsing analyses for HF subtypes, i.e., HF with reduced ejection fraction [cases=8,100, controls=245,306] and HF with preserved ejection fraction [cases=3,194, controls=182,297]. Results: Across collapsing models, two genes achieved study-wide significance ($p<2.6x10^{-7}$), with rare variants in both previously reported to associate with HF: *TTN* ("ptv" model: OR=2.1, $p=1.0x10^{-42}$) and *LMNA* ("flexnonsyn" model: OR=1.5, $p=2.2x10^{-7}$). We also observed suggestive associations for other known HF genes, such as *FLNC* ("ptv" model: OR=4.4, $p=4.9x10^{-7}$), *MYH7* ("flexnonsynmtr" model: OR=1.3, $p=1.3x10^{-5}$)and *FBN1* ("ptv" model: OR=38.4, $p=1.6x10^{-6}$). We identified additional suggestively associated genes, which are currently undergoing further validation as potential therapeutic opportunities. No study-wide significant genes were identified in the HF subtype analyses. Conclusions: We assessed the contribution of rare genetic variation in HF, leveraging one of the largest and most diverse collections of HF cases with sequence data. Although our study did not identify any novel study-wide significant genes in relation to HF - which could partly reflect the heterogenous nature of the disease - access to larger sample sizes in the future holds the promise of providing novel human genetics-validated therapeutic targets.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1754 Rare variant analysis of MRI-derived fat distribution phenotypes strengthens detected effects compared to larger meta-analysis.

Authors:

L. Ganel[1], B. Geraghty[1], K. Landheer[2], P. Dornbos[1], P. Akbari[1], J. Bovijn[1], N. Verweij[1], L. Lotta[1], M. Germino[2], J. Marchini[1], A. E. Locke[1]; [1]Regeneron Genetics Ctr., Tarrytown, NY, [2]Regeneron Pharmaceuticals, Tarrytown, NY

While body mass index is an important predictor of obesity-associated disease outcomes, the distribution of fat is independently predictive of cardiometabolic morbidity. To assess the genetic factors underlying these fat distribution traits, we used a 3-dimensional fat segmentation machine learning model to estimate the volumes of six fat depots from whole-body MRI data in UK Biobank data (N = 49,169). From these volumes we also generated three summary fat measures and three fat distribution ratios, totaling twelve fat measures for genetic analysis. After correcting for BMI, height, and several other demographic, ancestry, and technical covariates, we performed the largest gene-based rare variant analysis of these fat distribution traits to date. We identified 7 genes significantly associated with at least one fat measure by using a gene-based framework that combines burden, SBAT, SKATO and ACATV tests into a single p-value per gene (GENE_P). These gene-phenotype associations covered 6 traits and included several known fat distribution genes such as *PDE3B, ACVR1C, PLIN4, INSR*. Some canonical obesity genes, such as *MC4R* and *LEPR*, were not associated with any of our traits, likely due to the mechanisms by which these genes regulate energy balance. Of the 24 significant gene-trait pairs, 10 were driven by burden signals (either from marginal or joint modeling of burden masks), 9 by single variants, and 5 by variance component tests. Nine of these gene-trait pairs were from sex-specific tests, of which 8 were specific to males. Next, we compared our results for visceral/gluteofemoral fat ratio (VGFR) to a published meta-analysis of BMI-adjusted waist-hip ratio (WHRadjBMI) in over 600,000 individuals (Akbari *et al*., *Nat. Commun.* 2022). Of the 14 genes tested for fat measures that were identified by Akbari *et al*. for WHRadjBMI, two were statistically significant for VGFR after Bonferroni correction and nine at least reached nominal significance, despite the dramatic difference in sample size. Using robust regression to compare the VGFR and WHRadjBMI effect sizes, we observed that estimated effects were, on average, >50% larger for VGFR (β=1.51, SE 0.19, Wald test P-value 5.76×10^{-15}). Given that the WHR meta-analysis sample size was over 12 times greater, the fact that VGFR effect sizes for true positive genes tended to be significantly larger suggests that the imaging-based phenotype more accurately reflects the biological processes underlying obesity. Overall, these data support the utility of machine learning-based segmentation of imaging data for fine-scaled phenotyping in genetic studies.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1755 Rare variant association analyses reveal the significance of carbohydrate metabolism in severe adolescent idiopathic scoliosis

Authors:

W. Wen[1], S. Zhao[2,1], Z. Zhao[1], H. Zhao[1], X. Cheng[1], G. Li[1], J. Cai[1], Q. Li[1], K. Xu[1], N. Wu[1]; [1]Peking Union Med. Coll. Hosp., Beijing, China, [2]Baylor Coll. of Med., Houston, TX

Background: Adolescent Idiopathic Scoliosis (AIS) is the commonest scoliosis form, with an unknown etiology. Anomalous energy metabolism, often evidenced by a lower BMI, is frequently seen in AIS patients, suggesting its role in disease onset. However, the link between metabolism and AIS pathogenesis remains largely unclear.
Methods: We conducted whole-genome or whole-exome sequencing on 368 severe AIS (Cobb angle > 40°) patients from a Chinese Han cohort, and 4000 unrelated control samples. We performed a gene-based weighted rare variant burden test using the ACAT package, categorizing variants into distinct mask levels, with weight values assigned based on variant type and bioinformatic prediction results. Furthermore, we explored the cumulative impact of genetic variants via an aggregate mutational burden analysis of gene sets, using Molecular Signatures Database (MSigDB) pathways. To corroborate our findings, we also analyzed differential muscle expression using RNA-sequencing data from 14 AIS patients and 15 controls.
Results: Our gene-based burden analysis identified *SLC16A8*, a proton-coupled lactate transporter predominantly expressed in retinal pigment epithelial cells, as a novel candidate gene for severe AIS (P = 3.80E-06, FDR = 0.009). *SLC16A8* is implicated in the regulation of glucose metabolism via lactate accumulation. A large majority of AIS cases with deleterious SLC16A8 variants demonstrated a pronounced severe spinal curve (Cobb angle > 60°) and early-onset high myopia (> 600 degrees) preceding scoliosis. Pathway-based burden analysis revealed a significant enrichment in carbohydrate metabolism pathways, with the galactose metabolic process (GO:0006012, P = 2.93E-05) as the top signal, amino sugar and nucleotide sugar metabolism (KEGG: map00520, P = 4.04E-05), carbohydrate derived catabolic process (GO:1901136, P = 6.46E-05) and deoxyribose phosphate catabolic process (GO:0046386, P=7.06E-05) as other key signals. Patients with deleterious variants in carbohydrate metabolism-related genes demonstrated a significantly larger spinal curve (P = 0.015). Concordantly, genes related to the positive regulation of catabolic processes and response to nutrient levels displayed significantly divergent expression between AIS cases and controls, substantiating the preliminary genomic findings.
Conclusion: Our findings indicate that genetic variants influencing carbohydrate metabolism could involve in the development and progression of AIS, thereby contributing a novel insight to the understanding of this disease.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1756 Rare variant contribution to the heritability of coronary artery disease based on 22,443 whole genome sequences from the NHLBI TOPMed Program.

Authors:

G. Rocheleau[1,2,3], S. Clarke[4,5], G. Auguste[6], J. Rotter[7], C. Miller[6,8,9], T. Assimes[4,5,10], P. De Vries[11], R. Do[1,2,3], NHLBI TOPMed Atherosclerosis Working Group; [1]The Charles Bronfman Inst. for Personalized Med., Icahn Sch. of Med. at Mount Sinai, New York, NY, [2]Dept. of Genetics and Genomic Sci., Icahn Sch. of Med. at Mount Sinai, New York, NY, [3]Ctr. for Genomic Data Analytics, Icahn Sch. of Med. at Mount Sinai, New York, NY, [4]Dept. of Med. (Div. of Cardiovascular Med.), Stanford Univ. Sch. of Med., Stanford, CA, [5]VA Palo Alto Hlth.Care System, Palo Alto, CA, [6]Ctr. for Publ. Hlth.Genomics, Univ. of Virginia, Charlottesville, VA, [7]The Inst. for Translational Genomics and Population Sci., Dept. of Pediatrics, The Lundquist Inst. for BioMed. Innovation at Harbor-UCLA Med. Ctr., Torrance, CA, [8]Dept. of Biochemistry and Molecular Genetics, Univ. of Virginia, Charlottesville, VA, [9]Dept. of Publ. Hlth.Sci., Univ. of Virginia, Charlottesville, VA, [10]Stanford Cardiovascular Inst., Stanford, CA, [11]Human Genetics Ctr., Dept. of Epidemiology, Human Genetics, and Environmental Sci., Sch. of Publ. Hlth., The Univ. of Texas Hlth.Sci. Ctr. at Houston, Houston, TX

Introduction: Twin studies and large multi-population genome-wide association studies (GWAS) have reported heritability of coronary artery disease (CAD) in the range of ~35-60%. However, these heritability estimates are based on genotyped and imputed variants, possibly missing a substantial contribution from rare variants. Whole genome sequence (WGS) data enables discovery of rare variants and can contribute a portion of missing heritability of CAD that is not explained by common variants. **Material and Methods:** To measure the contribution of rare variants to the CAD heritability, we applied the GREML-LDMS-I approach to the whole genome sequences (WGS) of 4,949 cases and 17,494 controls of European ancestry from the National Heart, Lung and Blood Institute (NHLBI) Trans-Omics for Precision Medicine (TOPMed) program. Variant were binned according to their linkage disequilibrium (LD) scores (below or above the median over all variants) and their minor allele frequencies (MAF) ($MAF \leq 0.1\%$, $0.1\% < MAF \leq 1\%$, $1\% < MAF \leq 10\%$, $10\% < MAF \leq 50\%$). We further assessed variant contributions to CAD heritability overlapping with cis-regulatory elements (CREs) using a cell-specific chromatin atlas of 13 cell types of the human coronary artery generated by single-nucleus assay for transposable-accessible chromatin with sequencing (snATAC-seq) profiling. Finally, we compared the enrichment of protein-altering (e.g. frameshift, missense) versus non-protein-altering (e.g. synonymous, intronic) variants to CAD heritability. **Results:** After stringent quality control in both the sample and variant set, we estimated the CAD heritability (on the liability scale) at 28.8% (standard error = 13.9%), assuming a population prevalence of 8.2% in sample of genetically inferred European participants. We observed that ultra-rare variants ($MAF \leq 0.1\%$) in low LD contributed the most with 62% of the total observed heritability. In contrast, common variants ($MAF > 10\%$) in high LD contributed only 18%. Ultra-rare variants in low LD demonstrated a ~5 fold enrichment towards CAD heritability in CREs for natural killer and endothelial cells, fibroblasts, fibromyocytes and pericytes, although they represent only 1% of all variants. Protein-altering variants showed a disproportionate enrichment compared to non-protein-altering variants, particularly for common protein-altering variants which contributed 2.9% of the total observed heritability with less than 0.05% of all variants. **Conclusion:** Ultra-rare variants in low LD contribute a substantial fraction to the total heritability of CAD and are enriched in CREs in specific cell types of the human coronary artery.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1757 Rare variant enrichment analysis in pediatric European Moyamoya Angiopathy patients

Authors:

E. Cabello[1,2], K. Steindl[1], I. Ivanovski[1], P. Boonsawat[1], P. Zanoni[1], N. C. Braun[1], M. Papik[1], A. Bahr[1], N. Khan[3], A. Rauch[1,4]; [1]Inst. of Med. Genetics, Univ. of Zürich, Zürich, Switzerland, [2]PhD Program Biomedicine (BioMed), Life Sci. Zurich Graduate Sch. (LSZGS), Zürich, Switzerland, [3]Moyamoya Ctr., Univ. Children's Hosp., Univ. of Zürich, Zürich, Switzerland, [4]Moyamoya Ctr., Univ. Children's Hosp., Univ. of Zurich, Zürich, Switzerland

Introduction: Moyamoya angiopathy (MMA) is a rare condition characterized by stenosis of the intracreanial internal carotid artery and its branches resulting in transient ischemic attacks and strokes in children. Although MMA has been mostly described and studied in Asia, latest reports show a continuous increase in the incidence and prevalence of MMA cases worldwide. Previous studies, mainly focused on adult patients and/or Asians, have suggested several susceptibility genes, with *RNF213* being the only recurrent one. Specifically, the p.R4810K variant in this gene has been found in East Asian populations and, when homozygous, it goes along with an earlier onset and worse prognosis. In adult Caucasian MMA patients, diverse *RNF213* rare variants have been positively associated, but not p.R4810K. **Methods:** The whole exome was sequenced from blood samples in a pediatric cohort of 105 MMA patients and 234 healthy control individuals (88% Europeans). After quality control and excluding common variants (MAF >2%) and homologous regions, we performed a rare-variant enrichment test using the R package *SKAT*. To verify the independence of our results, we repeated the test using the same Moyamoya cohort against a different control cohort of 160 individuals. **Results:** As a positive control for the validity of our approach, the results revealed significant enrichment of rare variants in the known gene *RNF213* on the top of the list in both tests (adjusted p-value=4.04×10^{-7}). There is a significant enrichment of rare variants in 23 genes in the first test (234 controls) and 16 in the second one (160 controls), 4 of them from both studies being already associated with Moyamoya (*RNF213*, *NF1*, *PCNT*, and *OBSCN*), and 7 are common in both studies, including *RNF213*. The remaining 5 genes are strong novel candidate genes for MMA susceptibility. **Conclusion:** we present an international MMA case-control study focusing exclusively on pediatric patients, mostly of European origin. The results show 5 novel candidate genes that could improve the understanding of MMA.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1758 Rare variant pathways: Analyzing gene set burden of biological pathways associated with rare variants in Parkinson's disease.

Authors:

V. Pitz[1], A. Y. Fu[2], S. Bandres-Ciga[3], A. Singleton[3,2], M. Makarious[2], C. Blauwendraat[3,1]; [1]Integrative Neurogenomics Unit, Lab. of Neurogenetics (LNG), NIH, Bethesda, MD, [2]Molecular Genetics Unit, Lab. of Neurogenetics (LNG), NIH, Bethesda, MD, [3]Ctr. for Alzheimer's and Related Dementias (CARD), NIH, Bethesda, MD

Aim: This study aims to identify biological pathways associated with Parkinson's disease (PD) based on their variant class and cumulative gene-set rare variant burden.
Background: Age, environmental factors, and genetics play a crucial role in the development of PD. The emergence of large-scale initiatives allows for the exploration of rare genetic variants (MAF<1%) as causal or significant risk factors for PD. Burden tests enable the assessment of cumulative burden of rare and protein-altering variants within specific gene sets that constitute pathways between cases and controls. This allows for the identification of gene-associated burden patterns, shedding light on biological pathways involved in PD pathogenesis.
Materials & Methods: We used data from the Accelerating Medicines Partnership Parkinson's disease (AMP-PD), comprising 2,253 cases and 2,834 controls as well as data from the UK Biobank (UKB) comprising 1,105 cases, 6,701 proxies, and 38,051 controls. Based on SnpEff and LOFTEE predictions, variant classes were extracted based on 1) Missense, 2) Moderate or high impact, 3) High confidence loss of function, and 4) Combined Annotation Dependent Depletion (CADD) PHRED score>20. We utilized canonical pathway gene-sets from the Molecular Signatures Database (MSigDB) and custom scripts based on the burden test algorithms Sequence Kernel Association Test-Optimal (SKAT-O) and Combined Multivariate and Collapsing with Weighted-Adaptive Least Absolute Deviations (CMC WALD) for each variant class and variants with minor allele frequencies (MAFs) <1% and <0.1%. Case-control analyses were performed separately and were adjusted for sex, and principal components (PCs) 1-5 in AMP-PD, while UKB was adjusted for sex, Townsend score, and PCs 1-5.
Results: A total of 3,358 cases, 6,701 proxies, and 40,885 controls were analyzed. We conducted rare and ultra-rare variant pathway analyses on 3,090 canonical pathway gene sets: UKB contributed most gene sets with all 3,090 being present, 2,679 of which were found in AMP-PD. We report pathways significantly associated with PD, based on: passing multiple test correction at P<1.6e-05, generating robust statistical outputs, and remaining significant in a meta-analysis of both data sets.
Discussion: Despite many associations found between genes and PD, our understanding of the biological processes and networks remains incomplete. Conducting the largest rare variant burden pathway analysis in PD, this study provides a valuable tool for identifying important biological pathways. The results have the potential to aid in functional prioritization and the identification of potential therapeutic targets.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1759 Rare variants enhance the ability to identify associated phenotypes in disease networks

Authors:

J. Woerner, R. Venkatesh, V. Sriram, D. Kim; Univ. of Pennsylvania, Philadelphia, PA

Complex diseases exhibit genetic associations with multiple phenotypes, but the role of rare variants in these pleiotropic connections is not well understood. Disease-disease networks (DDNs) offer insight into the relationships between complex disorders through graphical representations where nodes represent disease phenotypes and edges represent phenotype-associated genes or variants that are shared across nodes. We hypothesize that integrating rare variant information in DDNs improves our understanding of cross-disease relationships compared to using common variants alone. We constructed single nucleotide polymorphism (SNP)-based common variant and gene-based rare variant DDNs for 93 diseases using phenome-wide association study (PheWAS) summary statistics from the UK Biobank. Common variants were defined as having a minor allele frequency (MAF) &gt 0.01, while rare variants had an MAF &lt 0.001. The common variant network was then supplemented with edges exclusively derived from the rare variant network, resulting in an augmented DDN. The intersection between common and rare variant edges made up 2.09% of the total edges of the augmented DDN, indicating that rare variants uncover cross-phenotype associations not captured by common variants. The augmentation of the common variant DDN led to a 22.7% increase in information, identifying 280 new associations. Egocentric networks focusing on circulatory, endocrine, and neoplasm phenotypes were derived from the augmented DDN. We identified rare-variant edges linking myeloproliferative disease with hypervolemia, as well as type 1 diabetes with mitral valve disease. These genetic connections, well-supported in the literature, were not identifiable using data from common variants alone. These findings underscore the importance of including rare variants in addition to common variants in network analyses of polygenic diseases. Future directions involve integrating additional phenotypes into the rare variant augmented DDN and employing a graph-based semi-supervised learning approach to evaluate its translational utility.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1760 Reanalyzing legacy GWAS dataset using TOPMed imputation yields novel loci for orofacial clefting.

Authors:

X. Zhang[1], T. H. Beaty[2], E. Feingold[3], S. M. Weinberg[4], M. L. Marazita[4], J. R. Shaffer[3,4], J. C. Carlson[3,1]; [1]Dept. of Biostatistics, Sch. of Publ. Hlth., Univ. of Pittsburgh, Pittsburgh, PA, [2]Dept. of Epidemiology, Johns Hopkins Univ., Baltimore, MD, [3]Dept. of Human Genetics, Sch. of Publ. Hlth., Univ. of Pittsburgh, Pittsburgh, PA, [4]Ctr. for Craniofacial and Dental Genetics, Dept. of Oral and Craniofacial Sci., Sch. of Dental Med., Univ. of Pittsburgh, Pittsburgh, PA

Background: Genotype imputation is critical to the discovery and fine mapping of loci discovered via genome-wide association studies (GWAS). With improved representation of diverse populations in available haplotype reference panels, there is an opportunity to revisit legacy genome-wide SNP array datasets to identify any possible associations which may have been masked due to poor imputation quality or low variant frequencies in the reference panel. Here we illustrate this principle by performing a GWAS of cleft lip with or without cleft palate (CL/P) in a sample of 4,488 participants of Asian and European ancestry from 1,494 case-parent trios imputed using the TOPMed reference panel.
Methods: We used a total of 4,294,290 total variants (those genotyped were from an Illumina Human610-Quad v.1_B BeadChip; non-genotyped SNPs were newly imputed with high quality ($r2 \geq 0.80$) using the TOPMed Imputation Server) to test for association with CL/P via the transmission/disequilibrium test as implemented in PLINK.
Results: We observed genome-wide significant signals ($p < 5e\text{-}8$) for two novel loci (near RBFOX1 on chromosome 16 and the IGLV region on chromosome 22) in addition to the known clefting loci IRF6, 8q24, NTN1, and MAFB. The lead variant in the RBFOX1 locus (rs17242358, chr16:6429186 A>G [hg38], $p = 2.13e\text{-}9$) is an intronic variant in RBFOX1 with overall minor allele frequency (MAF) of 0.06; this variant is common in East Asians (gnomAD MAF = 0.12) and rare elsewhere (MAF < 0.01). RBFOX1 is expressed in the human craniofacial complex (particularly in muscle precursor, mesenchyme, ectoderm, and neuronal cells) and in brain tissue while being differentially expressed during neural tube development in mice, possibly reflecting a plausible biological role in the development of the face in early embryogenesis. The association signal in the IGLV region was led by rs141924550 (chr22:22757242 G>A, $p = 1.27e\text{-}8$), which is a non-coding transcript exon variant in lncRNA ENST00000689050, upstream of IGLV2-14 and downstream of IGLV3-15.
Conclusion: We observed two novel loci associated with risk of CL/P by updating imputation. Additional studies are necessary to validate these associations and to determine the biological underpinnings of these associations, which may point to new biological mechanisms contributing to orofacial clefting. Moreover, this strategy may be a fruitful avenue for other researchers to take full advantage of hidden associations in legacy GWAS datasets, especially those derived from collaborations using more diverse ancestries than previously represented in older imputation panels.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1761 † Reduced transferability of tobacco and alcohol use polygenic scores across populations

Authors:

G. Saunders; Univ. of Minnesota, Minneapolis, MN

Polygenic risk scores (PRS) based on a recent genome-wide association study (GWAS) of nearly 2.7 million individuals have been found to explain up to 9.6% of the variance in alcohol and tobacco use. This level of predictive power, however, is only true for individuals of European ancestries, the same ancestries upon which the discovery GWAS was based. The relative accuracy of scores derived from this same GWAS were between 0.5% and 62% in individuals of non-European ancestry, with average relative accuracies of 12% for African ancestries, 26% for East Asian ancestries, and 31% for American ancestries. This is consistent with a growing body of literature showing that the predictive power of polygenic scores declines with increasing genetic distance from the discovery sample ancestry. There are several factors that may explain the reduced transferability of PRS across populations including differences in linkage disequilibrium patterns, minor allele frequencies, and causal effect sizes. With the potential of PRS to provide clinical, research, and personal utility, understanding of the causes of reduced accuracy, and relative contributions to the problem, has important implications for equitable potential of PRS. In the current study we use summary statistics from a large-scale GWAS of tobacco and alcohol use to describe the relative accuracy of European-based PRS across diverse ancestries using >300,000 individuals from the All of Us cohort. We use two recently derived methods in samples of admixed ancestries to compare genetic effects sizes across ancestry within an individual, as well as within ancestry across individuals, to evaluate evidence for difference in causal effects sizes. This analysis is combined with a decomposition of the contribution of differences in linkage disequilibrium patterns and allele frequencies to reduce PRS accuracy across populations. Together, preliminary results suggests that differences in allele frequencies and linkage disequilibrium patterns explain a majority of reduced PRS transferability, with some evidence for small contributions of causal effect size differences. Finally, we compare these results to other medical, anthropometric, and psychiatric phenotypes to show how the pattern of results differs across human traits and behaviors. The results of this work inform effects to understand the genetic architecture of complex traits across genetic ancestries and to create more accurate polygenic risk scores for individuals from ancestries that have largely been left out of genomic research.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1762 Replication analysis of blood trait associations in the All of Us dataset

Authors:

N. Ma[1], A. Massarat[2], Y. Li[3], J. Margoliash[4], K. Frazer[5], M. D'Antonio[6], M. Gymrek[7]; [1]Dept. of Med., Univ. of California San Diego, La Jolla, CA, 92093, CA, [2]Univ. of California San Diego, La Jolla, CA, [3]Univ. of California San Diego, San Diego, CA, [4]Univ. of California, San Diego, La Jolla, CA, [5]UC San Diego, San Diego, CA, [6]Dept. of Med., Div. of BioMed. Informatics, Univ. of California San Diego, La Jolla, CA, 92093, CA, [7]Univ California San Diego, La Jolla, CA

The All of Us (AoU) cohort has the potential to enable detailed study of the influence of genetics and other factors on complex traits across diverse and admixed populations. To evaluate the feasibility of performing genome-wide association studies (GWAS) in AoU, we performed a replication analysis of association signals we and others previously identified in the UK Biobank for blood traits and serum biomarkers, including platelet count and LDL cholesterol. We first developed a data preprocessing pipeline to obtain reliable phenotype values from lab measurements by filtering outlier values, converting values to standardized units, and summarizing measurements for each sample access multiple (in some cases hundreds of) time points. We further implemented a genotype quality control (QC) pipeline for filtering individuals with low quality data, problematic variants, and low quality genotype calls from genotypes obtained from whole genome sequencing (WGS) of AoU participants. After filtering and QC, 52,458 and 36,053 individuals remained for analysis for platelet count and LDL cholesterol, respectively. Using these filtered datasets, we successfully replicated some of the strongest known blood trait associations. For example, we found a significant association between variants near CBL, with platelet count (lead variant p=9.17e-10 in AoU compared to p=3.77e-83 in UKB) and replicated an association we previously identified between repeat length of a CGG repeat in the promoter of CBL with this trait (p=3.77e-83 in UKB). Log transformed p-values from variants in this region were strongly correlated between UKB and AoU (r=0.9253, p< 10e-2000). In another example, we replicated a significant association between variants at the APOB locus with LDL cholesterol (lead variant p=2.92e-23 in AoU compared to p=2e-235 in UKB) including an association of a CAG repeat in the 5'UTR of APOB (p=1.37e-279 in UKB). Overall, our results demonstrate that some genetic associations identified in a relatively homogeneous population (White British from UK Biobank) can be replicated in the diverse AoU population. However, the reduction in power suggests that methodologies beyond ordinary least squares may need to be developed to achieve the full potential of extracting genetic associations in diverse populations.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1763 Revisiting the association of autoantibodies with *HLA-DRB1* in rheumatoid arthritis: non-shared epitope alleles *09 and *15 associated with elevated levels of anti-CCP2 antibodies.

Authors:

L. Padyukov[1], L. Panaifo[1], L. Israelsson[1], M. Hanson[1], RACI consortium, J. Askling[1], L. Alfredsson[1], L. Klareskog[1], C-L. Too[2], D. Lagutkin[1]; [1]Karolinska Inst, Stockholm, Sweden, [2]Inst. for Med. Res., Kuala Lumpur, Malaysia

Introduction: The presence of autoantibodies, such as anti-citrullinated protein antibodies (ACPAs), is a significant diagnostic marker in rheumatoid arthritis (RA). Human leukocyte antigen (HLA)-DRB1 alleles, particularly shared epitope (SE) alleles, have been implicated in the pathogenesis of RA and ACPA production. However, the levels of these autoantibodies in association studies have been considered only in relation to diagnostic threshold with analyses of ACPA positive vs. ACPA negative categories. This study aimed to investigate the association between HLA-DRB1 alleles and the actual measured levels of anti-cyclic citrullinated peptide 2 (anti-CCP2) antibodies in RA patients.
Methods: We used EIRA cohort dataset to address the connection between HLA-DRB1 alleles and autoantibody levels. Analyzed data included anti-CCP2 antibody level, HLA-DRB1 haplotype, sex, and age for totally 4393 patients. Anti-CCP2 antibody levels were measured by commercial ELISA. The association between HLA-DRB1 allele presence and anti-CCP2 antibody levels was primarily assessed using the Mann-Whitney-Wilcoxon test and then with Tobit regression, adjusting for age and sex as covariates. For the replication we employed WTCCC+NARAC combined cohort and MyEIRA cohort with the same set of statistical methods. Additionally, we assessed association of HLA-DRB1 alleles with the levels of anti-cytomegalovirus (CMV), anti-Epstein-Barr virus (EBV), and anti-parvovirus antibody in a subset of the EIRA cohort as a control.
Results: We confirmed the strong association of the levels of anti-CCP2 with SE alleles with stronger effect for HLA-DRB1*04 and *10 compared to *01 allelic group. In addition, our results showed a statistically significant and reproducible association between the presence of HLA-DRB1 alleles *09 and *15 and elevated levels of anti-CCP2 antibodies in RA patients. We also observed a trend for association with the *16 allele, but the sample size of patients carrying this allele was insufficient to draw definitive conclusions. No significant associations were detected for anti-viral antibodies except for HLA-DRB1 alleles *04 and *15 in relation to anti-EBV IgG levels.
Conclusion: This study demonstrates that elevated levels of anti-CCP2 antibodies are significantly associated with non-SE HLA-DRB1 alleles *09 and *15 in RA patients, suggesting a potential role of non-SE HLA-DRB1 alleles in autoantibody production in RA patients. This elevation is likely to be RA-specific. Further research is pending to confirm the association with the *16 allele and to elucidate the underlying mechanisms driving these associations.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1764 Rhinovirus-activated epithelial cells likely drive genetic susceptibility to childhood-onset asthma

Authors:

D. Fernandez Salinas[1,2,3,4], S. Djeddi[1,2,3], G. X. Huang[5,6], V. Rezende da Costa Aguiar[1,2,3], C. Mohanty[7], C. Kendziorski[7], J. Boyce[5,6], J. Gern[8], N. Barret[5,6], M. Gutierrez-Arcelus[1,2,3]; [1]Div. of Immunology, Boston Children's Hosp., Boston, MA, [2]Dept. of Pediatrics, Harvard Med. Sch., Boston, MA, [3]Broad Inst. of MIT and Harvard, Cambridge, MA, [4]Undergraduate Program in Genomic Sci., Ctr. for Genomic Sci., Univ. Natl. Autonoma de Mexico, Cuernavaca, Mexico, [5]Jeff and Penny Vinik Ctr. for Allergic Disease Res., Brigham and Women's Hosp., Boston, MA, [6]Dept. of Med., Harvard Med. Sch., Boston, MA, [7]Dept. of Biostatistics and Med. Informatics, Univ. of Wisconsin-Madison, Madison, WI, [8]Dept.s of Pediatrics and Med., Univ. of Wisconsin Sch. of Med. and Publ. Hlth., Madison, WI

Asthma is a complex disease caused by genetic and environmental factors. Genome-wide association studies (GWAS) have identified hundreds of variants contributing to asthma susceptibility, most of them implicating non-coding risk alleles. Studies have identified T cells as the likely driver of asthma through functional genomics, however, the role of epithelial cells remains under-studied. Furthermore, epidemiological studies for asthma indicate that wheezing caused by infection with rhinovirus (RV), which is the most frequent cause of the common cold, increases the risk of developing asthma in children. Rhinovirus infects airway epithelial cells (AECs), which are known to have roles in type 2 inflammation.

Here we hypothesized that specific cell states of airway epithelial cells may play a role in mediating genetic susceptibility to asthma. We compiled bulk and single-cell transcriptomic and epigenomic datasets of AECs exposed to different stimuli such as cytokines and viruses. We used GWAS summary statistics from 4 different cohorts involved in multiple asthma endotypes and associated diseases: childhood-onset asthma (COA), adult-onset asthma, allergy/eczema, and all asthma to characterize AEC states mediating genetic risk. We additionally used 4 unrelated well-characterized traits as positive and negative controls. We used a single-cell disease-relevance score (scDRS, Zhang, et al. Nat Genet. 2022) that identifies cells over-expressing GWAS genes in a weighted manner. We also used Linkage Disequilibrium Score-regression in Specifically Expressed Genes (LDSC-seg, Finucane, et al. Nat. Genet. 2018) to identify cell-state specific annotations with enrichment of heritability.

First, we validated that T cell-specific open chromatin and gene expression is significantly enriched in genetic risk to asthma. Then, we found that genes upregulated in RV-infected epithelial cells were associated specifically with COA (with replication in 3 independent transcriptomic datasets). Single-cell data indicates that within epithelial cells, non-ciliated subsets predominantly up-regulate COA-associated genes, specifically at 24 and 42 hours post-infection. Interestingly, while influenza virus also induced significant upregulation of asthma-associated genes in epithelial cells, SARS-CoV2 did not. We found no evidence of stimulation with cytokines (IFNa, IFNg, IL13, or IL17) upregulating asthma-associated genes in epithelial cells.

Overall, our results indicate that part of the "missing regulatory effects" for childhood-onset asthma are likely hidden in RV-activated non-ciliated epithelial cells.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1765 Risk stratification for different subtypes of ischemic stroke based on polygenic risk score and carotid intima-media thickness

Authors:

T-H. HSIAO[1], H. Tung[2], C-S. Liao[1]; [1]Dept. of Med. Res., Taichung Veterans Gen. Hosp., Taichung City, Taiwan, [2]Neurological Inst., Taichung Veterans Gen. Hosp., Taichung City, Taiwan

Rationale Ischemic stroke is a common multifactorial disease. Different subtypes, based on the TOAST classification, presented the different underlying pathogenesis. Polygenic risk scores (PRS) calculate the weighted sum of the number of the genetic variants to predict the inherent risk of the specific disease. Carotid intima-media thickness (IMT) and plaques reflect the condition of atherosclerosis, which are associated with the risk of ischemic stroke. Thus, we combined PRS and IMT to identify the subjects with highest risk of developing stroke among whom not experienced it. **Methods** Stroke patients enrolled in the stroke registry file, Taiwan Precision Medicine Initiative (TPMI) project and received carotid sonographic exams were selected. The gender and age matched non-stroke controls were selected from TPMI dataset. We selected 14 disease-associated PRSs for analysis and also compared with the IMT values and the presence of carotid plaques in stroke and control groups. We identified the PRS associated with different subtypes of ischemic stroke. Then, the selected PRS were applied for the controls, to survey if they also process the predicting power to present thicker IMT and carotid plaques. **Results** We found stroke patients had significant thicker IMT and presence of carotid plaques compared with the non-stroke controls, expect older stroke patients in TOAST 2, TOAST 3 and TOAST 5 subtype. For stroke subtype, PRS of body fat percentage are associated with *large-artery atherosclerosis stroke (OR = 2.4, p = 0.037)*, small artery disease stroke (*OR = 3.0, p = 0.018*) *and* cardio-embolic stroke (*OR = 2.6, p = 0.019*). Stroke subjects with undetermined etiology were associated with PRS of coronary artery disease (*OR = 3.4, p = 0.017*), body fat percentage (*OR = 6.0, p = 0.002*) and hip circumference (*OR = 2.8, p = 0.025*). **Conclusion** Subjects with higher core of coronary artery disease or systolic blood pressure PRSs will have higher risk of developing stroke with undetermined etiology (TOAST 5), as well as thicker IMT and presence of carotid plaques during the age of 40-60. Therefore, we can use the carotid sonographic exam as the screening tool to monitor such high stroke risk subjects based on score of coronary artery disease and systolic blood pressure PRSs. Additionally, more strict blood pressure control and active coronary disease risk reduction are suggested for these high risk subjects.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1766 RNA Editing modulates inflammation in Parkinson's disease: analysis of adenosine-to-inosine RNA editing in human iPSC-derived astrocytes and neurons.

Authors:

A. Wagen[1,2], K. D'Sa[1,2], M-L. Choi[3,2,4], S. Strohbuecker[2], G. R. Young[5], R. H. Reynolds[1], S. Garcia-Ruiz[1], Z. Chen[1], E. Gustavsson[1], A. Ghareeb[2], J. Bayne[2], J. Brenton[1], M. Grant-Peters[1,2], S. G. Rodriques[1], S. Gagliano Taliun[6], M. H. Horrocks[7], S. Gandhi[1,2], M. Ryten[8]; [1]Univ. Coll. London, London, United Kingdom, [2]The Francis Crick Inst., London, United Kingdom, [3]Korea Advanced Inst. of Sci. and Technology, Daejeon, Korea, Republic of, [4]Univ. Coll. London, Ucl, United Kingdom, [5]MRC London Inst. of Med. Sci., London, United Kingdom, [6]Univ. of Montreal, Montreal, QC, Canada, [7]The Univ. of Edinburgh, Edinburgh, United Kingdom, [8]UCL Great Ormond Street Inst. of Child Hlth., London, United Kingdom

Inflammation is increasingly implicated in the early pathogenesis and of Parkinson's disease (PD), with widespread microglial and astrocytic activation, and interferon signalling in patient samples and disease models. Cytosolic double-stranded RNA (dsRNA), evolutionarily a sign of viral infection, can trigger these responses via multiple intracellular sensors to activate the innate immune cascade. This response is dampened by adenosine-deaminase-acting-on-RNA (ADAR), and specifically the ADAR1p150 isoform, which edits RNA by deaminating adenosine to inosine (A-to-I). This negative feedback is particularly important in humans, as the expansion of retrotransposons has resulted in an increase in endogenous dsRNA that might trigger inflammation. Here we explore the contribution of A-to-I RNA editing to PD in a human induced pluripotent stem-cell (hiPSC)-derived astrocyte cell-model treated with α-synuclein oligomers (α-synO), a toxic species in PD. Using α-synO, we treated hiPSC-derived astrocytes from 5 neurotypical donors, as well as astrocytes in co-culture with hiPSC-derived neurons. We explored single-cell and bulk RNA sequencing, as well as editing rates using a pipeline optimised with JACUSA2. We found that in astrocytes, treatment with α-synO triggered a type 1 interferon response, and expansion of an inflammatory astrocytic sub-population. *ADAR* (ADAR1) expression increased by 0.90 log2fold (p.value = $1.25e^{-5}$) with an increase in the proportion of the ADAR1p150 isoform from 39 to 68%. Consistent with these findings, treatment significantly increased the number of A-to-I editing sites (p.value < $2e^{-16}$), and the proportion of editing at each site: 27,262 sites had increased editing verse 1,489 sites with decreased editing. The increase in editing preferentially affected the 3' untranslated region (p.value < $2e^{-16}$), consistent with the cytosolic location of ADAR1p150. The co-culture model followed the same pattern, though the neuronal monoculture did not, indicating astrocytes as a driver of this process. These results support an important role for innate immune responses in the pathogenesis of PD and suggest that modulation of the interferon response through RNA editing could represent a novel therapeutic strategy.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1767 Robust prediction of cancer-associated Venous Thromboembolism (VTE) using routinely-collected clinical and panel-sequencing data

Authors:

I. Moon[1], H. Jeong[1], M. Ghassemi[1], A. Gusev[2]; [1]Massachusetts Inst. of Technology, Cambridge, MA, [2]Dana-Farber Cancer Inst., Boston, MA

Venous Thromboembolism (VTE), a frequent yet fatal complication, has a significant impact on patients with active cancer, leading to elevated mortality risk, compromised quality of life, and disruption of treatment regimens. The Khorana score, a current clinical standard for predicting VTEs, has shown inconsistent performance due to the heterogeneity of risk, characterized by factors such as tumor types and cancer stages. Here, we aim to improve the prediction of cancer-associated VTE across a wide range of cancer patients by leveraging a robust optimization framework and comprehensive clinical data.

We leveraged two large cohorts of cancer patients treated at the Dana-Farber Cancer Institute (DFCI) for prediction and association: 7,286 outpatient cancer patients with Khorana scores, and 29,565 patients with panel sequencing data.

We trained DeepSurv (Katzman et al., 2016), a neural network based time-to-event prediction algorithm, to predict time-to-VTE on the first group using patient covariates and routine lab test results. We noted a significant improvement over Khorana scores across 14 tumor types (overall AUC: DeepSurv 0.703 vs. Khorana 0.612). However, tumor-type specific model performance varied, with highest accuracy for skin cancers including melanoma (AUC: DeepSurv 0.909 vs. Khorana 0.697) and lowest accuracy for prostate cancers (AUC: 0.621 vs. 0.446). This heterogeneity severely limits the deployability of the model. To address this, we plan to use a distributionally robust optimization framework, which aims to optimize worst-case loss across predefined group-specific distributions and ensures consistent performance across the groups.

In our second group of patients, who had tumor panel sequencing data, we identified multiple somatic variants with significant associations to VTE incidence: including TP53 and KRAS mutations (H.R.: 1.24, p: 6.86e-13, and H.R.: 1.41, p: 5.18e-11, respectively). Leveraging these insights, we plan to develop an integrative model incorporating both the lab test results and the panel-sequencing data. We will investigate whether it captures additional heterogeneity in patients at an elevated VTE risk.

Our initial results suggest that a model with a rich set of clinical data improves the existing clinical standards for predicting VTE, while specific somatic variants show promise in capturing additional heterogeneity. For further improvements, we aim to utilize a robust optimization method and incorporate somatic variants into the model. We anticipate that our work will enable more tailored preventative measures for those at high VTE risk, while benefiting a wider demographic of cancer patients.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1768 Schizophrenia and long-read RNA-seq: Analysis of differential isoform usage in the dorsolateral prefrontal cortex.

Authors:

A. Abrantes[1], N. Ancalade[1], S. Sekle[1], F. Memic[2], A. Dijkstra[3], E. Tseng[4], G. Sheynkman[5], A. Smit[6], P. Sullivan[1,2], P. Giusti-Rodriguez[7]; [1]Univ. of North Carolina-Chapel Hill, Chapel Hill, NC, [2]Karolinska Inst.t, Stockholm, Sweden, [3]Amsterdam UMC, Amsterdam, Netherlands, [4]Pacific BioSci.s, Menlo Park, CA, [5]Univ. of Virginia, Charlottesville, VA, [6]Vrije Univ.it Amsterdam, Amsterdam, Netherlands, [7]Univ. of Florida Coll. of Med., Gainesville, FL

Over the past 15+ years, genetic studies have been notably successful in revealing the architecture of multiple psychiatric disorders, including schizophrenia. It is now widely acknowledged that schizophrenia is highly polygenic, consisting of both common variants of unknown significance (identified through genome-wide association studies [GWAS]) and rare coding variants. Despite these advances, the roles and the underlying mechanisms of specific genes implicated by GWAS are not well understood.

Multiple lines of evidence implicate alternative splicing in the pathophysiology of schizophrenia. In this scenario, single-nucleotide polymorphisms (SNPs) residing in GWAS implicated loci for schizophrenia could affect changes in the distribution of isoform expression within a gene. These genes would not necessarily be differentially expressed in a typical DEG study. Hence, we generated a comprehensive isoform survey of postmortem human dorsolateral prefrontal cortex (DLPFC) from schizophrenia cases and controls and identified case-control isoform-level differences. We combine the strength of PacBio SMRT long-read RNA sequencing in conducting detailed isoform census with the capacity of short-read RNA sequencing for the quantification of isoform abundances. We then implement a case vs. control differential isoform usage analysis (DIU) using a combination of established and in-house pipelines.

From several hundred thousand discovered long-read isoforms we curated a transcriptome with tens of thousands of high confidence novel isoforms. Many of these novel isoforms are differentially expressed in schizophrenia DLPFC vs healthy controls. Differentially expressed genes are enriched in gene sets related to synaptic structure and function, RNA binding and splicing, as well as cell types previously implicated in schizophrenia. We use publicly available splicing data, genotyping, proteomics, and single nucleus sequencing results from colleagues to verify and support our results.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1769 † Screening for polygenic risk for eight common conditions in an unselected and diverse primary care patient population

Authors:

S. Suckiel[1], E. R. Soper[1], K. E. Bonini[1], P. N. Marathe[1], J. A. Odgis[1], T. I. Gray[1], N. S. Abul-Husn[2], E. E. Kenny[1]; [1]Icahn Sch. of Med. at Mount Sinai, New York, NY, [2]23andMe, Sunnyvale, CA

Background: There is growing evidence of the predictive abilities of polygenic risk scores (PRS) for common diseases. Applying PRS clinically could improve health outcomes by identifying at-risk individuals and informing personalized medical care. Nevertheless, clinical PRS are just emerging, and there is little data on their application to identify genomic risk at population scale.
Methods: The NHGRI-funded eMERGE IV Network is implementing cross-ancestry PRS for eight common conditions to be reported to adult (18-75 years) participants as part of a comprehensive genome-informed risk assessment. The conditions include atrial fibrillation (AF), breast cancer (BC), chronic kidney disease (CKD), coronary heart disease (CHD), hypercholesterolemia (HC), obesity (OB), prostate cancer (PC), and type 2 diabetes (T2D). High-risk PRS thresholds were set at the top 2-10%, depending on the condition. We evaluated the prevalence of high-risk PRS in the first 1,131 participants recruited from primary care clinics at one eMERGE IV site (Mount Sinai). We also performed chi-square testing to compare rates of high-risk PRS by sex and examined disease occurrence in participants who had related high-risk PRS and were <50 years old.
Results: Participants were 71% (805) female and the mean age was 47 (SD=16). Most self-reported as Hispanic/Latino/Spanish (578, 51%), White/European American (234, 21%), or Black/African American/African (160, 14%). Of the 1,131 participants, 23% (260) had a high-risk PRS for one of the eight conditions, 3% (33) for two conditions, and 0.3% (3) for three conditions; 74% (835) did not have a high-risk PRS for any of the conditions. There were 6% (63) with high-risk PRS for CHD, 5% (52) OB, 4% (48) AF, 3% (38) CKD, 3% (32) T2D, and 2% (26) HC. There were no differences in high-risk PRS rates between males and females for these six conditions. In addition, 10% (31/326) of males had a high-risk PRS for PC, and 6% (45/805) of females had a high-risk PRS for BC. Among participants <50 years, no one with a related high-risk PRS reported having AF, BC, CKD, or PC, and only one had CHD or T2D. However, a considerable number of participants <50 with a related high-risk PRS had OB (10/22, 45%) or HC (6/15, 40%).
Conclusions: In a cohort of unselected adults recruited from primary care, nearly a quarter had a high-risk PRS result for at least one of eight conditions. Except for OB and HC, most participants <50 with high-risk PRS did not have the related condition. These findings suggest that clinical PRS can help identify at-risk individuals prior to disease onset when risk-management strategies could positively impact prevention.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1770 Search strategy for oligogenic models of cardiomyopathy

Authors:

S. Batalov[1], M. Bainbridge[2]; [1]Rady Children's Inst. for Genomic Med., San Diego, CA, [2]Rady Children s Hosp., San Diego, CA

Cardiomyopathies (CM) are the leading cause of cardiac death, after hypertension and coronary artery disease, with over 200,000 new cases per year. Cardiomyopathy, including hypertrophic, dilated and left ventricular non-compaction, is caused by >40 genes, some of which can cause multiple forms of CM. Genetic CM is inherited in an autosomal dominant fashion and, like many other autosomal dominant diseases is often highly variable in penetrance and severity. It has long been recognized that genetic interactions (epistasis) play a role in explaining this phenotype. Increasingly, in the past 5 years, there have been numerous clinical reports of digenic inheritance and genetic modifiers (secondary mutations) of CM in small pedigrees and cohorts. Previous studies provide compelling observations that in some cases multiple variants can act epistatically to cause CM or to cause more severe CM but demonstrating the causal impact of polygenic inheritance requires experimental evidence.
We present a prioritization algorithm to identify interacting variants that are good candidates for epistasis. RCIGM has a CAP/CIiA accredited clinical lab and generated DNA libraries and whole genome sequencing data for >2,500 samples, with CM cohort comprising ~15% of all clinical patients (selected using defined set of 636 CM-related HPO terms). These data have been aligned to reference genome and variants called using the DRAGEN pipeline, annotated using CASSANDRA software to determine the effect of each mutation (intronic, nonsynonymous etc.), the evolutionary conservation, and in silico predicted deleteriousness, using REVEL (and a combination MutPred, FATHMM, VEST, PolyPhen, SIFT, PROVEAN, MutationAssessor, MutationTaster, and additional methods). Variants are filtered based on MAF from gNomad and our own sample database, to be above 0.1% in at least one subpopulation but not to exceed 1% in any sub-population. Gene (protein-protein) interactions are inferred using STRINGdb (v.11.5) combined score. Together, these factors enrich for mutations that are uncommon but alter evolutionarily conserved amino acids in pathogenic mutation-rich regions. These methods are refined such that known epistatic interactors are prioritized (1,059 CM-related gene targets from literature: including Richter 2020, Szot 2018, initially gated to only near-coding positions, within 12 bp of all coding exons), with significance p-values < 10^-9 using permutation analysis.
Using the prioritization algorithm and the existing cardiomyopathy cohort at RCIGM as well as public databases we identified good candidates for functionalization in vivo/in vitro in mice (in progress).

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1771 SERPINE3 is an RPE-specific protease inhibitor critical for retinal function

Authors:

H. Li[1], D. McGaughey[2], L. Dong[1], J. Liu[1], T. Johnson[1], R. Hufnagel[3]; [1]NIH NEI, Bethesda, MD, [2]Johns Hopkins Univ, Baltimore, MD, [3]Natl. Eye Inst., NIH, Bethesda, MD

Purpose: Serpin family E member 3 (SERPINE3) is a putative serine proteinase inhibitor with unknown function in mammalian retina. Here, we define the cellular expression of SERPINE3 in human and mouse, define target proteases, and explore its function in a novel mouse model. **Methods:** Human and macaque expression data from eye-Integration and plae were analyzed. Protease assay using fluorescently-labeled peptide substrate was performed for screening the SERPINE3 against all major classes of proteases. *Serpine3-VenusGFP* knockin mice were generated using CRISPR-mediated homologous recombination in mouse zygotes, replacing exon 2 with a Venus reporter. For analyzing the function of Serpine3 in RPE regeneration, sodium-iodate injury model, anti-ZO1 staining and anti-cell proliferation specific markers' staining were performed. **Results:** *SERPINE3* is specifically expressed in human and macaque retinal pigment epithelium (RPE). Through compartment analysis, SERPINE3 protein was observed in cytoplasm and extracellular vesicles of human iPSC-derived RPE. In mouse eye, Serpine3 protein and *Serpine3-Venus*GFP reporter expression were localized exclusively to RPE, and the GFP+ cells were able to reenter cell cycle after sodium-iodate injury. *Serpine3*$^{Gfp/Gfp}$ homozygous mice are viable. Protease inhibition screen of 75 candidates identified Caspase 1, Caspase 2, Caspase 3, and Caspase 5 as putative SERPINE3 targets. **Conclusion:** In mammalian retina, SERPINE3 is expressed specifically in the RPE and a high level in the peripheral mature RPE. The subtype of Serpine3-GFP+ RPE cells are able to reenter cell cycle under the sodium iodate treatment. Deletion of Serpine3 in mouse accelerates sodium iodate-induced RPE degeneration. SERPINE3 targets include multiple caspases implicated in retinal inflammation and cell death structural retinal abnormalities, supporting a protective role of SERPINE3 to prevent RPE degeneration.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1772 Severe COVID-19 infection Increases Long-term Cardiovascular Risk Through Gene-Pathogen Interaction with ABO Blood Type

Authors:

J. Hilser[1], N. Spencer[1], K. Afshari[1], F. Gilliand[1], H. Hu[1], A. Deb[2], A. Lusis[2], W. Tang[3], J. A. Hartiala[1], S. Hazen[3], H. Allayee[1]; [1]Univ. of Southern California, Los Angeles, CA, [2]Univ. of California Los Angeles, Los Angeles, CA, [3]Cleveland Clinic, Cleveland, OH

Background: COVID-19 is associated with increased post-acute risk of major adverse cardiac events (MACE), such as myocardial infraction (MI), stroke, or all-cause mortality. However, it is not known whether increased cardiovascular disease (CVD) risk in COVID-19 patients persists beyond one year after infection and whether genetic factors modulate such risk. **Methods:** COVID-19 cases were identified in the UK Biobank based on positive PCR tests for SARS-CoV-2 infection (n=8,167) or ICD-10 codes for COVID-19 (n=1,872) between February 1, 2020 and December 31, 2020. Controls (n=217,591) were defined as subjects without a PCR^+ test for SARS-CoV-2 infection or who were never assigned a COVID-19 ICD-10 code during the same interval. Cox proportional hazards models were used to evaluate association between COVID-19 and long-term risk of MACE with follow-up to October 31, 2022 (1003 days). A candidate gene approach was used to test if COVID-19 modulated risk of thrombotic events (MI or stroke) through interactions with ABO blood type and previously identified variants for COVID-19 severity (rs11385942), susceptibility to SARS-CoV-2 infection (rs73062389), or coronary artery disease (CAD) (rs4977574). **Results:** Among all cases and controls, COVID-19 was associated with significantly increased risk of MACE, which was more pronounced among severe COVID-19 cases (HR=4.14, 95% CI 3.77-4.55; $P=2.0x10^{-194}$). Compared to patients with CVD but without COVID-19, risk of MACE among severe COVID-19 cases without CVD was even higher (HR=1.23, 95% CI 1.09-1.38; $P=1.0x10^{-3}$). Severe COVID-19 also increased risk MI or stroke to a greater in subjects with non-O blood types (HR=2.29, 95% CI 1.85-2.84; $P=2.5x10^{-14}$) than subjects with blood type O (HR 1.27, 95% CI 0.92-1.76; P=0.15), resulting in a significant interaction between the *ABO* locus and COVID-19 ($P\text{-int}=3.2x10^{-3}$). By comparison, COVID-19 did not modulate risk of incident MI or stroke as a function of variants for severe COVID-19 (P-int=0.75), SARS-CoV-2 infection susceptibility (P-int=0.59), or CAD (P-int=0.46). **Conclusions:** Risk of MACE in COVID-19 patients remained significantly elevated nearly 3 years after SARS-CoV-2 infection, with severe COVID-19 representing a CAD risk equivalent. Increased risk of thrombotic events was particularly evident among severe COVID-19 with non-O blood types and would represent one of the first known gene-pathogen exposure interactions for CVD outcomes. Additional studies will be needed to determine the underlying biological mechanism(s) for the genetic interaction between *ABO* blood type and severe COVID-19 and whether risk of MACE persists beyond 3 years in COVID-19 patients.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1773 Sex chromosome overdosage impairs neuronal synapses in human iPSC-derived neurons and brain organoids.

Authors:

A. Adamo, V. Astro, L. V. Cortez Medina, K. Cardona Londono, R. Alghamdi; King Abdullah Univ. of Sci. and Technology, Thuwal, Saudi Arabia

Sex chromosome overdosage in Klinefelter Syndrome (KS; 47,XXY), Jacob Syndrome (JS; 47,XYY), and high-grade sex-chromosome aneuploidies (48,XXXY and 49,XXXXY) leads to a shared broad spectrum of clinical features, including motor, speech, and language delays. The severity of intellectual disabilities directly correlates with the number of extra sex chromosomes. However, the in vitro modelling of these conditions is, to date, largely unexplored. We investigated the impact of sex-chromosome aneuploidies during early neurodevelopment using an induced pluripotent stem cells-based disease-modelling approach. We postulated that the neurodevelopmental deficits in KS, JS and high-grade SCAs subjects lie in the overdosage of escape genes located within the pseudoautosomal region (PAR) of the X and Y chromosomes. We used a paradigmatic cohort of KS, JS, high-grade SCAs, and 46,XY iPSCs to obtain 2D neuronal cultures and 3D cortical brain organoids. Combining a transcriptomic and functional approach, we demonstrated that sex chromosome overdosage negatively impacts the electrophysiological properties of iPSC-derived neurons in a dosage-sensitive manner. Our unique iPSC cohort represents an ideal platform to investigate the molecular and cellular impact of X and Y overdosage during early development and could serve as a clonal cellular model to study X inactivation erosion.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1774 Sex differences in genetic architecture of clinical laboratory tests: A lab wide analysis in the electronic health record

Authors:

F. Blostein[1], A. Lake[2,1], M. Niarchou[1], J. Kim[1], P. Straub[1], B. Bose[3], J. L. Winters[1], D. Velez Edwards[1], B. Stranger[3], L. Davis[1]; [1]Vanderbilt Univ. Med. Ctr., Nashville, TN, [2]Vanderbilt Univ. Sch. of Med., Nashville, TN, [3]Northwestern Univ., Chicago, IL

Sex differences exist in the prevalence, incidence, and presentation of heritable, complex diseases. Similarly, sex differences exist in the distributions of some clinical laboratory tests, and in their utility as disease biomarkers. To date, the genetic etiology of these differences is understudied and poorly understood. We investigate sex differences in genetic architecture for over 1,000 clinical laboratory tests among approximately 94,000 genotyped patients in the Vanderbilt University Medical Center electronic health record (BioVU). We perform sex-stratified genome-wide association analyses for clinical laboratory median values and variance, followed by sex*SNP interaction analyses for those SNPs that have significant main effects in at least one sex. Because sex and gender correlate, and sociocultural forces create differences in health care behaviors and treatment, we additionally evaluate differences in ordering of clinical laboratory tests for male and female patients and assess epidemiological methods to address potential ascertainment bias (e.g., inverse probability weighting). Accounting for and understanding sex differences in the genetic architecture of laboratory test values can improve their sensitivity and specificity as clinical biomarkers. Further, understanding sex differences in the genetic etiology of clinical laboratory tests may elucidate sex differences in disease mechanisms.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1775 Sex differences in the genetic architecture of Alzheimer's disease cognitive endophenotypes.

Authors:

J. M. Eissman[1,2], **A. N. Smith**[1], S. Walters[1], S. Mukherjee[3], M. L. Lee[3], S-E. Choi[3], P. Scollard[3], E. H. Trittschuh[4,5], J. B. Mez[6], W. S. Bush[7], B. W. Kunkle[8], A. C. Naj[9,10], K. A. Gifford[1], M. Bilgel[11], A. B. Kuzma[10], The Alzheimer's Disease Neuroimaging Initiative (ADNI), Alzheimer's Disease Genetics Consortium (ADGC), The Alzheimer's Disease Sequencing Project (ADSP), M. L. Cuccaro[8], C. Cruchaga[12,13], M. A. Pericak-Vance[8], L. Farrer, A[6,14,15], L-S. Wang[10], G. D. Schellenberg[10], R. P. Mayeux[16,17,18], J. L. Haines[7], A. L. Jefferson[1], W. A. Kukull[19], C. D. Keene[20], A. Saykin, J[21,22], P. M. Thompson[23], E. R. Martin[8], M. S. Albert[24], S. C. Johnson[25], C. D. Engelman[25], L. Ferrucci[26], D. A. Bennett[27], L. L. Barnes[27], J. A. Schneider[27], S. M. Resnick[11], R. A. Sperling[28], P. K. Crane[3], T. J. Hohman[1,2], L. Dumitrescu[1,2]; [1]Vanderbilt Memory & Alzheimer's Ctr., Nashville, TN, [2]Vanderbilt Genetics Inst., Nashville, TN, [3]Dept. of Med., Univ. of Washington, Seattle, WA, [4]Dept. of Psychiatry and Behavioral Sci., Univ. of Washington Sch. of Med., Seattle, WA, [5]VA Puget Sound Hlth.Care System, GRECC, Seattle, WA, [6]Dept. of Neurology, Boston Univ. Chobanian & Avedisian Sch. of Med., Boston, MA, [7]Cleveland Inst. for Computational Biology, Dept. of Population and Quantitative Hlth.Sci., Case Western Reserve Univ., Cleveland, OH, [8]John P. Hussman Inst. for Human Genomics, Univ. of Miami Miller Sch. of Med., Miami, FL, [9]Dept. of Biostatistics, Epidemiology, and Informatics, Univ. of Pennsylvania Perelman Sch. of Med., Philadelphia, PA, [10]Penn Neurodegeneration Genomics Ctr., Dept. of Pathology and Lab. Med., Univ. of Pennsylvania Perelman Sch. of Med., Philadelphia, PA, [11]Lab. of Behavioral NeuroSci. Natl. Inst. on Aging NIH, Baltimore, MD, [12]Dept. of Psychiatry, Washington Univ. Sch. of Med., Saint Louis, MO, [13]NeuroGenomics and Informatics Ctr., Washington Univ. Sch. of, Saint Louis, MO, [14]Dept. of Biostatistics, Boston Univ. Sch. of Publ. Hlth., Boston, MA, [15]Dept. of Med. (BioMed. Genetics), Boston Univ. Chobanian & Avedisian Sch. of Med., Boston, MA, [16]Columbia Univ., New York, NY, [17]The Taub Inst. for Res. on Alzheimer's Disease and The Aging Brain, Columbia Univ., New York, NY, [18]The Inst. for Genomic Med., Columbia Univ. Med. Ctr. and The New York Presbyterian Hosp., New York, NY, [19]Dept. of Epidemiology, Sch. of Publ. Hlth., Univ. of Washington, Seattle, WA, [20]Dept. of Lab. Med. and Pathology, Univ. of Washington, Seattle, WA, [21]Dept. of Radiology and Imaging Services, Indiana Univ. Sch. of Med., Indianapolis, IN, [22]Dept. of Med. and Molecular Genetics, Sch. of Med., Indiana Univ., Indianapolis, IN, [23]Keck Sch. of Med., Univ. of Southern California, Los Angeles, CA, [24]Dept. of Neurology, Johns Hopkins Univ. Sch. of Med., Baltimore, MD, [25]Alzheimer's Disease Res. Ctr., Univ. of Wisconsin Sch. of Med. and Publ. Hlth., Madison, WI, [26]Johns Hopkins Univ. Bloomberg Sch. of Publ. Hlth., Baltimore, MD, [27]Rush Alzheimer's Disease Ctr., Rush Univ. Med. Ctr., Chicago, IL, [28]Dept. of Neurology, Massachusetts Gen. Hosp./Harvard Med. Sch., Boston, MA

Sex and gender differences are present across multiple domains of cognition in aging and Alzheimer's disease (AD). For example, clinically healthy women show cognitive advantages in episodic memory and verbal fluency tasks, whereas men show advantages in visuospatial tasks. However, women tend to lose cognitive advantages to men during disease, experiencing more rapid cognitive decline. Additionally, genetic studies of AD cognitive endophenotypes have identified loci that associate with cognition, but these studies have not considered sex/gender. In this study, we investigated if genetic contributors to cognition in older adults differed by sex, leveraging harmonized memory, executive functioning (EF), language, and global cognition scores across 9 cohort studies of aging and AD. For each domain, we conducted sex-stratified and sex-interaction genome-wide meta-analyses, gene-set analyses, and genetic correlation tests among a cross-ancestral sample of over 35,000 older adults (mean age at baseline=73 years, 57% female, 14% AD cases, average number of observations per participant>4, 88% non-Hispanic White and 12% non-Hispanic Black). Linear regression associations covaried for age and ancestry principal components. Beyond the well-known *APOE* locus, we identified 3 sex-specific genome-wide loci. First, *HS3ST4*, associated with global cognitive decline among males (rs10852291: β_M=-5.29x10^{-3}, P_M=4.73x10^{-8}, β_{int}=6.26x10^{-3}, P_{int}=2.77x10^{-6}) but not among females (β_F=7.70x10^{-4}, P_F=0.35). This locus also showed a male-specific association with fluid intelligence in the UK Biobank (β_M=-2.30x10^{-2}, P_M=7.49x10^{-3}; β_F=1.44x10^{-3}, P_F=0.86). Second, *TMEM245/CTNNAL1* interacted with sex on baseline EF (rs112083348: β_{int}=-0.10, P_{int}=2.67x10^{-8}; β_M=0.05, P_M=5.18x10^{-4}; β_F=-0.05, P_F=4.92x10^{-5}). This locus was also associated in a UK Biobank sex-aware GWAS of an EF-related task (β_M=1.63x10^{-2}, P_M=6.07x10^{-3}; β_F=5.88x10^{-3}, P_F=0.28). Furthermore, rs112083348 is an eQTL for *CTNNAL1* in two brain regions affected in AD: the hippocampus and the dorsolateral prefrontal cortex. Finally, *VRK2* associated with language decline among females (rs11898834: β_F=2.75x10^{-3}, P_F=9.99x10^{-10},

$\beta_{int}=2.94x10^{-3}$, $P_{int}=6.93x10^{-5}$) but not among males ($\beta_M=-7.50x10^{-5}$, $P_M=0.89$). Notably, rs11898834 is an eQTL for *VRK2* in whole blood, and is a top locus in a sex-aware GWAS on developmental stuttering. Overall, our results are evidence of a sex-specific component to the genetic architecture of cognitive performance. We look forward to our next steps of incorporating bulk and single-cell transcriptomic analyses into our sex-aware exploration of cognition.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1776 Sex differences in the pleiotropy of hearing problems with imaging-derived phenotypes: a brain-wide imaging study.

Authors:

J. He, B. Cabrera-Mendoza, G. A. Pathak, D. Koller, R. Polimanti; Yale Univ., New Haven, CT

Background: Hearing problems (HP) are one of the major health burdens in the elderly. While changes in the peripheral auditory system play an important role, genetic variation associated with brain structure and function could also be involved in HP predisposition. **Methods**: We analyzed a large-scale HP genome-wide association study (GWAS; Ntotal=501,825, 56% females) and GWAS data related to 3,935 brain imaging-derived phenotypes (IDPs) assessed in up to 33,224 individuals (52% females). To investigate systematically HP pleiotropy with brain structure and function, i.e., loci affecting both traits, we conducted genetic correlation, latent causal variable (LCV), Mendelian randomization (MR), and multivariate logistic regression analyses. Additionally, we performed local genetic correlation and multi-trait colocalization analyses to identify genomic regions and loci implicated in the pleiotropic mechanisms shared between HP and brain IDPs. **Results**: We observed a widespread genetic correlation of HP with multiple IDPs in the sex-combined analysis (Ntraits=171) and in females (Ntraits=120) and males (Ntraits=89), separately. Applying Bonferroni correction accounting for the number of IDPs tested, the LCV analyses showed that some of these genetic correlations could indicate cause-effect relationships. For seven of them, the possible causal effects were supported by an independent MR approach: vessel volume in the sex-combined analysis, hippocampus volume, cerebellum grey matter volume, primary visual cortex volume, and rfMRI-ICA100 node 46 in females, and global mean thickness and mean orientation dispersion index in superior corona radiata in males. The local genetic correlation analyses identified 13 pleiotropic regions of HP with respect to these seven IDPs. We also observed a colocalization signal for the rs13026575 variant between HP, primary visual cortex volume, and *SPTBN1* transcriptomic regulation in females. **Conclusion**: This study provides evidence that brain structure and function may have a role in HP predisposition via possible cause-effect relationships and shared regulatory mechanisms.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1777 Sex-specific Associations of Gene Expression in Brains with Alzheimer's Disease Pathology and Cognitive Performance within X Chromosome.

Authors:

M. Clifton[1], M. Seto[1,2], R. Buckley[2,3,4], K. Gifford[1], A. Jefferson[1], P. De Jager[5,6], D. Bennett[7], Y. Wang[7], L. Barnes[7], J. A. Schneider[7], T. J. Hohman[1,8], L. Dumitrescu[1,8]; [1]Vanderbilt Memory and Alzheimer's Ctr., Vanderbilt Univ. Med. Ctr., Nashville, TN, [2]Ctr. for Alzheimer's Res. and Treatment, Dept. of Neurology, Brigham and Women's Hosp./Harvard Med. Sch., Boston, MA, [3]Dept. of Neurology, Massachusetts Gen. Hosp./Harvard Med. Sch., Boston, MA, [4]Melbourne Sch. of Psychological Sci., Univ. of Melbourne, Melbourne, Australia, [5]Ctr. for Translational & Computational Neuroimmunology, Dept. of Neurology, Columbia Univ. Med. Ctr., New York, NY, [6]Cell Circuits Program, Broad Inst., Cambridge, MA, [7]Rush Alzheimer's Disease Ctr., Rush Univ. Med. Ctr., Chicago, IL, [8]Vanderbilt Genetics Inst., Vanderbilt Univ. Med. Ctr., Nashville, TN

Alzheimer's disease (AD) disproportionately affects women, who make up two-thirds of all clinical cases of AD. Despite known sex differences in AD neuropathology, the response to pathology, and the genetic predictors of clinical AD, sex differences in the brain transcriptomic signatures of AD endophenotypes have not been fully characterized, especially the contribution of the X chromosome. We leveraged bulk RNA-sequencing data from three brain regions [dorsolateral prefrontal cortex (DLPFC), posterior cingulate cortex (PCC), and caudate nucleus (CN)] from the Religious Orders Study and Rush Memory and Aging Project (ROS/MAP). Propensity scoring was used to match male to female participants due to disproportionate sample sizes. As a result, there were 791 participants (50% male; mean age at death=88 years). Gene-expression from approximately 500 genes on the X chromosome were included sex-stratified regression models assessing tissue-specific associations with amyloid and tau burden at autopsy, along with longitudinal global cognition. Age at death and post-mortem interval were included as covariates. Sex-specific genes were defined as those associated with a trait in only one sex (FDR-corrected $P<0.05$). Of the 724 significant associations, 33% were sex-specific. We observed more female-specific (26%) compared to male-specific (7%) effects. Interestingly, across all tissues, fewer sex-specific associations were observed with amyloid (3%) compared to tau tangles (14%) or longitudinal cognition (16%), which supports the hypothesis that sex differences are more likely to appear downstream after the onset of amyloid accumulation. Several genes showed particularly strong effects among females in two of the three brain tissues, including DYNLT3 ($P_M>0.11$, $P_F<8.67x10\text{-}4$) and ATRX ($P_M<0.1$, $P_F>0.003$) with tau tangles. Additionally, female-specific associations were observed with POU3F4 expression with both cognition ($P_M>0.06$, $P_F<0.002$) and tau tangles ($P_M>0.15$, $P_F<0.003$) in both DLPFC and CN. Conversely, male-specific associations were observed between TRMT2B ($P_M<6.51x10\text{-}4$, $P_F>0.07$) and tau tangles within the DLPFC and CN. Our results highlight X chromosome transcriptomic associations with AD endophenotypes, including ATRX, which has been related to hippocampal-dependent memory impairments in male mice. Similar sex-specific patterns with AD were observed in independent AMP-AD cohorts for DYNLT3, POU3F4, and TRMT2B. These findings highlight the exciting potential of the often overlooked X chromosome to select new targets for mechanistic evaluation.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1778 † Sex-stratified analysis of adipose cell-types and cell states identify genes underlying transitions from obesity to its key comorbidity, type 2 diabetes

Authors:

Z. Wang[1,2], S. Lee[1], M. Alvarez[1], S. Rajkumar[1], A. Kar[1], K. Gelev[1], S. Heinonen[3], B. W. d. Kolk[3], U. Säiläkivi[4], T. Saarinen[4], K. H. Pietiläinen[3,5], P. Pajukanta[1,2,6]; [1]Dept. of Human Genetics, David Geffen Sch. of Med. at UCLA, Los Angeles, CA, [2]Bioinformatics InterDept.al Program, UCLA, Los Angeles, CA, [3]Obesity Res. Unit, Res. Program for Clinical and Molecular Metabolism, Faculty of Med., Univ. of Helsinki, Helsinki, Finland, [4]Dept. of Abdominal Surgery, Abdominal center, Helsinki Univ. Hosp. and Univ. of Helsinki, Helsinki, Finland, [5]Hlth.y Weight Hub, Abdominal center, Helsinki Univ. Hosp. and Univ. of Helsinki, Helsinki, Finland, [6]Inst. for Precision Hlth., David Geffen Sch. of Med. at UCLA, Los Angeles, CA

Sex affects gene expression in subcutaneous adipose tissue (SAT). However, it is not known which genes underly cell-type and subcell-type (i.e. cell state) level sex-specific differences in SAT gene expression, and whether they contribute to transitions from obesity to its key comorbidity, type 2 diabetes (T2D). To address this knowledge gap, we conducted a large single nucleus RNA-sequencing study on SAT from 69 Finns with obesity (n=20 males, n=49 females), including 6 males and 7 females with obesity and normal glucose levels (ObNoC); 6 males and 12 females with obesity and prediabetes (ObPrD); and 6 males and 22 females with obesity and T2D (ObT2D). In our adipocyte differential expression (DE) analysis between all 20 males and 49 females, we observed 508 genes (adjp<0.05) upregulated in males. We hypothesized that the sex-specific genes reflect differences in cell states by sex, and that these differ by obesity comorbidities and transitions to comorbidities (TTC). We first identified differentially abundant cellular neighborhoods (DA) between the males and females using Milo (controlling cell counts and ages; FDR<0.1). In the DA analysis (n=69), we found prominent sex-specific DAs in all adipose cell-types, particularly in adipocytes. The adipocyte DAs more abundant in females had 1,263 DE genes (DEGs) with transcription factor (TF) motif enrichment in a key adipogenesis TF, *CEBPD*, using Homer. Next, we performed stratified analysis by comorbidity. In ObNoC, the DEGs from DAs more abundant in males (n=181 DEGs) or more abundant in females (n=539 DEGs) did not show gene ontology (GO) or TF motif enrichments. In ObPrD, the T cell DAs more abundant in males contained 3 DEGs, *TXK*, *LEF1*, and *BACH2*; and endothelial cell DAs more abundant in males had 953 DEGs with a TF motif enrichment in *NFIC*. In ObT2D, the adipocyte DAs more abundant in males contained 2,213 DEGs with GO enrichments in protein modification and autophagy, and TF motif enrichments in *PKNOX1* and *TFAP2C*. In the TTC analysis, we observed DAs in the following transitions in females: ObNoC to ObT2D; and ObPrD to ObT2D. From ObNoC to ObT2D, the adipocyte DAs more abundant in ObNoC had 923 DEGs with TF motif enrichments in *HINFP*, *RELB*, and a T2D GWAS gene, *TFAP2B*, while the 586 DEGs from the adipocyte DAs more abundant in ObT2D showed TF motif enrichment in *TCF3*, and were enriched for female glycated hemoglobin GWAS SNPs using Magenta (FDR<0.05). From ObPrD to ObT2D, the adipocyte DAs more abundant in ObPrD had 735 DEGs. In conclusion, our analyses identify sex-specific genes at the SAT cell-type and cell-state resolution, as well as female DAs with DEGs for transitions from obesity to its key comorbidity, T2D.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1779 † Sex-stratified rare-variant association analysis using 419,000 UK Biobank exomes reveals sex-biased genetic associations with disease and quantitative traits.

Authors:

E. Oerton[1], Q. Wang[2], A. Nag[1], K. Smith[1], K. Carss[1], S. Petrovski[1]; [1]Ctr. for Genomics Res., Discovery Sci., R&D, AstraZeneca, Cambridge, United Kingdom, [2]Ctr. for Genomics Res., Discovery Sci., R&D, AstraZeneca, Waltham, MA

Due to physiological or environmental differences between females and males ('female' and 'male' here defined by sex karyotype), genetic associations may appear to be stronger in one sex than the other, or may even be absent in one sex entirely. These sex differences can reveal important information about the gene-phenotype relationship that may be obscured by standard sex-combined analyses. In this work, we apply our rare-variant collapsing PheWAS pipeline to 419,000 European-ancestry UK Biobank exomes, analysing females and males separately in order to identify sex-biased autosomal gene-phenotype associations. As well as 4,852 binary (case-control) and 1,264 quantitative traits, we apply this pipeline to Nightingale metabolite and Olink protein measurements, in the first biobank-scale study of sex-biased metabolomic and proteomic quantitative trait loci (QTLs). Females and males for each trait (including both cases and controls for binary traits) were downsampled to ensure equivalent power in both sexes. Defining 'sex-biased' as being genome-wide significant ($p<10^{-8}$) in one sex, with a non-significant ($p>=0.05$) p-value in the other and/or a significant difference in effect size as assessed by the Breslow-Day (binary traits) or Cochran's Q (quantitative traits) tests, we find 106 binary sex-biased traits. These include an association between protein-truncating variants in *DSP* and dilated cardiomyopathy which is highly significant in females with an odds ratio of 32.4 ($p = 3.92x10^{-12}$), but non-significant in males (OR 3.4, $p = 0.261$). We also find 1,035 quantitative sex-biased traits across 461 genes, including 385 mQTLs and 438 pQTLs. 55 of the sex-biased associations are genome-wide significant only in the sex-stratified analysis ($p<10^{-8}$ in stratified vs $p>=10^{-6}$ in combined), despite having only half the sample size of the sex-combined analysis. This illustrates the value of conducting sex-stratified analysis, unveiling biological insights that may remain concealed when considering males and females as a combined group. These sex-biased genetic associations shed light on potential sex-specific pleiotropy and enhance our understanding of sex-biased disparities in disease biology; further, they aid in identifying individuals most likely to benefit from therapeutic interventions targeting related gene targets.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1780 Shared genetic architecture and pleiotropy across uterine fibroids and hypertension.

Authors:

A. Pigg Akerele[1,2,3], J. N. Hellwege[2], J. A. Piekos[3], N. K. Khankari[2], T. L. Edwards[2], D. R. Velez Edwards[2]; [1]Meharry Med. Coll., Nashville, TN, [2]Vanderbilt Univ. Med. Ctr., Nashville, TN, [3]Vanderbilt Univ., Nashville, TN

Uterine fibroids (leiomyomata, UF) are among the most common gynecologic diseases in females of reproductive age, having an estimated cumulative prevalence of 70%. Women with fibroids experience debilitating symptoms (like severe pain and infertility) and a host of comorbidities (like polycystic ovarian syndrome, endometriosis, anemia, and hypertension (HTN)). Complications stemming from UF are a major indication for pelvic surgery and causes an economic burden of tens of billions of dollars per year. There is a need to better understand the genetic liability of risk for this phenotypically heterogeneous and complex disorder.

Multiple genetic loci have been identified to be associated with disease susceptibility for UF and shared risk factors, like HTN. In efforts to understand causal relationships between UF and HTN, we conducted a Mendelian Randomization analysis and evaluated the genetic correlations across blood pressure (BP) trait loci and UF. We used data from a cross-ancestry genome-wide association study (GWAS) meta-analysis of UF including published and unpublished data (44,205 cases and 356,552 controls), as well as data from a cross-ancestry GWAS meta-analysis of BP phenotypes (including diastolic BP [DBP], systolic BP [SBP], and pulse pressure [PP], N=447,758). We also evaluated genetic heritability and correlation of BP phenotypes and UF with linkage disequilibrium score regression (LDSC). Bi-Directional Two-Sample Mendelian Randomization (MR) was conducted to assess potential causal relationships between BP traits and UF. Genetic instruments for the MR analysis were selected from summary-level data of BP traits and UF by linkage disequilibrium clumping of genome-wide significant SNPs ($p<5e-8$) with an r2 threshold of 0.1.

LDSC results indicated a positive genetic correlation between DBP and UF (0.140, $p = 0.0004$), and SBP and UF (0.076, $p = 0.016$), and PP and UF (0.008, $p > 0.05$). MR using BP traits as exposures and UF as the outcome showed that DBP and pulse pressure both increase risk for UF ($b = 0.024$, $p = 0.002$ and $b = -0.010$, $p = 0.0008$, respectively). Having UF as the exposure and BP traits as the outcomes indicated a relationship between UF and DBP and UF and PP ($b = 0.46$, $p = 0.005$ and $b = -0.51$, $p = 0.015$, respectively). SBP as the outcome and as the exposure did not provide significant results. Our results provide evidence that HTN has shared genetic effects with UF that may contribute to the severity of the disease. Future directions of this study are to expand LDSC and MR analysis to other reproductive conditions. Through this work we will elucidate the shared genetic architecture and pleiotropy among women's gynecologic disorders and hypertension.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1781 Shared genetic architecture between congenital kidney defects and autism spectrum disorders.

Authors:

S. Krishna Murthy[1], H. Rasouly[1], S. Bheda[1], M. Marasa[1], N. Vena[1], E. Fiaccadori[2], A. Materna-Kiryluk[3], G. Masnata[4], V. Tasic[5], M. Saraga[6], K. Kiryluk[1], G. Ghiggeri[7], S. Sanna-Cherchi[1], A. Gharavi[1]; [1]Columbia Univ., New York City, NY, [2]Univ. di Parma, Parma, Italy, [3]Uniwersytet Medyczny w Poznaniu, Poznan, Poland, [4]Univ. of Cagliari, Cagliari, Italy, [5]Univ. Children's Hosp., Skopje, Macedonia, The Former Yugoslav Republic of, [6]Univ. Hosp. of Split, Split, Croatia, [7]Giannina Gaslini Inst., Genova, Italy

Background: Shared pathogenesis of congenital anomalies of the kidney and urinary tract (CAKUT) and autism spectrum disorder (ASD) has been suggested. Copy-number variants associated with ASD were reported to be enriched in CAKUT patients. Unlike ASD, CAKUT is often identified prenatally. Identification of syndromes including both ASD and CAKUT has the potential to improve prenatal diagnosis of syndromes associated with ASD.**Methods**: A total of 281 trios with CAKUT and no family history of disease were analyzed for de-novo variants enrichment using denovolyzer. In addition, 1,723 unrelated probands and 22,252 unrelated controls were matched based on genetic ancestry to perform gene-set burden analysis. Multiple gene-sets were pre-defined including i) a gene-set of 632 constrained genes (pLI score > 0.9, LOEUF>0.35) ii) highly expressed (top decile) in nephron-progenitor cells (NPCs) in 18wks human embryonic kidneys and iii) a gene-set of 547 constrained genes associated with ASD extracted from AutDB, a publicly available web-portal for on-going collection of genes linked to the disorder. Only rare loss-of-function variants (LoFs, gnomAD allele frequency $<10^{-5}$) were included in the analysis. Since the 2 cohorts were independent, we calculated the Fisher combination p-value for each gene-set enrichment analysis.**Results**: In the 281 trios, we observed 9 de-novo LoFs in the NPCs gene-set compared to the 1.2 expected (7.56-fold enrichment, p-value= $4.56x10^{-6}$). In the case-control analysis, we identified 78 probands with LoFs variants in the NPCs gene-set compared to 503 controls (1.81-fold enrichment, p-value= $3.53x10^{-5}$). The NPCs gene-set Fisher combination p-value reached $3.79x10^{-9}$. Using the ASD gene-set, we observed 10 de-novo LOFs compared to the 1.7 expected (5.74-fold enrichment, p-value= $1.76x10^{-5}$). The genes driving this enrichment included genes known to be associated with kidney anomalies and ASD (*PSMD12; KAT6B*), as well as genes known to be associated with ASD, but not with CAKUT (*KMT2A, TLK2, MAGEL2, ADNP, SIK1*). In the case-control analysis, we identified 88 probands with LoFs variants in the same gene-set compared to 660 controls (1.52-fold enrichment, p-value=$1.5x10^{-3}$). The Fisher combination p-value reached $5.38x10^{-7}$.**Conclusions**: A significant enrichment for rare LoFs in genes associated with ASD was observed in individuals with CAKUT. Future studies analyzing enrichment for LoFs in genes associated with CAKUT could further inform the scope of this overlap.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1782 Shared genetic effect of kidney function on bipolar disorder and major depressive disorder: a large-scale genome-wide cross-trait analysis

Authors:

S. Yu[1], B. Liao[1], X. Jin[1], Y. Lin[2]; [1]Dept. of Urology, Inst. of Urology (Lab. of Reconstructive Urology), Chengdu, China, [2]Hlth.Management Ctr., Gen. Practice Med. Ctr., and Med. Device Regulatory Res. and Evaluation Ctr., West China Hosp., Sichuan Univ., Chengdu, Sichuan, PR. China, Chengdu, China

Background:
Epidemiological studies have revealed a significant association between impaired kidney function and certain mental disorders, particularly bipolar disorder (BIP) and major depressive disorder (MDD). However, the evidence regarding shared genetics and causality is limited due to residual confounding and reverse causation.
Methods:
In this study, we conducted a large-scale genome-wide cross-trait association study to investigate the genetic overlap between 5 kidney function biomarkers (eGFRcrea, eGFRcys, blood urea nitrogen (BUN), serum urate, and UACR) and 2 mental disorders (MDD, BIP). Summary-level data of European ancestry were extracted from UK Biobank, Chronic Kidney Disease Genetics Consortium, and Psychiatric Genomics Consortium.
Results:
Using LD score regression, we found moderate but significant genetic correlations between kidney function biomarker traits on BIP and MDD. Cross-trait meta-analysis identified 1 to 19 independent significant loci that were found shared among 10 pairs of 5 kidney function biomarkers traits and 2 mental disorders. Among them, 3 novel genes: *SUFU*, *IBSP*, and *PTPRJ*, were also identified in transcriptome-wide association study analysis (TWAS), most of which were observed in the nervous and digestive systems (FDR < 0.05). Pathway analysis showed the immune system could play a role between kidney function biomarkers and mental disorders. Bidirectional mendelian randomization analysis suggested a potential causal relationship of kidney function biomarkers on BIP and MDD.
Conclusions:
In conclusion, the study demonstrated that both BIP and MDD shared genetic architecture with kidney function biomarkers, providing new insights into their genetic architectures and suggesting that larger GWASs are warranted.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1783 Shared genetic signal discovery in multi-trait analysis of lung and liver fibrosis**.**

Authors:

G. Parcesepe[1,2], E. Joof[1,2], R. J. Allen[1,2], J. K. Quint[3], I. D. Stewart[3], G. Aithal[4,5], R. G. Jenkins[3], L. V. Wain[1,2], DEMISTIFI consortium; [1]Univ. of Leicester, Leicester, United Kingdom, [2]NIHR Leicester BioMed. Res. Ctr., Leicester, United Kingdom, [3]Imperial Coll. London, London, United Kingdom, [4]Univ. of Nottingham, Nottingham, United Kingdom, [5]NIHR Nottingham BioMed. Res. Ctr., Nottingham, United Kingdom

Fibrosis is the pathological scarring of organ tissues, and is estimated to account for 45% of deaths in the developed world. There are no drugs that can halt, let alone reverse fibrosis. Our goal was to identify genes that could represent new targets for the development of anti-fibrotic therapy and that would benefit patients with one or more fibrotic disease. Taking lung and liver fibrosis as exemplars, we applied genetic association methods that combine diseases with shared underlying fibrotic pathology to improve power to identify shared risk variants and genes. Organ-specific fibrosis Genome-wide Association Studies (GWAS) were conducted in UK Biobank utilising a recently published consensus list of fibrotic diseases of the lung and liver. LD Score Regression was used to calculate genome-wide correlation and multi-trait analysis of GWAS was performed (MTAG) with $P<5x10^{-8}$ threshold for significance. MTAG exploits correlation between traits, to improve effect estimates, and gives trait-specific association statistics. The genome-wide genetic correlation between lung and liver fibrosis was 0.470 (P=0.047). When combining the GWAS using MTAG, there was increased significance of previously reported signals across the genome for both traits, including the signal at *IVD* previously associated with idiopathic pulmonary fibrosis (IPF) ($P=3.31x10^{-5}$ to $P=8.63x10^{-6}$), and the *SERPINA1* locus (alpha-1 anti-trypsin deficiency and non-alcoholic fatty liver disease (NAFLD))) ($P=1.29x10^{-6}$ to P=6.29x10-^{7}). Conversely, there was a decrease in the significance of signals at other known loci, including the *MUC5B* locus (IPF)) ($P=6.33x10^{-48}$ to $P=5.25x10^{-35}$) and the *PNPLA3* locus (NAFLD, alcoholic liver disease, and chronic hepatitis C)) ($P=5.18x10^{-42}$ to $P=1.70x10^{-35}$). An increase in association significance for signals in either trait suggests that these may represent shared genetic risk factors for lung and liver fibrosis. Signals that were attenuated when traits were combined may represent organ-specific signals. This could provide valuable insight into the shared and distinct fibrotic mechanisms in the liver and lung that can inform development of new anti-fibrotic therapies.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1784 Shared polygenic risk and causal inference between kidney function and inflammatory biomarkers: a large-scale genome-wide cross-trait analysis

Authors:

Y. Lin[1], H. Jin[2], G. Liao[3], L. Du[4], J. Li[5], L. Liang[1]; [1]Harvard Sch. of Publ. Hlth., Boston, MA, [2]West China Hosp., Sichuan Univerisity, Chengdu, China, [3]State Key Lab. of Oral Diseases, Natl. Clinical Res. Ctr. for Oral Diseases, West China Hosp. of Stomatology, Sichuan Univ., Chengdu, China, [4]West China Hosp., Sichuan Univerisity, CHengdu, China, [5]Brigham & Women s Hosp., Harvard Med. Sch., Boston, MA

With an increasing risk of mortality and morbidity, chronic kidney disease was expected to become the fifth leading cause of death globally. Various studies indicated that kidney dysfunction was prone to have a high risk of inflammation in the human body, which presented as disorders of diverse biomarkers containing, CRP, IL-6, leukocytes and its subtypes. However, most findings were limited due to small sample sizes, lack of biomarker categories and environmental confounding. We aimed to evaluate the shared genetic architecture between kidney function and systemic chronic inflammation. We recruited 11 kidney function biomarkers from UK Biobank and CKDgen(N=~400,000 to ~1,200,000) and 137 chronic inflammatory biomarkers from UK Biobank, deCODE, INTERVAL, KORA and other publicly released population databases(N = 30,000 to 400,000). Based on the large-scale genome-wide analyses(GWAS), we performed linkage-disequilibrium score regression to detect a genetic correlation between these 2 kinds of traits. Using Causal Analysis Using Summary Effect Estimates(CAUSE), we examined the bi-directional causal inferences between these traits. Traditional two-sample Mendelian randomization approaches, including IVW, MR-Egger and MBE, were additionally applied as sensitivity analyses. A total of 72 significantly genetically associated pairs between kidney function and inflammation biomarkers were identified. Cystatin C showed significant associations with 27 inflammatory factors, while eGFRcys showed significant associations with 45 inflammatory factors (FDR<0.01). Among them, the most significant associations were observed with Leptin (Cystatin C: Rg = 0.35, FDR = 6.53E-13; eGFRcys: Rg = -0.38, FDR = 7.06E19). Serum creatinine showed significant associations with 8 inflammatory biomarkers, while eGFRscr showed significant associations with 12 inflammatory biomarkers. The most significant association with serum creatinine was TNFaR2 (Rg = 0.38, FDR = 1.91E-13), while the most significant association with eGFRscr was TNFaR1 (Rg = -0.498,FDR = 7.58E-21). Inflammatory biomarkers,namely VCAM1, IL2Rb, Leptin, and TNFaR1 had a causal effect on Cystatin C, while TNFaR1, IL18Bpa, MPIF1, and IL2Rb had a causal effect on serum creatinine (FDR <0.05). In summary, we identified shared genetic architecture associated with kidney function and chronic inflammation in large population-based studies. We further characterized the shared polygenic risk of these biomarkers to provide insights into the molecular mechanism of the pathogenesis of impaired kidney function or to immune and inflammatory systems.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1785 Shared sex-specific functional genetic risk factors in self-reported clinical depression and Alzheimer's disease.

Authors:

T-C. Wang[1], A. C. Naj[2], W. Bush[3], L. Dumitrescu[4], P. Below[1]; [1]Vanderbilt Univ., Nashville, TN, [2]Univ. of Pennsylvania, Philadelphia, PA, [3]Case Western Reserve Univ, Cleveland, OH, [4]Vanderbilt Univ. Med. Ctr., Nashville, TN

In Alzheimer's disease (AD), roughly two-thirds of affected individuals are females, compared to males. Similar sexually dimorphic disease prevalence has been observed in individuals with self-reported clinical depression. Studies have suggested that individuals with depression have a higher risk of developing AD. In this study, we investigated the overlapping sex-specific functional genetic architecture of depression and AD to reveal shared biological mechanisms. We performed tissue-specific transcriptome-wide association analyses (SPrediXcan) on GWAS summary data for AD and sex-stratified depression. These analyses were to determine if there was an overlap in genes associated with AD and depression while investigating whether genetically regulated expression (GReX) of these genes exhibits a consistent direction of effect on both phenotypes. The AD GWAS included 111,326 clinically diagnosed/proxy AD cases and 677,664 controls analyzed by Bellenguez et al. The sex-stratified depression GWAS summary data were based on UK Biobank participants, including 8,166 male cases and 43,675 male controls in addition to 16,921 female cases and 49,020 female controls. The SPrediXcan analysis included eQTLs from 19 GTEx tissues, consisting of the colon sigmoid, liver, whole blood, adrenal gland, colon transverse, pituitary, and 13 brain-associated tissues. The AD SPrediXcan analysis yielded 3,128 significant tissue-specific associations (false discovery rate [FDR] < 0.05). The analysis of the male depression SPrediXcan model resulted in 24 such associations surpassing the significant threshold (FDR < 0.05), whereas no significant gene associations were observed in females. GReX of three genes, *TMEM106B*, *PPP1R18*, and *ZSCAN9*, was associated with both AD risk and male depression. Interestingly, the association of predicted expression of *TMEM106B* in whole blood was observed in both datasets but with opposite directions of effect (*TMEM106B* in AD: effect size=-0.057, P_{FDR}=0.013; in male depression: effect size=0.019, P_{FDR}=0.043). *TMEM106*B encodes a protein that regulates lysosomal function, and it has been implicated in AD and depression studies. Emerging evidence suggests that *TMEM106B* plays a crucial role in regulating microglial proliferation and survival in response to demyelination. Overall, our results suggest that there are genes that contribute to both depression and AD in a sex-specific manner. We plan to follow up with our current results to identify the causality of these associations between tissue-specific gene expressions and phenotypes, and in sex-stratified AD summary statistics.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1786 Significance of diversity and inclusion in the genomic study of speech and language disorder

Authors:

A. Hashim[1], M. H. Raza[2]; [1]Univ. of Kansas, Lawrence, KS, [2]The Univ. of Kansas, Lawrence, KS

Speech and language-related disorders like stuttering, speech sound disorder (SSD), and specific language impairment (SLI) affect a human's ability to communicate. The prevalence of these disorders is 5% to 10% among 2-5 year old children in the United States. Twins' studies demonstrate the highest heritability estimates of these communication disorders to 0.90. However, the genetic basis of these disorders is not well understood. Family studies identified several gene loci, and through next-generation sequencing, several candidate genes have been proposed. Population-based genetic studies also suggested the genetic and epigenetic basis of these disorders. The majority of the data collected is from individuals of European descent and the precision medicine, that is produced by the data, cannot be applied to individuals of different backgrounds due to limited diversity and inclusion in the genetic studies. The diversity in human genetic research is expanding to various research fields, but some areas like speech-language impairment still lack the inclusion of individuals from different origins and ethnicities. The diversity and inclusion in these areas are essential to the utility of clinical interventions in precision medicine. My research goals are to enhance the understanding of epigenetic and its effects on genomic research of speech-language disorders in diverse populations. Precision medicine can be improved with the addition of data from different population groups. I also plan to investigate how the quality of life of individuals with speech and language-related disorders can be improved by studying people of different descents.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1787 Single-cell analysis reveals a population of Th17 tissue-resident memory T cells in the small bowel of Crohn's disease.

Authors:

Y. Lee[1], J. Baek[1], S. Park[1], Y. Kim[1], S. Hwang[2], J. Lee[2], S. Park[3], S-K. Yang[3], B. Han[4], K. Song[1], Y. Yoon[2], B. Ye[3], H-S. Lee[1]; [1]Dept. of Biochemistry and Molecular Biology, Univ. of Ulsan Coll. of Med., Seoul, Korea, Republic of, [2]Div. of Colon and Rectal Surgery, Dept. of Surgery, Inflammatory Bowel Disease Ctr., Asan Med. Ctr., Univ. of Ulsan Coll. of Med., Seoul, Korea, Republic of, [3]Dept. of Gastroenterology, Inflammatory Bowel Disease Ctr., Asan Med. Ctr., Univ. of Ulsan Coll. of Med., Seoul, Korea, Republic of, [4]Dept. of BioMed. Sci., BK21 Plus BioMed. Sci. Project, Seoul Natl. Univ. Coll. of Med., Seoul, Korea, Republic of

Crohn's disease (CD) is a chronic inflammatory disorder that affects all layers of the bowel wall, characterized by patchy inflammation with skip lesions. Tissue-resident memory T cells (Trms) have been implicated in CD, yet their specific phenotype and molecular features in the context of CD pathogenesis remain poorly understood. To address this, we conducted single-cell RNA sequencing on T cells isolated from small bowel tissues of patients with CD, aiming to elucidate the characteristics of CD-predominant Trms. Samples were collected from 7 patients with CD and 4 patients with ulcerative colitis (UC), with small bowel tissues from patients with UC serving as controls. We isolated the lamina propria (LP) and submucosa/muscularis propria-enriched fraction (SM/MP) from each patient's small bowel tissues to assess the distribution of T-cell subsets and transcriptional signatures across different locations. Subsequently, we sorted T cells from each sample and performed single-cell RNA sequencing. Differential expression analysis was conducted using the Seurat V4 package, and transcription factor module activities were calculated using the SCENIC package with the RcisTarget database. We identified a total of 58,123 T cells, comprising 24,987 LP T cells and 33,136 SM/MP T cells, across 22 samples. These T cells clustered into 16 distinct clusters, with 3 clusters identified as Trm clusters. The Trm cluster exhibited high expression levels of genes associated with tissue residency, including *ITGAE*, *ITGA1*, and *CD69*. The CD4+ Trm subset, characterized by a Th17 transcriptional profile with notable expression of *IL17A*, *IL22*, *CCR6*, and *CCL20*, was designated as Th17 Trm. Among all clusters, only Th17 Trm cluster showed a significantly higher proportion in patients with CD compared to patients with UC in both the LP and SM/MP regions. Th17 Trm cells demonstrated pronounced upregulation of master transcription factors *RORC* and *PRDM1*, as well as *IL23R*, encoding a receptor for IL-23. When comparing the transcriptional landscape of CD and UC Th17 Trm cells, we observed a significant increase in genes related to IFN-γ production/response in patients with CD compared to patients with UC. In conclusion, our findings provide compelling evidence for a consistent increase in the proportion of Th17 Trm cells across different layers of the small bowel in patients with CD. These CD-predominant Th17 Trm cells display transcriptional characteristics reminiscent of Th1 cells, suggesting their potential contribution to the pathogenesis of CD.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1788 Single-cell level analysis of obese individuals' two main fat depots discovers differences in cell-type marker genes and their genetic risks for obesity.

Authors:

M. Sukhatme[1], U. Arasu[2], A. Kar[1], S. T. Lee[1], M. Alvarez[1], I. Selvarajan[2], T. Örd[2], D. Kaminska[3,4], V. Männistö[5], J. Pihlajamäki[3,6], M. U. Kaikkonen[2], P. Pajukanta[1,7]; [1]Dept. of Human Genetics, David Geffen Sch. of Med. at UCLA, Los Angeles, CA, [2]A. I. Virtanen Inst. for Molecular Sci., Univ. of Eastern Finland, Kuopio, Finland, [3]Inst. of Publ. Hlth.and Clinical Nutrition, Univ. of Eastern Finland, Kuopio, Finland, [4]Dept. of Med., Div. of Cardiology, UCLA, Los Angeles, CA, [5]Inst. of Clinical Med., Univ. of Eastern Finland and Kuopio Univ. Hosp., Kuopio, Finland, [6]Dept. of Med., Endocrinology and Clinical Nutrition, Kuopio Univ. Hosp., Kuopio, Finland, [7]Inst. for Precision Hlth., David Geffen Sch. of Med. at UCLA, Los Angeles, CA

It is not known whether obesity influences cell-type level gene expression and its genetic regulation differently in the two key human fat depots, subcutaneous (SAT) and visceral adipose tissue (VAT), and how that relates to the main obesity comorbidity, type 2 diabetes (T2D). We hypothesized that there are differences in cell-type level VAT and SAT expression profiles and that regional variants of the VAT and SAT-specific cell-type marker genes may contribute differently to the genetic risk for obesity. To this end, we performed a dual-tissue single nucleus RNA-sequencing analysis from 7 obese individuals from the Finnish KOBS cohort. After uniform quality control using DIEM and DecontX and conducting clustering, integration, and cell-type assignments using Seurat, CCA, and SingleR, we identified shared and tissue-specific marker genes for each SAT and VAT cell type. Through a GWAS enrichment analysis with MAGENTA, we found that marker genes specific to SAT adipocytes and macrophages were enriched at a 95% enrichment level for BMI (p=0.002) and T2D (p=0.033), whereas the adipocyte marker genes shared between SAT and VAT were enriched for an abdominal obesity proxy, waist-hip-ratio adjusted for BMI (WHRadjBMI) (p=0.0095). No significant GWAS enrichments were identified using the VAT-specific cell-type marker genes. To further investigate the genetic obesity risk related to the adipocyte expression in SAT vs. VAT, we built regional polygenic risk scores (PRS) for WHRadjBMI in the UK Biobank cohort using PLINK for SNPs in the cis-regions of the shared and SAT/VAT-specific adipocyte marker genes. We observed that the regional PRS built from the local SNPs around the SAT-specific adipocyte genes explained 0.785% of the variance in WHRadjBMI (permuted p=0.017). Overall, this suggests that SNPs in the cis-regions of adipocyte markers specific to SAT adipocytes play a larger role in obesity than regional VAT-specific marker gene SNPs. Our follow-up pathway enrichment analysis with WebGestalt for the SAT adipocyte-specific marker genes identified 22 enriched pathways (FDR<5%) centered on fatty acid and triglyceride metabolism. The adipose master transcription factor SREBF1 is found in 19 of the 22 pathways, supporting the key role of SAT in lipogenesis and fatty acid metabolism. In conclusion, our dual tissue single-cell level omics results, along with GWAS enrichment and PRS analyses, suggest that there are differences in SAT and VAT adipocyte marker genes and the risks their regional variants cause for obesity. Specifically, we observed that adipocytes and macrophages in SAT were more implicated in traits associated with obesity than their counterparts in VAT.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1789 Single-cell multi-omic dissection of opioid use disorder reveals gene regulatory circuitry changes across multiple brain regions

Authors:

B. James[1,2], R. M. Linville[1,2], K. Galani[1,2], L-L. Ho[1,2], Z. Liu[1,2], N. Sun[1,2], S. Pineda[1,2], D. Mash[3], B. Ruzicka[4,5], B. Madras[5,4], D. H. Gabuzda[6,7], M. Heiman[1,2], M. Kellis[1,2]; [1]Massachusetts Inst. of Technology, Cambridge, MA, [2]Broad Inst. of MIT and Harvard, Cambridge, MA, [3]Dept. of Neurology, Miller Sch. of Med., Univ. of Miami, Miami, FL, [4]McLean Hosp., Belmont, MA, [5]Dept. of Psychiatry, Harvard Med. Sch., Boston, MA, [6]Dept. of Cancer Immunology and Virology, Dana-Farber Cancer Inst., Boston, MA, [7]Dept. of Neurology, Harvard Med. Sch., Boston, MA

Opioid use disorder (OUD) is a chronic relapsing disorder engaging addictive and emotional centers of the brain. The opioid crisis in the US is an epidemic, with over 100,000 deaths in 2022 alone, and over a million deaths in the last two decades. OUD is largely mediated by changes in the brain's reward system, including the mesocortiolimbic pathway that transmits dopamine from the ventral tegmental area (VTA) to both the the prefrontal cortex (PFC) and nucleus accumbens (NAc). While mechanisms of addiction have been extensively studied along this pathway in rodents, the role of specific cell types and brain regions in human OUD are not fully understood in humans. To address this, we profiled postmortem human PFC and NAc using single-cell multiomics from 26 OUD and 26 control individuals. With over a million snRNA cells and a million snATAC cells, most of which have paired barcodes, we identify cell type-specific epigenomic and transcriptomic profiles of the human NAc and PFC. We initially focused on expression patterns of opioid receptors. Genetic studies have identified the OPRM1 locus, the gene encoding the mu opioid receptor, as the most significant genetic risk locus for OUD. Notably, we identify a subtype of medium spiny neurons (MSNs) of the NAc that express highest levels of *OPRM1*, while in the PFC OPRM1 expression is highest in VIP+/CLSTN2+ inhibitory neurons. This MSN subtype was previously annotated in mouse brain by Otof expression, and by a cross-species comparison, we find conserved markers including Adarb2, Sema3a, and Foxp2. In human, this cell subtype is marked by thyrotropin releasing hormone degrading enzyme (TRHDE) expression, and we validate this finding by *in situ* hybridization. Next, we identify gene regulatory networks that are dysregulated across OUD, and use multi-omics to prioritize genomic loci. We also identify the peaks of these eccentric MSNs as highly enriched for OUD genetic variants. We identify large scale changes across neurons, glia, and vasculature, as well as epigenomic changes upstream of these. Overall, we develop a comprehensive map of OUD in the human brain and use this to map regulatory changes happening at several levels.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1790 † Single-cell RNA-seq analysis of microglia reveals functional heterogeneity in the substantia nigra of Parkinson's disease donors.

Authors:

R. Kubler[1,2], D. Mattei[1], K. BP[1], O. Narcis[1], M. Rosen[1], B. Jang[1], L. De Witte[2], T. Raj[1]; [1]Icahn Sch. of Med. at Mount Sinai, New York, NY, [2]Donders Inst. for Brain, Cognition and Behaviour, Njmegen, Netherlands

Background: Parkinson's disease (PD) is characterized by the presence of α-synuclein-containing Lewy bodies and the selective loss of dopaminergic neurons in the substantia nigra. Recent studies have identified numerous genetic risk variants associated with PD. Our previous work identified PD risk loci that alter gene expression in microglia and monocytes. However, previous studies predominantly captured oligodendrocytes when transcriptionally profiling the substantia nigra of PD patients. The functional role of PD risk variants in microglia of the substantia nigra, therefore, remains unclear. We sought to address this gap by using single-cell RNA-seq to characterize transcriptional microglia subtypes and their enrichment for PD variants in microglia isolated from fresh and frozen substantia nigra tissues of PD patients and controls. **Methods:** We sequenced 110,000 oligodendrocyte-depleted brain nuclei (37,000 *in-silico*-filtered microglia) from 16 donors and an additional 60,000 fresh microglia cells from 19 donors. **Results:** Our analysis revealed 19 distinct microglia identities with diverse transcriptional phenotypes, including phagocytosis, chemotaxis, neuroinflammatory, and disease-associated microglia (DAM) signatures. We found that expression of inflammatory and DAM genes was increased in PD microglia. *GPNMB*, a PD microglia expression QTL gene, was a marker for microglia cluster M2, an identity strongly enriched for DAM I genes. Cluster M5 was significantly enriched for Nalls et al. (2019) PD brain eQTL genes and DAM II markers. It further overexpressed *TMEM163* and *CTSB*, two microglia-specific eQTL and splicing QTL genes colocalized with PD risk loci. Chemotactic microglia cluster M9, enriched for DAM I and PD-DAM genes, was significantly overexpressing *TMEM163* and *P2RY12*, also a colocalized PD microglia eQTL gene. We further identified a highly heterogeneous cluster (M0) enriched for epigenetic and cell differentiation markers but not mapping to any previously published microglia phenotypes. Pseudotime analysis revealed a central cluster, dividing into distinct (anti-)inflammatory DAM clusters, highly enriched for M0 microglia. **Conclusions:** Our findings provide novel insights into the heterogeneity and dynamics of microglia functionality in the substantia nigra. In particular, DAM clusters expressed PD QTL genes as markers. We further indicated the presence of a novel, heterogeneous microglia cluster (M0) that might serve as a transitioning stage for microglial phenotypic differentiation. How these microglia states in the substantia nigra of PD patients contribute to disease progression remains to be elucidated.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1791 *SOX7*: Novel autistic gene identified by analysis of multi-omics data and human brain organoids.

Authors:

J. Zhao[1], S. Gonzales[2], N. Choi[3], P. Acharya[3], M-Y. Lee[3]; [1]Miami Dade Coll. Kendall Campus and Sch. for Advanced Studies, Miami, FL, [2]Florida Intl. Univ., Miami, FL, [3]Univ. of North Texas, Denton, TX

Background: Genome-wide association studies (GWAS) based on DNA information have identified thousands of mutations associated with autism spectrum disorder (ASD). However, more than 99% of identified mutations are non-coding. Thus, it is unclear which of these mutations might be functional and thus potentially causal variants. Transcriptomic profiling using total RNA-sequencing has been one of the most utilized approaches to link protein levels to genetic information at the molecular level. The transcriptome captures molecular genomic complexity that the DNA sequence solely does not. Some mutations alter a gene's DNA sequence but do not necessarily change expression and/or protein function. To date, few common variants reliably associated with the diagnosis status of ASD despite consistently high estimates of heritability. **Objectives:** It is necessary to integrate DNA and RNA testing together to identify true causal genes. **Methods:** We performed gene-based association studies with adaptive test using GWAS summary statistics with two large GWAS datasets (ASD 2019 data: 18,382 ASD cases and 27,969 controls [discovery data]; ASD 2017 data: 6,197 ASD cases and 7,377 controls [replication data]) which were obtained from the Psychiatric Genomics Consortium (PGC). In addition, we investigated differential expression for genes identified in gene-based GWAS with a RNA-seq dataset (GSE30573: 3 cases and 3 controls) using DESeq2 package. We further investigate function of identified genes with autistic brain organoids. **Results:** We identified 5 genes significantly associated with ASD in ASD 2019 data (*KIZ-AS1*, $p=8.67\times10^{-10}$; *KIZ*, $p=1.16\times10^{-9}$; *XRN2*, $p=7.73\times10^{-9}$; *SOX7*, $p=2.22\times10^{-7}$; *PINX1-DT*, $p=2.14\times10^{-6}$). Among these 5 genes, gene *SOX7* (p=0.00087), *LOC101929229* (p=0.009), and *KIZ-AS1* (p=0.059) were replicated in ASD 2017 data. *KIZ* (p=0.06) was close to the boundary of replication in ASD 2017 data. Genes *SOX7* (p=0.0017, adjusted p=0.0085), *LOC101929229* ($p=5.83\times10^{-7}$, adjusted $p=1.18\times10^{-5}$), and *KIZ* (p=0.00099, adjusted p=0.0055) indicated significant expression differences between cases and controls in the RNA-seq data. Human brain organoids support the gene *SOX7* associated with ASD. *SOX7* encodes a member of the *SOX* (SRY-related HMG-box) family of transcription factors pivotally contributing to determining of the cell fate and identity in many lineages. The encoded protein may act as a transcriptional regulator after forming a protein complex with other proteins leading to autism. **Conclusion:** Gene *SOX7* in the transcription factor family could be associated with ASD. This finding may provide new diagnostic and therapeutic strategies for ASD.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1792 Spatial omics of acute myocardial infarction reveals novel sites of immune cell infiltration.

Authors:

F. Wuennemann[1,2], F. Sicklinger[3,4], K. Bestak[1,2], J. Nimo[5], F. Coscia[5], M. Ibarra[1,2], M. Chazotte[1,2], T. Thiemann[4], S. Pohl[4], F. Leuschner[3,4], D. Schapiro[1,2]; [1]Inst. for Computational Biomedicine, Faculty of Med., Univ. Hosp. Heidelberg and Heidelberg Univ., Heidelberg, Germany, [2]Inst. of Pathology, Univ. Hosp. Heidelberg, Heidelberg, Germany, [3]Dept. of Internal Med. III, Univ. of Heidelberg, Heidelberg, Germany, [4]German Ctr. for Cardiovascular Res. (DZHK), Partner Site Heidelberg, Heidelberg, Germany, [5]Spatial Proteomics Group, Max-Delbrück-Centrum for Molecular Med. in the Helmholtz Association (MDC), Berlin, Germany

Myocardial infarction (MI), commonly known as a heart attack, continues to be the leading cause of death worldwide. Recent studies have highlighted the significance of tissue inflammation during the early acute phase of MI, which plays a vital role in the healing process following a heart attack. During the inflammatory phase of MI immune cells infiltrate the infarct region to remove dying and necrotic cells. Here we investigated the dynamics of immune cell infiltration during the first four days in a minimally invasive mouse model of acute myocardial infarction using combinatorial single-molecule FISH (smFISH) and sequential immunofluorescence (seqIF). We evaluated several approaches for cardiac cell segmentation and established computational pipelines to process and quantify both transcriptomic and antibody based imaging modalities for cardiac tissue. Our time series of immune infiltration during MI using spatially resolved methods revealed significant invasion of monocytes and neutrophils during the first 2 days via the endocardium, i.e. the inner lining of the heart. We characterized the proteome of the endocardial layer at 1 day post infarction using laser-microdissection coupled to high sensitivity proteomics and identified von-Willebrand-Factor (vWF) as an important player in immune cell recruitment. Our results reveal the endocardium as a significant and important novel target for immune cell infiltration during the inflammatory phase of MI.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1793 Statistical physics of fitness landscape discovers risk genes and pathways specific to male and female with Alzheimer's Disease.

Authors:

K. Lee, M. Samieinasab, F. Nazem, I. Al-Ramahi, Y. Lagisetty, J. Botas Miret, C. Wang, K. Wilhelm, P. Katsonis, J. Asmussen, J. Botas, O. Lichtarge; Baylor Coll. of Med., Houston, TX

Identifying genes linked to complex diseases is a challenge. Association studies demand large cohorts, and multimodal omics integration is context sensitive. This study tests a third approach in Alzheimer's Disease, exploiting freely available but massive evolutionary and phylogenetic coupling data on sequence variation and speciation. In the context of fitness landscapes, these couplings lead to computable values for the influence of genes in a population. We may then compare a gene's influence in sequenced cases vs controls cohorts and test the hypothesis that significant deviations identify genes linked to disease risk.

In 4768 AD cases and 4689 controls, from NIAGADS, 122 genes had significantly greater influence in cases or in controls. These genes overlapped ($p = 3.10^{-5}$) or interacted ($z = 4.85$) with AD GWAS genes. They interacted mutually ($p = 0.0033$) and with AD-related processes ($n = 58$, $p = 7.10^{-13}$). In a live *Drosophila* tauopathy model, they modulated neurodegeneration: the knockdown or overexpression of 64 genes ameliorated or worsened age-dependent neuronal dysfunction ($p < 0.05$). These data support the association of these genes to AD.

Next, robustness to down-sampling suggested we could analyze smaller, sex-separated cohorts. We found 82 genes in males (2010 cases, 1938 controls) and 69 genes in females (2758 cases, 2751 controls) overlapping by 15 ($p < 10^{-53}$). Many male (36) and female (34) genes were new and significantly connected to known AD processes ($p < 10^{-7}$). The 136 genes, combined, overlapped ($p < 0.0001$) and interacted ($z > 3.5$) with AD genes.

Critically, all these genes predict AD risk after training when tested on never seen before held-out subjects. Male-specific genes trained in males had AUCs of 0.74 (0.67 without *APOE*) in males, and 0.65 (0.59) in females. Female-specific genes trained in females had AUCs of 0.74 (0.65) in males and 0.68 (0.61) in females. Sex-neutral genes obtained from and trained on both males and females reached peak AD risk AUCs of 0.80 (0.75) in males, 0.71 (0.67) in females.

These data support the identification of new AD-associated genes informing relevant molecular pathways and improving risk evaluation from germline variants. This performance required fewer than 5000 cases and 5000 controls, and it was consistently better in males where risk prediction from germline variants compared favorably to the state-of-the-art. These results are consistent with key genetic and environmental differences leading to AD in males and females, and support usage of evolutionary and phylogenetic information to probe complex human diseases.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1794 Structural variation in Keratoconus

Authors:

M. Rydzanicz[1], A. Wysocka[2], K. Jaskiewicz[2], M. Maleszka-Kurpiel[3,4], M. Kabza[5], M. Gajecka[5,2]; [1]Dept. of Med. Genetics, Med. Univ. of Warsaw, Warsaw, Poland, [2]Inst. of Human Genetics, Polish Academy of Sci., Poznan, Poland, [3]Chair of Ophthalmology and Optometry, Poznan Univ. of Med. Sci., Poznan, Poland, [4]Optegra Eye Hlth.Care Clinic, Poznan, Poland, [5]Chair and Dept. of Genetics and Pharmaceutical Microbiol., Poznan Univ. of Med. Sci., Poznan, Poland

Background: Keratoconus (KTCN) is a degenerative disorder of the eye characterized by progressive stromal thinning, that results in the conical shape of the cornea. We limited our research to corneal epithelium (CE) samples, as characteristic clinical changes in KTCN occur in this layer. We defined 3 topographic regions (TRs, *central*, *middle*, *peripheral*) based on the epithelial thickness and corneal topography maps. The preliminary studies were undertaken on a group of KTCN patients and non-KTCN individuals. WGS was performed to assess KTCN-specific structural variants (SVs). The aim of this project was to characterize KTCN-specific SVs affecting gene function, to unravel their role in KTCN pathogenesis.
Material and methods: CE samples were collected during a standard procedure of crosslinking in KTCN patients [12 experimental CE samples (4 unrelated patients x 3 topographic regions)] and in controls undergoing a photorefractive keratectomy to correct vision errors [6 control samples (from 2 individuals with mild myopia phenotype x 3 TRs)]. The extraction of DNA, RNA, and protein was performed. Quality and quantity of samples were verified before WGS, performed with the TruSeq Nano DNA HT LibraryPrep Kit (Illumina), HiSeqX platform (Illumina), and 30X coverage. The reads after bioinformatic processing were mapped to reference genome GRCh38. Structural variants calling has been performed using Manta and cnMOPS, and tools' outputs were combined into a single consensus. SVs which had at least 90% mutual overlapping with any region detected by at least one bioinformatic tool were used in further analysis.
Results: Implemented pipeline enabled the recognition of long deletions localized in non-coding regions of the genome and in gene introns. We removed the SVs present in non-KTCN samples and obtained 779 variants for KTCN patients. After a literature search, GeneCard, and gnomAD analysis, as well as BAM files verification, we received the definitive number of 24 SVs. Assessing the previously published KTCN-specific *loci*, we did not observe differences in the number and the length of deletions comparing KTCN patients and control individuals. The number of recognized deletions per person (in both examined groups) ranges from 4159 to 4381.
Conclusions: The identified genomic features further confirm the heterogeneous genetic background of KTCN. The preliminary results should be verified using long-read sequencing.
Support: The National Science Centre grant no. 2018/31/B/NZ5/03280.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1795 Subtyping complex traits using tissue-specific polygenic risk scores.

Authors:

A. Kim, S. Grimes, S. Chhetri, R. Keener, M. Arvanitis, A. Battle; Johns Hopkins Univ., Baltimore, MD

Disease subtyping can accurately categorize clinical patients into disease subgroups and address the heterogeneity of disease phenotypes. Complex traits can be highly polygenic and often have a pathophysiology that involves multiple tissues; variants affecting each of these tissue types can contribute to the risk. Although integrative methods for subtyping based on genetics and electronic health records exist, current approaches do not incorporate the tissue-specific genetic architecture of complex traits. Here, we leveraged functional genomic annotations and tissue or cell type-specific annotations to build 25 tissue-specific polygenic risk score (TS-PRS) models of body mass index (BMI) for 336,335 individuals of white British ancestry in UK Biobank. Using the reference catalog of human epigenomes from the Roadmap Epigenomics Consortium to identify enhancer regions that are unique to a tissue or cell type, we identified tissue-specific variants from BMI genome-wide association study from the GIANT consortium and estimated PRS using LDpred-funct. Among the 25 TS-PRS models, the pluripotent stem cell and brain PRS models explained the most phenotypic variance (R^2=0.046 and R^2=0.043, respectively), supporting the known substantial brain component of BMI. To further evaluate our framework, we generated TS-PRS models for a related phenotype, waist-to-hip ratio adjusted for BMI, which is thought to manifest through mesenchymally derived tissues. Indeed, the mesenchymal stem cells PRS model explained the most phenotypic variance (R^2=0.031), demonstrating that our method identifies appropriate tissues for complex traits. To identify the molecular mechanisms that are correlated with a subset of tissues a complex trait is primarily manifesting through, we calculated the Pearson correlation coefficients between BMI TS-PRS and liver enzyme levels, in addition to anthropometric and lipid measurements. Our pluripotent stem cells and brain BMI TS-PRS models were statistically significantly associated with liver enzyme levels. The association between brain BMI TS-PRS and liver enzyme suggests that tissue subtyping may be able to provide mechanistic insight for individuals by which obesity plays a role in the risk of diseases associated with elevated liver enzyme levels. Our TS-PRS method yields mechanistic insights into complex traits and may help identify subgroups of individuals whose genetic susceptibility to the trait is mediated in a tissue-specific manner. We expect this method to scale across diverse complex traits and improve our understanding of their underlying tissue-specific biology to better diagnose and treat affected individuals.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1796 Survival GWAS with inverse probability weighing to estimate population-unbiased polygenic risk scores of psychiatric disorders using the iPSYCH case-cohort of Danish nation-wide register data.

Authors:

L. Huang[1], D. Helenius[1], V. Appadurai[1], W. Thompson[2], T. Werge[1]; [1]Inst. of Biological Psychiatry, Roskilde, Denmark, [2]Univ. of California, San Diego, La Jolla, CA

[Aim] We aim to estimate population-unbiased survival polygenic risk scores (PRS) to predict the absolute risk of psychiatric disorders such as schizophrenia, bipolar disorder, autism and ADHD, which addresses both selection bias and age-of-onset, that can be reliably implemented into clinical practice. **[Background]** Case-control GWAS is known to suffer from selection bias which may inflate the risk estimates. In addition, case-control GWAS does not account for age-of-onset, which is a hallmark for psychiatric disorders. The PRS constructed from case-control GWAS may give rise to biased risk estimates that could lead to invalid or premature clinical interventions if implemented. **[Study design and data]** In our study, we used the iPSYCH case-cohort that is built from Danish nation-wide psychiatric register data with full genotyping profile from Danish Neonatal Screening Biobank. iPSYCH2012 case-cohort includes individuals born in Denmark between 1981 and 2005, diagnosed with one or more of the investigated disorders or selected as a random control. iPSYCH2015 augmented this cohort by increasing the birth year up until 2008 with 4 additional years of follow-up. **[Method]** We conduct survival GWAS on 4 psychiatric disorders using Cox proportional hazard model to address age-of-onset, and Barlow's inverse probably weighing to adjust for strong case-ascertainment by upweighing controls, to provide estimates comparable to a true cohort study. The adjusted robust variance is calculated by a jack-knife estimator to account for increased variance introduced by weighing, which provides more precise estimates. **[Results]** Analysis is ongoing. We have computed the weighted hazard ratio estimates from survival GWAS on the iPSYCH2012 cohort, which will then be used to compute PRS in the iPSYCH2015 cohort to estimate the disease risk. **[Discussion]** We propose Barlow's weighted survival PRS as a robust alternative to case-control PRS in predicting absolute risk of disease.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1797 Systolic blood pressure, preeclampsia, and leiomyoma genes - a multivariable Mendelian randomization and mediation analysis

Authors:

N. Vaitinadin[1], M. Shi[2], J. Mosley[3], A. Giri[3]; [1]Vanderbilt Univ Med. Ctr., Nashville, TN, [2]VUMC, Nashville, TN, [3]Vanderbilt Univ. Med. Ctr., Nashville, TN

Introduction Preeclampsia (PE) (sudden-onset hypertension (>20 weeks gestation) with proteinuria, maternal organ/uteroplacental dysfunction or angiogenic imbalance) affects ~4 million women annually, causing death of >70,000 women and 500,000 babies, and long-term sequelae like heart failure. Epidemiologic literature suggests rise in diastolic blood pressure (DBP) is more predictive for PE than systolic BP (SBP). It is not clear which of the known genetic associations of PE is an independent risk factor. **Methods** Publicly available European ancestry genome-wide association study (GWAS) summary statistics (SS) were used to develop single nucleotide polymorphism (SNP)-based genetic instruments (P-value$\leq 5x10^{-08}$, $r^2<0.001$) for known associations of PE - SBP (mm Hg) and DBP (mm Hg) (458,577 subjects), body mass index (BMI, kg/m^2) (339,224), and high density lipoprotein (HDL, mg/dl) (188,577), along with SS for PE or eclampsia (7212 cases, 194266 controls), the outcome. We performed 1) univariate two-sample Mendelian randomization (MR), using inverse variance weighted method (IVWM) for results; MR-Egger, and weighted-median methods for sensitivity analyses, 2) multivariable (MV)MR to determine independent risk factors after multivariable adjustment, 3) analysis to quantify proportional change in effect size after adjustment, and 4) FUMA based analysis of independent risk factor SNPs to annotate uterus associated genes using GTEx v8. MR results are reported as odds of PE per ten-unit change in SBP/DBP/HDL, and per standard deviation (SD) change in BMI. **Results** In 1) IVWM analysis, SBP (OR= 2.23 [95% CI= 1.82, 2.72]), DBP (3.32 [2.46,4.48], HDL (0.22 [0.09,0.55]), and BMI (1.58 [1.34,1.86]) associated with the odds of a diagnosis of PE; 2) MVMR analysis, only SBP (1.82 [1.11, 3.00] remained significant, after multivariate adjustment for BMI, DBP, and HDL; 3) proportion-explained analysis, ~18% reduction in the OR of SBP, after adjustment for BMI, DBP, and HDL; and 4) in FUMA based analyses, out of 7 gene associations, the top 2 uterus associated genes for SBP SNPs were *ZNF859*, and *CDC25A*. **Conclusions** Genetic predisposition to elevated SBP, rather than DBP, associated with increased odds of PE, independent of BMI, DBP, and HDL; adjustment attenuated risk, suggesting shared genetic mechanisms between SBP, DBP, BMI, and HDL. SBP is an independent risk factor for PE. SBP SNPs map to uterine *ZNF859* and *CDC25A*, known leiomyoma associated genes, suggesting a pleiotropic role in PE development.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1798 Temporal landscape of the Alzheimer's Disease plasma proteome: an 1800-sample cohort study

Authors:

A. Siddiqui[1], H. Guturu[1], A. Alavi[1], B. Lacar[1], G. Venkataraman[1], D. Hornburg[1], M. de Geus[2], P. Kivisakk[2], S. Batzoglou[1], S. E. Arnold[2]; [1]Seer, Redwood City, CA, [2]Massachusetts Gen. Hosp., Boston, MA

Introduction

The pathophysiological mechanisms that drive the development and progression of Alzheimer's Disease (AD) are still not fully understood, and ongoing research continues to identify biomarkers that can be used to diagnose the disease earlier, predict progression, and help in the development of more effective treatments. Here, we leveraged recent advances in plasma proteomics to conduct an unprecedented deep and unbiased study of the temporal landscape of the AD plasma proteome.

Methods

Our cohort is comprised of subjects with three different conditions: normal cognition, AD, and other neurodegenerative diseases, which included 1,802 samples from > 1,000 unique individuals from whom between 1 and 6 blood draws were analyzed over the course of 12 years. The plasma samples were randomized based on conditions, where all longitudinal samples from the same individual were analyzed on the same plate. The samples were processed with the Proteograph™ XT workflow and then analyzed using a high-throughput Liquid Chromatography (LC) coupled to a Data Independent Acquisition (DIA) on a single Thermo Fisher Scientific Orbitrap™ Exploris™ 480 Mass Spectrometer (MS) in 11 weeks. Further, whole exome sequencing data from 149 subjects of this study were generated for proteogenomic analyses including protein variant identification, and array genotyping is planned for all individuals.

Results

LC-MS data were acquired on 1,802 samples. We assessed several DIA spectral library search strategies, including library-free analysis with match between runs and a project-specific gas-phase fractionation (GPF) library. We identified 5,253 protein groups using the library-free search and 4,007 using the GPF library (FDR < 0.01). Preliminary analysis showed a strong biological signal, demonstrating the statistical power of the data. Using 10-fold cross validation and a customized AutoML framework, we could separate controls from dementia-affected individuals (AUC = 0.81). Further, we identified both known AD-associated biomarkers (e.g., SPP1, CRP, F12, and C9) as well as novel biomarkers, supporting the potential for new insights into AD.

Conclusions

A comprehensive characterization of the temporal landscape of AD of this scale and depth using unbiased proteomics has not been previously published, and this work represents a new opportunity to extend our knowledge of AD and other neurodegenerative disease. Our preliminary analysis is promising, and we plan to extend this work to better understand the trajectory of AD-associated protein changes over time and the biological pathways that underlie the observed differences.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1799 Tensor decomposition of multi-dimensional splicing events across multiple tissues to identify splicing-mediated risk genes associated with complex traits

Authors:

Y. Yan[1], R. Chen[2,3], Y. Tan[2,3], A. Tiwari[2,3], X. Zhong[2,4], B. Li[2,3]; [1]Dept. of Biostatistics, Vanderbilt Univ., Nashville, TN, [2]Vanderbilt Genetics Inst., Nashville, TN, [3]Dept. of Molecular Physiology and Biophysics, Vanderbilt Univ., Nashville, TN, [4]Dept. of Med., Div. of Genetic Med., Vanderbilt Univ. Med. Ctr., Nashville, TN

Identifying risk genes associated with complex traits remains challenging. The integration of gene expression data with Genome-Wide Association Study (GWAS) through Transcriptome-Wide Association Study (TWAS) methods has discovered candidate risk genes for a variety of complex traits. Splicing, which explains a comparable heritability of complex traits as gene expression, has not been fully explored yet due to the challenge of multidimensionality of splicing events. It is even more challenging to model splicing events across tissues. To fully take the advantages of multiple splicing events in a gene and the shared splicing across tissues, we employed tensor decomposition in conjunction with sCCA (sparse Canonical Correlation Analysis) to extract meaningful information from high-dimensional multiple splicing events across multiple tissues. We then build gene-based splicing predictive models using GTEx data and apply the models to GWAS summary statistics of Alzheimer's disease (111,326 cases and 677,663 controls) to identify splicing-mediated risk genes associated with AD. In total, we identified 174 significant risk genes at Bonferroni correction. Gene Ontology analysis showed that the risk genes identified are enriched in AD-related functions, e.g., amyloid-beta-related pathways, endocytosis, and immunity functions. Compared to the models using single-tissue of the brain frontal cortex, our results demonstrated significant enrichment of AD-related pathways and identified additional AD risk genes that were not detected in the brain tissue analysis alone, while preserving most of the top genes identified in brain tissue. As AD genetics is mostly implicated in microglia that occupy only a small proportion of the cells in the brain, using transcriptomics data in brain tissues for AD may not capture the full spectrum of AD genetics. Our across-tissue modeling is able to extract the splicing information relevant to AD for risk gene discovery. Furthermore, our across-tissue splicing models can be applied to other complex traits to help identify splicing-mediated disease risk genes.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1800 Testing an autism spectrum disorder polygenic score in individuals with African-European admixed ancestries

Authors:

A. Flynn-Carroll[1], K. Hou[1], J. B. Williams[2], J. K. Lowe[1], L. M. Hernandez[3], M. E. Kumagai[1], B. Pasaniuc[1], D. Geschwind[4]; [1]UCLA, Los Angeles, CA, [2]Univ. at Buffalo, Buffalo, NY, [3]Univ. of California, Los Angeles, Los Angeles, CA, [4]Univ California Los Angeles, Santa Monica, CA

Autism spectrum disorder (ASD) is a common neurodevelopmental syndrome defined by moderate to severe disfunction in social skills, communication, and repetitive behavior. Heritability estimates for ASD range from 64-91%, indicating a strong genetic component. Polygenic scores (PGS) have previously been used to quantify an individual's risk of ASD, but most of these scores have been developed in populations of predominantly European ancestries, and their accuracy has not been well documented in non-European populations. Individuals with African ancestry are particularly underrepresented in genetic studies of neuropsychiatric disorders. Moreover, admixed individuals are routinely excluded from large scale genomic studies as there is a lack of developed methods to help account for population sub-structure. Here, we test the performance of a PGS developed in a European population, on an admixed group of individuals with both African and European ancestry. We further test a method using ancestry partial scores to better adapt this score for use in an admixed population.

We tested the ASD PGS from Grove et al. 2019 in individuals of African-European admixed ancestries from the Simons Foundation Powering Autism Research (SPARK) dataset. SPARK was formed by a collection of 31 university-affiliated research clinics from across the United States who collected the genetic and phenotypic data of individuals diagnosed with ASD and their families. The SPARK sample population of 68,316 has 4,608 individuals with African ancestry (7%) and 50,970 with European ancestry (75%). The African ancestry subset had a significantly decreased odds ratio for the PGS as compared to that of the European subset, indicating that the PGS explained less of the variation in disease status of individuals with African ancestry, though the polygenic score was not well powered in either population (EUR OR 1.02, 95% CI:1.02-1.03 & AFR OR 0.99, 95% CI: 0.98-1.01). In addition, we tested the accuracy of ancestry partial scores using the same European-derived PGS weights to calculate a European-specific and African-specific score. We created an unrelated subset of 3,758 individuals with African-European ancestries (average African proportion = 50%), 2,143 of whom have an ASD diagnosis. There was no significant increase in explanatory power in the African admixed individuals with the partial scores (Total OR 1.02, 95% CI:0.98-1.05 & AFR partial OR 1.00, 95% CI: 0.95-1.05). Future work should focus on increasing sample size and developing an ASD PGS in African American and other admixed populations to better understand the role of common variants in ASD diagnosis outside of a European population.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1801 The ancestral origin of rare variants contributing to Inflammatory Bowel Disease in Hispanics

Authors:

T. Haritunians[1], J. L. Mccauley[2], E. A. Torres[3], L. Gomez[2], D. Li[1], M. Daly[4,5], C. Stevens[4], J. Leavitt[6], O. M. Damas[2], M. A. Quintero[2], J. Cho[7], M. T. Abreu[2], D. P. B. McGovern[1], A. H. Beecham[2]; [1]Cedars-Sinai Med. Ctr., Los Angeles, CA, [2]Univ. of Miami, Miami, FL, [3]Univ. Puerto Rico, San Juan, Puerto Rico, [4]The Broad Inst., Cambridge, MA, [5]Univ. of Helsinki, Helsinki, Finland, [6]GastroHlth., Miami, FL, [7]Icahn Sch. of Med. at Mount Sinai, New York, NY

Genetic studies of inflammatory bowel disease (IBD) have largely focused on the identification of common variants contributing to disease pathogenesis over the last several decades. Rapidly expanding sequencing technologies and extensive collaborative networks have made it possible to investigate the role of rare variation on IBD and its individual components, Crohn's disease (CD) and ulcerative colitis (UC). A recent large-scale sequencing study of more than 30,000 CD patients and 80,000 population controls identified 45 rare coding variants associated with CD. We sought to establish the relationship of each of these variants to CD, UC, or IBD in a Hispanic cohort of 1660 IBD cases (927 CD; 689 UC; 44 undetermined IBD) and 5614 controls and to determine the role of ancestral allele origin on risk. In total, 37 of the variants were present and passed quality control of whole genome sequence data. Local ancestry was evaluated with RFMixV2. Single variant associations controlling for population substructure were assessed with Regenie, and local-ancestry aware regression was performed with Tractor to assess ancestral-specific allele frequencies and effects. For CD, we observed replication ($p<0.05$) for 8 of 37 variants. The strongest associations were observed for known *IL23R* protective R381Q and *NOD2* risk R702W. Additional known coding variants replicating for CD included *LRRK2* N2081D, *TYK2* A928V, and the *CARD9* splice variant. We identified an association ($p<0.05$) with only two variants, *IL23R* R381Q and *CCR7* M7V, with UC in our cohort. Almost all the associated rare variants were observed on a European background within the admixed sample. One variant, *SNX20* R109H, demonstrated a striking Amerindian-specific effect for UC (p=3.7E-06; compared to mixed model regression for all ancestral origins p=0.38) with higher allele frequency observed amongst Amerindian alleles (MAF = 0.06) compared to European alleles (MAF = 0.01) within our admixed cohort. Further work is needed to assess the linkage disequilibrium (LD) between this variant and established *NOD2* risk variants across populations. We have demonstrated that although European-identified rare CD risk variants are largely of European origin within the Hispanic admixed sample, there is some evidence for heterogeneity of effect across ancestral backgrounds due to differing allele frequency and LD structure. Incorporating ancestrally diverse populations is critical to furthering our understanding of IBD risk and accelerating individualized treatment for all patients.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1802 The Causal Relationship between Type 2 Diabetes mellitus and Chronic Periodontitis - A Mendelian Randomization study.

Authors:

M. Olujitan[1], A. Alade[2], E. Aladenika[3], A. Oladayo[1], W. Awotoye[3], T. Busch[2], S. Miller[2]; [1]Univ. of Iowa, Coralville, IA, [2]Univ. of Iowa, Iowa City, IA, [3]Univ. of Iowa, Iowa, IA

Introduction: Periodontitis is a complex, multifactorial inflammatory disease of the tooth-supporting structures. It has far-reaching consequences on quality of life and is a matter of public health concern. Several Epidemiological studies have implicated type 2 Diabetes Mellitus (T2DM) as a prominent risk factor for the disease, with some even suggesting a bidirectional relationship between T2DM and Periodontitis. However, due to the complex nature of these conditions, the underlying mechanisms that may explain this interrelationship are still poorly understood. Our study aimed to investigate genetic liability to periodontitis through T2DM. **Methods:** We employed the Mendelian Randomization technique which uses genetic variants as Instrumental Variables to assess causal relationships between two traits. We accessed summary statistics on Single Nucleotide Polymorphisms (SNPs) that showed genome-wide significance for T2DM (62 SNPs) and extracted the summary statistics for the T2DM-associated SNPs on periodontitis. This data is available through the *MRC Integrative Epidemiology Unit (IEU) OpenGWAS* database. Our analyses utilized the Inverse-Variance Weighted (IVW) method, and the MR-Egger method was used to assess horizontal pleiotropy. In addition, we performed a case-control association analysis in PLINK using genotype data from T2DM-significant SNPs obtained from 252 cases and 904 controls at the College of Dentistry, University of Iowa. **Results:** Our results showed a causal relationship between DM and Chronic Periodontitis (IVW method; OR 1.09, 95%CI - [1.01, 1.19], p = 0.03), suggesting that the genetic predisposition to Diabetes mellitus increases the risk of Chronic Periodontitis by 9% in this population of individuals. The MR-Egger method showed no evidence of directional pleiotropy. Additionally, our case-control analysis showed no significant SNP associations. **Conclusion:** These findings further give credence to the observed association between DM and Periodontitis which has been reported in previous epidemiological studies and contribute to a better understanding of the complex relationship between DM and periodontitis.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1803 The Cellular Composition of the Subgenual Anterior Cingulate Cortex in Bipolar Disorder, Schzophrenia and Major Depressive Disorder

Authors:

I. Adoremos, N. Akula, A. Schulmann, F. McMahon, A. Pavan, S. Marenco; NIH, Bethesda, MD

Gene expression differences in the subgenual anterior cingulate cortex (sgACC) from people with psychiatric disorders have been studied, but little is known about changes in cell type composition that drive these differences. Using various RNA sequencing data sources, this study aims to identify cell-specific biomarkers for various psychiatric disorders by estimating changes in neuronal cell types in the sgACC, an important component of limbic circuitry. We employed deconvolution, a computational method to estimate cell type-specific proportions from bulk tissue with a single-cell reference dataset. Postmortem sgACC samples from 185 donors (55 controls, 44 SCZ, 35 BD, 51 MDD) underwent RNA sequencing. Bisque deconvolution software was implemented, due to its ability to manage noisy biological data. Cell type fractions for endothelial cells, microglia, astrocytes, oligodendrocytes, oligodendrocyte precursor cells (OPC), inhibitory neurons and excitatory neurons were acquired for each sample in the deconvolution analysis. ANOVA was subsequently implemented to identify, for each cell type, whether diagnosis impacts proportion estimates, subsequent pairwise t-tests to identify key differences in cell type proportions across each diagnostic group, and Tukey's tests to adjust p values for multiple comparisons. OPCs showed the most significant impact in individuals with SCZ, with ANOVA results ($p < 0.001$, $F = 6.43$), a significant t-test for SCZ ($t = 3.00$, $p = 0.003$), and Tukey's test (p-adjusted= 0.03). Inhibitory neurons also showed significant ANOVA results ($p = 0.009$, $F = 3.99$), suggesting reduced neurons in SCZ individuals. Astrocytes showed a marginally significant ANOVA result ($p = 0.05$, $F = 2.65$), with SCZ individuals having higher proportions. Tukey's tests showed no deviation for inhibitory neurons (p-adjusted= 0.14) and astrocytes (p-adjusted = 0.24) in SCZ individuals from controls. Results suggest that the changes in cell type proportions are most pronounced in SCZ individuals, specifically among OPC, inhibitory neurons, astrocytes. In conclusion, these findings suggest associations between psychiatric diagnoses and cell-type proportions in postmortem sgACC. Limitations of this study include potential variability introduced by the postmortem nature of samples, limitations of the deconvolution method in capturing cellular heterogeneity, and the adequacy of the sample size for detecting smaller effects. Future work could benefit from larger and more diverse samples, improved deconvolution techniques, and novel biomarker identification to offer potential therapeutic targets.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1804 The ClinGen Kidney Disease Clinical Domain Working Group: Defining the genetic architecture of kidney disease.

Authors:

A. Byrne[1], H. Dziadzio[1], E. Edoh[1], A. Girod[1], V. Gitau[1], P. Robertson[1], T. P. Sneddon[2], R. Webb[1], M. G. Sampson[3,1,4], Z. Stark[5,6], R. Lennon[7,8], A. J. Mallett[9,10,11], ClinGen Kidney Disease Clinical Domain Working Group; [1]Broad Inst. of MIT and Harvard, Cambridge, MA, [2]Univ. of North Carolina Hlth., Chapel Hill, NC, [3]Boston Children's Hosp., Boston, MA, [4]Harvard Med. Sch., Boston, MA, [5]Murdoch Children's Res. Inst., Melbourne, Australia, [6]Australian Genomics, Melbourne, Australia, [7]The Univ. of Manchester, Manchester, United Kingdom, [8]Royal Manchester Children's Hosp., Manchester, United Kingdom, [9]Townsville Hosp. and Hlth.Service, Townsville, Australia, [10]James Cook Univ., Townsville, Australia, [11]The Univ. of Queensland, Brisbane, Australia

Kidney disease affects one in ten people with 10% of individuals having a monogenic cause. Genomic medicine holds enormous promise for transforming the diagnosis, care, and treatment of patients and their families, but in order to reach its full potential, systematic evaluation of the clinical validity of gene-disease relationships (GDRs) and accurate assessments of variant pathogenicity are required. To address this need, the ClinGen Kidney Disease Clinical Domain Working Group (CDWG) was established in 2019, convening Gene Curation Expert Panels (GCEPs) and Variant Curation Expert Panels (VCEPs) across five broad disease areas to cover the full spectrum of nephropathies: Kidney Cystic and Ciliopathy Disorders, Tubulopathies, Glomerulopathies, Congenital Anomalies of Kidney and Urinary Tract (CAKUT), and Complement-Mediated Kidney Diseases.

GCEPs addressing all five disease areas are working to evaluate GDRs following the framework defined by ClinGen, with a total of 420 genes identified as 'in-scope'. To date, 120 GDRs have been assessed resulting in 82 Definitive, 5 Strong, and 11 Moderate classifications, with the remaining GDRs being classified as Limited (16), Disputed (5), and No known disease association (1), highlighting the importance of such rigorous assessment before including genes in clinical testing panels. VCEPs addressing the two most common genetic kidney diseases, Polycystic Kidney Disease (*PKD1* and *PKD2*) and Alport Syndrome (*COL4A3*, *COL4A4* and *COL4A5*), are also underway, working to define gene- and disease- specific modifications to the ACMG/AMP variant classification framework. Once complete, these refined criteria are expected to help increase consistency in the application of the framework for these genes, and to improve the accuracy of variant classification. Additionally, leadership within the Kidney Disease CDWG are contributing to efforts, both internal and external to ClinGen, to establish consensus recommendations for kidney disease nomenclature and ontology. At present, the inaccuracies and inconsistencies in disease naming can hinder clinical recognition, impacting patients by restricting eligibility and access to appropriate testing, monitoring, and treatment.

Together, the work of the Kidney Disease CDWG to define the clinical relevance of genes and variants across the spectrum of nephropathies will help ensure diagnostic accuracy as genomic testing moves to be commonplace in the nephrology clinic.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1805 The combined effect of clonal hematopoiesis of indeterminate potential and mosaic chromosomal alterations on cardiovascular health outcomes in individuals diagnosed with cancer in the UK Biobank.

Authors:

M. Sun[1], J. Sandoval[1], L-P. Lemieux Perreault[2], A. Barhdadi[1], M-C. Cyr[1], L. Busque[3], J-C. Tardif[1], M-P. Dube[1]; [1]Montreal Heart Inst., Montreal, QC, Canada, [2]Beaulieu-Saucier Pharmacogenomics Ctr., Montréal, QC, Canada, [3]Maisonneuve-Rosemont Hosp, Montreal, QC, Canada

Background: Individuals diagnosed with cancer face an elevated risk for cardiovascular disease. Recent studies have linked clonal hematopoiesis (CH) to cardiovascular outcomes in the broader population. This investigation aims to ascertain the influence of clonal hematopoiesis of indeterminate potential (CHIP) on the short-term and long-term cardiovascular health and survival outcomes within a cohort solely comprised of cancer patients. Additionally, we examined the clinical relevance of co-existing mosaic chromosomal alterations (mCAs) with CHIP gene mutations in relation to cardiovascular and survival outcomes. **Methods**: Our study comprised 19,566 cancer-diagnosed individuals from the UK Biobank. Whole-exome sequencing data, derived from blood DNA, served to identify CHIP mutations, while genotyping array data was employed to delineate mCAs. Primary endpoints included time to cardiovascular-related emergency admission, death from cardiovascular disease causes, death from coronary artery disease causes, and death from any cause, based on a median of 7.2 years of follow up. Cox regression models were used with adjustments for age, sex, smoking status, chemotherapy, radiotherapy, prevalent cardiovascular disease, the duration between recruitment date and cancer diagnosis date, and genetically inferred genetic ancestry. **Results**: Of the 19,566 individuals diagnosed with cancer, 1,166 (6.0%) were found to carry at least one CHIP mutation. Of all patients, 302 (1.2%) individuals carried both CHIP and mCA, whereas 864 (4.4%) carried CHIP alone, and 3,648 (18.6%) carried mCA alone. Multivariable analyses showed that CHIP mutation carriers had a higher risk of death from cardiovascular diseases compared to non-carriers (hazard ratio [HR]: 1.66, 95% confidence interval [CI]: 1.14-2.42, *P*=0.008). CHIP alone led to an increased risk of death from cardiovascular causes compared to individuals with neither type of CH (HR: 1.74, 95% CI: 1.12-2.87, *P*=0.014). **Conclusions**: Our findings suggest that CHIP mutations independently contribute to an increased risk of death from cardiovascular causes among cancer patients, regardless of mCA co-occurrence.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1806 The contribution of common and rare exonic variants to cognitive performance across early childhood and adolescence

Authors:

D. Malawsky, M. Koko, W. Huang, M. Hurles, H. Martin; Wellcome Sanger Inst., Hinxton, United Kingdom

Twin studies and SNP-based estimates suggest the heritability of cognitive performance (CP) increases across development, with estimates of narrow sense heritability increasing by >50% between early childhood and adulthood. However, the contribution of rare variants to CP has been relatively understudied. Here, we investigate the contribution of common and rare exonic variants to CP in >6,000 unrelated individuals from the Avon Longitudinal Study of Parents and Children. Between 702 and 5,125 of the individuals with exome sequence and SNP array data had IQ measured at any given time point (at age 4, 8, or 17). We imputed missing IQ values at different ages using a range of variables including development scores at 18 months, parental socioeconomic status, birth weight, and cognitive-behavioral traits measured across childhood. Cross-validation of the imputation showed high imputation accuracy ($r>0.6$). We find that the Shet burden, a measure of the exome-wide burden of likely deleterious loss-of-function variants (LoFs) that takes into account genic constraint, is negatively associated with IQ across all ages ($p<10^{-4}$). In a linear mixed modeling framework and based on imputed IQ data, we find the variance explained by Shet burden decreases with age (age*Shet interaction $p<10^{-4}$). Compared to those with median Shet burden, individuals in the top 5% have an average decrease of 6 IQ points at age 4, which is attenuated to a 3 point difference at age 17. This trend was also evident using the non-imputed IQ data, albeit less significant due to reduced sample size ($p=0.001$). To evaluate the convergence of rare and common variant effects, we evaluated the relative contribution to Shet burden from top-ranked genes in an educational attainment (EA) GWAS. We found an enriched contribution from 916 genes fine-mapped in EA GWAS to the association between Shet burden and IQ at all ages, accounting for 60% of the total variance explained and ~4 times more variance than a randomly-chosen set of 916 genes with the same Shet distribution. In contrast to the rare variant effect, we find that EA and CP polygenic indices (PGIs) explained increasing variance with age (age*PGI interaction $p<10^{-10}$). In conclusion, our study provides insight into the contribution of common and rare exonic variants to CP across development in a comprehensive manner, highlighting the diminishing impact of deleterious LoF variants and increasing impact of common variants to the heritability of IQ across ages. We are now seeking to replicate these findings in two other UK birth cohorts of comparable sample sizes with new exome-sequence data, Millenium Cohort Study and Born in Bradford.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1807 The effect of *GBA* variants on Turkish patients with Parkinson's disease.

Authors:

T. Abali[1], M. K. Yekedüz*[2], R. Yilmaz[3,4], G. Kayis[1], E. Y. Unutmaz[5], A. V. Karacan[1], S. N. Kibrit[1], S. Ceylaner[6], M. C. Akbostancı[3,4], F. T. Eminoğlu[2,7]; [1]Ankara Univ. Sch. of Med., Ankara, Turkey, [2]Ankara Univ. Sch. of Med., Dept. of Pediatric Metabolism, Ankara, Turkey, [3]Ankara Univ. Sch. of Med., Dept. of Neurology, Ankara, Turkey, [4]Ankara Univ. Brain Res. Ctr., Ankara, Turkey, [5]Ankara Univ., Faculty of Language, History and Geography, Dept. of Psychology, Ankara, Turkey, [6]Intergen Genetics Ctr., Ankara, Turkey, [7]Ankara Univ. Rare Diseases Application and Res. Ctr., Ankara, Turkey

Objective We investigated genetic and clinical aspects of Turkish PwP with and without *GBA* variants.
Background Genetic risks play an important role in Parkinson's disease (PD) which is the second most common neurogenerative disease in elderly population. One risk for PD is variants in the *GBA* gene, which encodes the lysosomal enzyme glucocerebrosidase. The effect of heterozygous *GBA* variants on patients with PD (PwP) is well described; however, comprehensive data on this topic from Asia Minor has yet to be reported.
Method Deep-phenotyping of the PwP was achieved by evaluating disease-related information, and comprehensive motor, and non-motor symptom assessments including cognitive, psychiatric and autonomous nervous system symptoms. *GBA* variants were investigated by whole exome sequencing. The classification of the detected variants was done according to ACMG criteria-impact for Gaucher disease. Plasma beta-glucocerebrosidase enzyme and its substrate Lyso-Gb1 levels in PwP with a detected variant were also achieved. Clinical assessments were compared between the groups of PwP with a pathogenic *GBA* variant (PwP-pat-*GBA*) and without any variant (PwP-no-*GBA*). PwP with benign or unknown variants were excluded. Enzyme and substrate levels were also compared between PwP-pat-*GBA* and benign variants (PwP-ben-*GBA*).
Result Of the 379 PwP, 30 were in the PwP-pat-*GBA* group, 25 were PwP-ben-*GBA*, and 324 were PwP-no-*GBA*. Comparison of PwP-pat-*GBA* and PwP-no-*GBA* groups showed that the age of onset, was significantly lower in the PwP-pat-*GBA* (mean, (SD), 52.7 (10.5) vs. 57.3 (11.9); p=0.042). Also, the levodopa equivalent daily dose (mg/day) was significantly higher in the PwP-pat-*GBA* (973.8 (506.8) than PwP-no-*GBA* (728.3 (434.5), p=0.004) indicating a need for higher dose of medication for the former group. In addition, the presence of hallucinations and psychosis was higher in PwP-pat-*GBA* (n=10 (34.5%)) than in PwP-no-*GBA* (n = 46 (14.4%); p=0.005). Orthostatic hypotension was also significantly more prevalent in PwP-pat-*GBA* (n=16 (53.3%)) than PwP-no-*GBA* (n=102 (31.9%); p=0.017). Other assessments showed no significant differences. Comparison of the groups for enzyme and substrate levels showed no significant difference despite the lower enzyme and higher Lyso-Gb1 levels in the PwP-pat-*GBA* group.
Conclusion We report the largest data on *GBA* variants in PwP from Asia Minor. These results are compatible with previous studies and underline that PwP with pathogenic *GBA* variants have an earlier age of onset and perform worse in some clinical assessments. More studies are needed to elucidate the exact mechanisms linking lysosomal *GBA* and PD.
*co first author

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1808 The Genetic Architecture of Bronchopulmonary Dysplasia in Extremely Preterm Infants.

Authors:

M. Guardado[1], C. Chapin[1], D. Torgerson[2], R. Hernandez[2], P. Ballard[1], M. Steurer[1]; [1]Univeristy of California, San Francisco, San Francisco, CA, [2]UCSF, San Francisco, CA

Extremely premature infants are at high risk of developing Bronchopulmonary Dysplasia (BPD), a developmental lung disease characterized by the continuing requirement for supplemental oxygen and respiratory support at 36 weeks postmenstrual age. 10,000-15,000 infants develop BPD in the United States each year. Rates of BPD vary between racial/ethnic groups, with infants of mothers who self-report as Black or African-American having lower rates of BPD than infants of mothers who self-report as non-Hispanic white. Genetic studies have identified several variants associated with BPD, however, there is continuing debate based on the degree to which BPD is a heritable disease. Existing twin-based studies have produced conflicting results regarding the heritability of BPD. Using genotype arrays and genome-wide imputation, we analyzed ~34 million single nucleotide polymorphisms (SNPs) across 751 unrelated individuals with/without BPD from two clinical cohorts, spanning three racial/ethnic US population groups. We estimated the narrow sense heritability of BPD by binning alleles based on their frequency and performed frequency-binned heritability analysis to describe the common-to-rare variant architecture of BPD. We estimate that the heritability of BPD is ~20%, with rare variants (MAF<0.01) contributing to the signal that is found. The study's findings suggest that BPD has a heritable component. By demonstrating the heritability of BPD, the study supports the notion that genetic variations play a role in the risk and development of the disease. The signal of increased heritability of BPD due to rare variants may explain some of the challenges faced by genetic association studies, which are known to have a lower power to detect causal effects for rare variants than common variants. The study suggests the need for whole-genome sequencing in BPD studies to capture rare genetic variant's effect on BPD effectively.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1809 † The genetic architecture of pain intensity in the Million Veteran Program

Authors:

S. Toikumo[1,2], R. Vickers-Smith[3,2], Z. Jinwala[1], H. Xu[1], D. Saini[1], E. Hartwell[1,2], M. V. Pavicic[4], K. A. Sullivan[4], K. Xu[5,6], D. A. Jacobson[4], J. Gelernter[5], C. T. Rentsch[7,6], Million Veteran Program, E. Stahl[8], M. Cheatle[1], H. Zhou[5], S. G. Waxman[5,6], A. C. Justice[5,6], R. L. Kember[1,2], H. R. Kranzler[1,2]; [1]Univ. of Pennsylvania, Philadelphia, PA, [2]Crescenz VAMC, Philadelphia, PA, [3]Univ. of Kentucky, Lexington, KY, [4]Oak Ridge Natl. Lab., Oak Ridge, TN, [5]Yale Univ Sch. of Med., New Haven, CT, [6]VA Connecticut Hlth.care System, New Haven, CT, [7]London Sch. of Hygiene & Tropical Med., London, United Kingdom, [8]Regeneron Genetics Ctr., Tarrytown, NY

Chronic pain is a common problem, with more than one-fifth of adult Americans reporting pain daily or on most days. It adversely affects quality of life and imposes substantial personal and economic costs. Efforts to treat chronic pain using opioids played a central role in precipitating the U.S. opioid crisis. Despite an estimated heritability of 25-50%, the genetic architecture of chronic pain is not well characterized, in part because study samples have largely been limited to individuals of European ancestry. To help address this knowledge gap, we conducted a cross-ancestry meta-analysis of pain intensity in 598,339 participants in the Million Veteran Program, which identified 125 independent genetic loci, 82 of which are novel. Ancestry-specific GWASs identified 86 independent loci in European Americans, 1 independent locus (nearest gene *PPARD*; chr 6) in African Americans, and 2 independent loci (nearest genes *RNU6-461P*; chr 3 and *RNU6-741P*; chr 15) in Hispanic Americans. Pain intensity was genetically correlated with other pain phenotypes, level of substance use, substance use disorders, other psychiatric traits, education level, and cognitive traits. In Mendelian randomization analyses, pain intensity had a significant positive bidirectional causal effect with genetically predicted opioid use, depressed affect subcluster, major depressive disorder, neuroticism, use of drugs to treat peptic ulcer, and smoking cessation (coded as current smoking). Integration of the GWAS findings with functional genomics data showed enrichment for putatively causal genes (n = 142) and proteins (n = 14) expressed in brain tissues, specifically in GABAergic neurons. Drug repurposing analysis identified anticonvulsants, beta-blockers, and calcium-channel blockers, among other drug groups, as having potential analgesic effects. Our results provide insights into key molecular contributors to the experience of pain and highlight potential drug targets.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1810 The genetic regulatory architecture of Alzheimer's disease in iPSC-derived microglia from diverse ancestries.

Authors:

S. Moura[1], L. Bertholim Nasciben[1], A. Ramirez[1], J. Shepherd[1], L. Coombs[1], D. Van Booven[1], S. Simon[1], P. Whitehead[1], L. Adams[1], T. Starks[2], P. Mena[1], M. Illanes-Manrique[3], S. Tejada[1], G. Byrd[4], M. Cornejo-Olivas[5], B. Feliciano-Astacio[6], L. Wang[1], A. Griswold[1], J. Young[1], M. Pericak-Vance[1], D. Dykxhoorn[1], J. Vance[1]; [1]Univ. of Miami, Miami, FL, [2]Wake Forest Sch. of Med., Winston-Salem, NC, [3]Inst. Natl. de Ciencias Neurológicas, Lima, Peru, [4]NC A&T State Univ, Greensboro, NC, [5]Inst. Natl. de Ciencias Neurologicas, Lima, Peru, [6]Central Univ. of the Caribbean, Bayamon, Puerto Rico

Most Alzheimer's disease (AD)-GWAS studies have focused on Non-Hispanic Whites. However, recent studies have begun to study AD genetic risk in diverse populations, including admixed populations (e.g., Hispanics and African Americans). We have previously shown the brain genomic regulatory architecture (GRA) varies among different ancestries, particularly in astrocytes and microglia, potentially leading to different genetic susceptibility in AD. To further dissect the ancestry-specific GRA in AD, we assessed the GRA and transcriptome of iPSC-derived central nervous system cells, including microglial cells, from diverse ancestries. These analyses will complement those from the brain providing a more complete understanding of the ancestry-specific regulatory architecture underlying AD. Specifically, we characterized the GRA in iPSC-derived microglia (iMG) from individuals with high global Amerindian (AI), African (AF), or European (EU) ancestries. We derived iPSCs lines with >90% genomic content from different ancestries - EU, AF, and AI from AD patients and unaffected controls. iPSC lines were validated for pluripotency and chromosomal stability and were differentiated into MG. iMG cells were validated via immunocytochemistry for cell type-specific markers. To study the regulatory architecture of these admixed populations and the associated impact on gene expression, we performed bulk ATAC-seq and RNA-seq, followed by differential accessibility and expression analyses. We identified a total of 2,466 differentially expressed genes (DEGs) and 1,743 differentially accessible genes across ancestries. Interestingly, several DEGs were involved in cholesterol biosynthesis and metabolism, lipid metabolism, lysosomal activity, endocytosis, and Aβ-clearance - all highly relevant processes in AD pathology. Moreover, we observed that 10 of the 78 AD-GWAS risk loci were differentially expressed. Specifically, *ABCA1, ABI3, CTSB,* and *TREM2* between AI and AF, *AGRN* and *APP* between AF and EU, and *APOE, CR1, JAZF1,* and *NTN5* between AI and EU. Surprisingly, we observed that only *CTSB* is also differentially accessible between AI and AF. Lastly, on the chromatin accessibility level, we observed the most differences between AI and AF, followed by AF and EU, and AI and EU. We provide novel insights into ancestry-specific risk factors and candidate genes in AD pathophysiology. Specifically, we report novel data on chromatin accessibility and transcriptomics data in AF and AI microglia, a key player in AD pathology. Our data suggest the possibility of differential effect of AD-associated variants in the different ancestries.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1811 The impact of a G6PD variant on the risk of developing diabetic complications

Authors:

N. Arora[1], P. Straub[2], M. Shuey[2], N. Cox[2], J. HIRBO[2]; [1]Tulane Univ. Sch. of Med., New Orleans, LA, [2]Vanderbilt Univ. Med. Ctr., Nashville, TN

Glycated hemoglobin (HbA1c), a key indicator for diabetes diagnosis and prognosis, has been shown to have significantly higher average levels among individuals of African ancestry compared to individuals of European ancestry. However, HbA1c may be falsely lowered in individuals with the rs1050828 variant causing glucose-6-phosphate dehydrogenase (G6PD) deficiency, a hemolytic red blood cell disorder that disproportionately affects individuals of African descent, leading to an extended undiagnosed period of diabetes while still being hyperglycemic. The effects of a prolonged undiagnosed period on the risk of developing diabetic related complications in patients with diabetes and the rs1050828 variant have not been previously quantified. To determine the risk of complications in patients with diabetes and the rs1050828 variant compared to patient with diabetes without this variant, a longitudinal cohort of individuals with a coded Type 1 (T1D) or Type 2 (T2D) diabetes diagnosis with or without G6PD deficiency was compiled from BioVU from 2007 to June 2020, along with certain comorbid conditions they developed. We utilized poisson regression modeling to determine the difference in risk of diabetic complications between individuals of African ancestry with (n=1038) and without (n=4867) the G6PD deficiency variant. We estimated a 21.6% increased risk of developing diabetic complications in patients with T1D or T2D of African ancestry with the G6PD variant (rs1050828) compared to those without the variant (p=2.75E-06). In addition, there was a 60.5% increased risk of developing ketoacidosis (p=0.00218) and 76.2% increased risk of developing ophthalmic manifestations (p=0.000827) for those with the variant relative to those without in subjects with T1D. Because the rs1050828 variant causing G6PD deficiency falsely lowers HbA1c and increases the risk of developing diabetic related complications, especially ophthalmic manifestations in T1D, we showcase the need to routinely test individuals of African ancestry for G6PD variants and utilize multiple measures of glycemic control in addition to HbA1c for diabetes diagnosis and prognosis.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1812 The influence of genetic interaction between CHRNB4 andThe influence of genetic interaction between CHRNB4 and depressive symptoms on smoking dependence

Authors:

W. Cui; Zhejiang Univ., Hangzhou, China

Smoking dependence is a chronic mental disorder contributed by both genetic and environmental factors. Depression increases the risk of becoming a smoker, primarily because of difficulty in quitting. Smoking behavior can be influenced by the interaction between genetic variants and depressive symptoms. However, among individuals with mild to severe depression, the genetic differences in the susceptibility to smoking dependence and the underlying mechanisms remain to be elucidated. By using gene-environment interaction analysis, we found that the genetic interaction between CHRNB4 and CES-D, a measure of depressive symptoms, contributes to smoking dependence. Besides, the smoking motivation analysis revealed a significant association between CHRNB4 variants and the pleasant feelings after smoking. Moreover, through gene-gene interaction analysis, significant interactions between CHRNB4 and glutamate receptor genes (GRIA1, GRIN2B and GRM1) were determined in smokers, but not in non-smokers. Further, in animal studies, by using fear conditioning test, significant difference has been determined in Chrnb4 KO and WT mice under a chronic restraint stress (CRS) model, indicating a sustaining negative affect in mood disorders may be involved in the regulatory effect of Chrnb4 on nicotine dependence.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1813 The large effect of *APOC3* A43T SNP on serum triglycerides is partially independent from that of serum APOCIII levels In Amerindians

Authors:

W-C. Hsueh[1], L. J. Baier[2], R. L. Hanson[1]; [1]NIDDK, Phoenix, AZ, [2]NIH, Phoenix, AZ

Introduction: We previously identified a major locus for serum triglyceride (TG) levels in Amerindians, which largely reflects the effect of a functional SNP (A43T, or rs147210663) in *APOC3*. The T43 allele frequency ~3% in our Amerindian cohort but is extremely rare in non-Amerindian populations. It truncates APOCIII, causing loss of function and a lowering of circulating APOCIII levels. Its effect reduces TG level by ~1 SD unit (per allele), which is among the largest effects for any complex trait reported to date.
Hypothesis: To further explore the functional mechanism underlying this association, we analyzed the relationships among A43T, TG and APOCIII levels; we hypothesized that this SNP's effect on TG may be mediated through low levels of APOCIII, as measured in the peripheral blood.
Methods: We measured serum APOCIII level in 4,569 Amerindian subjects who were a subset of a cohort study in which the A43T genotype, TG and APOCIII levels were measured. There were 210 carriers of a T43 allele. All analyses were modeled using linear regressions based on cross-sectional data, with age, sex, diabetes status, specimen age, batch effects, and population admixture included as covariates. The natural logarithms of APOCIII and TG levels were taken to reduce skewness.
Results: The phenotypic correlation between APOCIII and TG was 0.50. We found a strong association between the A43T SNP and TG levels (β=-0.90±0.065 SD unit per copy of the T43 allele, p=3.5e-47, explaining 4.4% variance) that was only partially attenuated with further adjustment for APOCIII levels (β=-0.49±0.058 SD unit, p=1.4e-21, explaining 2.0% variance); this suggests partial mediation at 46%. On the other hand, there was also a strong association between the A43T SNP and APOCIII levels (β=-0.78±0.068 SD unit, p=3.1e-35, explaining 3.3% variance) which remained significant after adjustment for the TG effect (β=-0.31±0.061 SD unit, p=8.7e-10, explaining 0.82% variance).
Conclusion: Our results suggest that the *APOC3* A43T SNP effect on TG may be partially independent from APOCIII; however, further studies are needed to explore whether APOCIII distribution in lipids or plasma influence the observed relationship. Given partial mediation, these findings may lend further strength to the possibility of making APOCIII a therapeutic target for CVD interventions.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1814 The Major Depression Working Group of the Latin American Genomics Consortium (LAGC-MDD): Introduction, Insights, and Pipelines.

Authors:

R. Peterson[1], N. Corral-Frias[2], D. Nunez[3], V. Ota[4], J. Cárcamo[1], C. Machado[1], L. Hernandez Amaya[5], M. Velásquez Toledo[6], S. Nagamatsu[3], E. Dunn[7], B. Torres-Hernández[8], C. Lattig[9], P. Giusti-Rodriguez[10], J. Montalvo Ortiz[3]; [1]SUNY Downstate Hlth.Sci. Univ., Brooklyn, NY, [2]Univ. de Sonora, Sonora, Mexico, [3]Yale Sch. of Med., New Haven, CT, [4]UNIFESP, Sao Paulo, Brazil, [5]Univ. de Los Andes, Bogota, Colombia, [6]Univ. de los Andes, Bogotá, Colombia, [7]Massachusetts Gen. Hosp., Boston, MA, [8]Univ. of Puerto Rico-Med. Sci. Campus, San Juan, PR, [9]Univ. de los Andes, Bogota, Colombia, [10]Univ. of Florida Coll. of Med., Gainesville, FL

Major depressive disorder (MDD) is considered a complex disorder arising from both genetic and environmental factors and interactions between them. Twin studies, from primarily European ancestries, estimate the heritability of MDD as approximately 37%. Although there have been advancements in the etiological understanding of MDD, several limitations remain, including the lack of ancestral diversity in cohorts and large heterogeneity among the depression phenotypes assessed. Given that MDD is one of the most common mental health disorders in Latin America and rates are on the rise, it is imperative that there are large-scale focused consortia efforts to advance the understanding of this debilitating disorder in these under-studied populations. The Latin American Genomics Consortium (LAGC) was founded in 2019 in order to accelerate capacity building, improve representation of Latin American individuals in genetic studies, forge methodological advancements, and provide a forum for networking across countries. Building on these ideals, in 2022 the LAGC-MDD Working Group was formed and seeks to connect depression researchers internationally in order to perform the first large-scale genome-wide association study (GWAS) meta-analysis of depression phenotypes in individuals of admixed American descent from Latin America, the United States, and abroad. The purpose of this presentation is to provide an introduction to the LAGC-MDD Working Group (OSF: https://tinyurl.com/mr3xpwby) while providing insights from our genomic analytic pipelines and phenotypic harmonization efforts. Currently, we have aggregated data from across 33 datasets. These cohorts were collected from 5 countries including Brazil, Colombia, Mexico, Peru, and the US. The approximate meta-analysis sample size is 135,000 cases and 240,000 controls. Importantly, we explicitly assess and model heterogeneity across depression phenotypes from psychometric and genomic approaches (e.g., gSEM, Tractor) to parse heterogeneity and maximize power of our datasets for GWAS. These approaches are expected to improve polygenic risk prediction especially when coupled with environmental risk factors. Although 33 datasets have been aggregated, the majority of samples are from the US. Data collection and capacity building efforts in Latin American countries are critically needed and the LAGC is actively working towards this. To improve depression rates in Latin America and beyond, we will need to advance our knowledge of the risk architecture of MDD. The best way forward is to ensure precise measurement and ancestral inclusivity among samples - major goals of the LAGC-MDD working group.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1815 The Polygenic Genetics Landscape of Nephrolithiasis: Identifying Index SNPs and Their Associations by Linkage Disequilibrium

Authors:

E. Su[1], A. Salem[1], D. Hoying[1], J. Chang[1], M. Baum[2], D. Bodner[3], F. Hildebrandt[4], F. Schumacher[5], C-H. Wu[1]; [1]Case Western Reserve Univ. / Univ. Hosp., Cleveland, OH, [2]Boston Children's Hosp./Harvard Med. Sch., Boston, MA, [3]Univ. Hosp., Cleveland, OH, [4]Boston Children's Hosp, Boston, MA, [5]Case Western Reserve Univ, Cleveland, OH

Nephrolithiasis, alternatively referred to as urolithiasis, kidney stones, or urinary stones, are hardened formations composed of minerals and salts that form within the urinary system. The causes of nephrolithiasis are multifaceted, and the likelihood of developing this condition is influenced by a combination of genetic, metabolic, and environmental factors. Research indicates a substantial involvement of genetic factors and familial background in the occurrence of kidney stones.We reviewed all genome-wide association studies (GWAS) pertaining to nephrolithiasis, and compiled a list of SNPs that reached genome-wide significance ($p < 5 * 10\text{-}8$). However, none of the SNPs reaching genome-wide significance were directly replicated. We then grouped these SNPs into linkage disequilibrium (LD) blocks by identifying their nearest gene, and used a LD correlation coefficient threshold of $r2 > 0.4$ as the criteria of significant linkage. LDpair tool and GRCh38 genome build were employed for the LD computation. The SNP with the stronger association with nephrolithiasis was identified as the index SNP in each LD block. 47 SNPs were found to have reached genome-wide significance. 32 SNPs out of the 47 were identified to share a nearest gene with one or more other SNPs. With previously-described methods, 11 nearest genes and 15 index SNPs were identified being rs10917002 and rs1256328 from ALPL, rs780093 from GCKR, rs56235845 from RGS14, rs1155347 near KCNK5, rs12669187 from MINDY4, rs1000597 near MINDY4, rs7328064 from DGKH, rs578595 from WDR72, rs35747824 from PDILT, rs2079742 near BCAS3, rs13041834 and rs17216707 near CYP24A1, and rs199565725 and rs12626330 from CLDN14. Of note, ALPL, DGKH, WDR72, and CYP24A1 overlap with known monogenic genes. Although there was no SNP that was directly replicated among the 47 genome-wide significant SNPs associated with nephrolithiasis, our study highlights 15 index SNPs that are in LD blocks that have been implicated more than once among these GWAS. These index SNPs may provide perspectives for polygenic loci of nephrolithiasis, and may guide future research for the causes and treatment of nephrolithiasis

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1816 The prefrontal cortex transcriptomic landscape of the comorbidity between post-traumatic stress disorder and opioid misuse

Authors:

J. Martinez-Magana[1], S. T. Nagamatsu[2], D. L. Núñez-Ríos[1], J. Krystal[2], M. Girgenti[2], Traumatic Stress Brain Research Group, J. Montalvo-Ortiz[1]; [1]Yale Univ., Orange, CT, [2]Yale Univ., New Haven, CT

Opioid misuse is highly comorbid with posttraumatic stress disorder (PTSD). Individuals with comorbid opioid misuse and PTSD exhibit higher severity of symptoms, worse disease course, and poorer treatment adherence and response. Understanding the neurobiology of opioid misuse, PTSD, and its comorbidity could help us identify novel targets for interventions and treatments. Our study aimed to apply multiple bioinformatics approaches to comprehensively characterize the transcriptional landscape of opioid misuse, PTSD, and its comorbidity in four human prefrontal cortex (PFC) regions. We found that individuals affected with the comorbidity exhibit a higher number of differentially expressed genes in the PFC than those without the comorbidity. Further, we observed a set of genes that showed a discordant expression profile between those with comorbidity and those without, specifically in the orbitofrontal cortex and subgenual prefrontal cortex. These discordant genes were enriched in early developmental stages and across multiple brain regions. By conducting a single-cell clustering analysis, we found that these genes were enriched in early and adult neuronal developmental cell clusters, specifically immature neurons, Cajal-Retzius, and interneurons, and glutamatergic neurons, respectively. Lastly, using a systems biology approach, we identified a highly co-expressed set of synaptic genes with discordant effects and potential druggable action, mainly pointing to the cholinergic system. Our findings helped disentangle the PFC transcriptomics signatures of opioid misuse, PTSD, and its comorbidity and revealed potential signatures and profiles specific to each trait.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1817 The QGPRS Study: An ensembl learning classifier integrating multi-trait polygenic scores improves risk stratification for T2D in Qatar Biobank cohort

Authors:

I. Ahmed[1], S. Taheri[2], M. Ziab[1], A. Bhat[1], J. Lakshmi[1], H. Mbarek[3], R. Badji[3], O. Chagoury[2], K. Fakhro[1], A. Akil[1]; [1]Sidra Med., Doha, Qatar, [2]Hamad Med. Corp., Doha, Qatar, [3]Qatar Genome, Doha, Qatar

Background: Type 2 diabetes (T2D) is a common disease with multiple genome-wide association studies having reported >500 genetic variants associated with the disease. Translating these associations into specific predictive models has remained a challenge to predict and prevent disease risk. Polygenic scores (PRS) have been shown to provide enhanced power for detecting disease susceptibility and stratifying the high-risk individuals. PRS is an individual-level score that encapsulates the impact of multiple genetic variants on an individual's phenotype. Therefore, the score can be utilized as an outcome predictor in clinical prediction and disease screening programs as it provides an assessment of the entire genetic risk of an individual for a particular trait. When compared to single-trait PRS models, incorporating information from multiple genetically correlated traits can improve prediction accuracy and power for PRS analyses.**Objective:** In this study we aimed to identify genomic variants associated with T2D risk in Qatar Biobank (QBB) cohort and construct machine learning (ML) models with multiple PRS predictors for an enhanced evaluation of the risk.**Methods:** We carried out genome-wide association studies (GWASes) with 12 different T2D related traits to identify genomic variants linked to clinical characteristics of diabetes. A cohort of 14278 QBB participants served as the basis for each GWAS. Multiple weighted PRS models were constructed using the associated set of variants for each trait and combined risk from all 12 predictors was evaluated using ML models.**Results:** Our predictive models displayed a strong ability to stratify cases from controls, and our best model achieved an accuracy of 0.85, area under a receiver operating characteristic curve (AUC) of 0.92, an area under the precision-recall curve (AUC-PR) of 0.8522, and an F1 score of 0.757.**Conclusion:** PRS models have the potential to identify individuals that are at-risk for developing T2D and associated complications in our population. Stratifying people with T2D into groups based on their predisposition for problems could help identify those who would benefit most from an early intervention.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1818 † The Question That Must Be Asked: Is Behavioral Genetics a Null Field?

Authors:

S. Pittelli; El Dorado County, California, Hilton Head, SC

The field of Behavioral Genetics continues to grow as our understanding of DNA has improved, along with our ability to study it. Large amounts of money and resources have been invested in understanding the genetic basis for mental disorders, human intelligence and character traits. To date, however, after over a half century, there has been very little success in discovering causal genetic pathways. Despite this, there seems to be a singular focus on larger and larger genetic studies, with the hope of finding the elusive "missing heritability."To date, there is little discussion of the null. Few question the entire premise of the field of behavioral genetics, that genetic variation among individuals and groups are the primary or at least a significant causal influence. The purpose of this paper is to aggressively question the assumptions of behavioral genetics and to make the case that it is largely a null field, or at least that the influence of genetic variation on behavior is minimally significant. The following questions will be considered: - Have twin studies created a Frankenstein which is now too big to fail?- Are Genome Wide Association Studies (GWAS) likely to end in the same way as candidate gene studies, with optimistic results fading over time and leaving us with more false positive findings, albeit in a more watered down fashion?- Are ever-increasing databases getting us closer to substantial findings or simply creating more heads of a hydra?- Are polygenic scores anything more than a rough measure of population stratification?- Would the findings in GWAS best an actual null?- Does the lack of a standardized statistical approach create a risk of "p-hacking"?- Do the UK Biobank and the other large databases in use allow for a representative sample of a population, or do these databases suffer from their own biases that lead to false positive results, repeated continuously?- Is the current "meta-analysis" approach to GWAS an unconscious attempt to avoid the problem of unreplicated results?- Is the field of behavioral genetics causing more harm than good? A healthy field of study would allow for these types of questions. Currently, there appears to be a consensus mentality, created not by debate or evidence, but by a determination to preserve the field. It's time to question the assumptions of the field of Behavioral Genetics, whether that allows for its advancement or leads to its demise.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1819 The role of tenascin-X in idiopathic gastroparesis: Whole genome sequencing analysis.

Authors:

P. Moszczynski, S. Smieszek, J. Carlin, C. Xiao, G. Birznieks, M. Polymeropoulos; Vanda Pharmaceuticals, Inc., Washington, DC

Tenascin-X (TNX) is a glycoprotein involved in the regulation of tissue structure via anti-adhesive interactions with collagen in the extracellular matrix. Deficiency often manifests as Ehlers-Danlos syndrome (EDS), involving joint hypermobility, skin hyperelasticity, pain and gastrointestinal (GI) dysfunction. TNX gene localizes to the major histocompatibility complex class III region on chromosome 6. TNX deficient mice have accelerated gastric emptying and markedly increased vagal afferent responses to gastric distension. These observations in knockout mice correlated closely with GI symptoms in genetically confirmed TNX-deficient patients with EDS, including disrupted motility and bowel habit and visceral pain.

We previously reported a significantly higher frequency of TNX variants in idiopathic gastroparesis patients as compared with ancestry, sex and age matched controls. TNX enrichment remained significant after Bonferroni correction (*P*-value< 0.001). A 160-fold (p= 8.18743E-26) increase of EDS cases was also reported in this set of gastroparesis patients. All EDS cases were of Caucasian ancestry and females with idiopathic gastroparesis.

In the current study, we replicated this enrichment of TNX variants in the largest whole genome sequencing set of gastroparesis cases (n>1700). We explored the type and frequency of identified TNX variants and focused on all coding/splicing variants in TNX with a minor allele frequency (MAF)<0.05. We see 83 non-synonymous variants with at least 2-fold enrichment. Several of these variants were previously reported in association with EDS in CLINVAR and the functional roles of the remaining ones are to be confirmed. TNX variant, TNX:NM_019105:exon26:c.A9044G:p.K3015R, is identified in our cohort with a high CADD score of 23.3 and is seen in two patients with EDS and idiopathic gastroparesis. We have also identified a subset of patients with copy number variations, specifically with TNX microdeletions. This analysis is ongoing.

These results suggest a statistically significant enrichment of moderate to rare TNX non-synonymous variants in the idiopathic gastroparesis set of patients. Furthermore, they suggest an enrichment of such gene variants in the entire cohort of patients with idiopathic gastroparesis, not only in individuals with confirmed EDS status. TNX appears to be associated with stabilizing the newly produced collagen fibrils, suggesting that all recessive forms of EDS affect post-translational modification of collagens. This finding would not be delineated with a classic GWAS approach yet may explain a proportion of underlying pathophysiology manifesting as gastroparesis.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1820 The role of the genetic influence of dietary preferences on type 2 diabetes and cardiometabolic traits.

Authors:

M. Lee, B. F. Voight; Univ. of Pennsylvania, Philadelphia, PA

Type 2 diabetes mellitus (T2D) is a common glycemic disorder which places patients at elevated risk for cardiovascular co-morbidities and premature mortality. Despite success in genetic studies for T2D, full elucidation of the underlying disease mechanisms is still challenging because susceptibility is also mediated by environmental risk factors (e.g., diet, physical activity), where correlation and causality have not been fully characterized. Our current understanding of the environmental factors underlying T2D risk is driven mainly by observational studies, motivating new approaches to explore this component in more detail. Diet is closely associated with cardiometabolic health, and it is a major environmental risk factor for T2D. Diet however does carry a genetic basis, as shown previously via twin studies and more recently, association studies for macronutrient intake, dietary habits, and food preferences. To characterize the role of diet on susceptibility to T2D, we performed integrated analyses using genome-wide association summary statistics between 42 dietary preference traits, T2D, and 21 cardiometabolic risk factors for T2D. Linkage Disequilibrium Score Regression revealed a shared genetic architecture between diet and T2D, with identification of 289 genetically correlated pairs ($P < 5.41 \times 10^{-5}$). Dietary preferences were largely correlated with adiposity and lipid traits. Pairwise colocalization detected 1,259 loci where diet and T2D share putative causal variants. Univariable mendelian randomization (MR) observed the causal relationships of dietary preferences with T2D, but multivariable MR suggested that many of these associations are potentially explained by cardiometabolic and/or educational factors. Overall, our findings elucidate the genetic relationship between diet, anthropometry, and liability to T2D.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1821 The roles of genetic susceptibility and sedentary behavior in the development of all-cause dementia in the UK Biobank

Authors:

Y. Klimentidis; Univ. of Arizona, Tucson, AZ

Background: Recent work suggests that sedentary behavior is a risk factor for late-onset Alzheimer's disease (LOAD) and related dementias. However, it is not clear how this relationship differs according to an individual's level of genetic susceptibility. **Aim:** We used polygenic risk scores (PRS) of LOAD and of all-cause dementia in addition to a measure of time spent watching television to assess associations with incident all-cause dementia in the UK Biobank. **Methods:** We used GWAS of LOAD (n_{case}=21,982, $n_{control}$=41,944) from a large-scale meta-analysis, and GWAS of all cause dementia from FinnGen (n_{case}=12,042, $n_{control}$=254,976) along with LDpred software to generate a PRS for each disease outcome among UK Biobank participants >60 years old. We used self-reported time spent watching television as a measure of sedentary behavior. We examined the genetic correlation between the LOAD and the all-cause dementia GWAS. Associations of each PRS and time spent watching television with incident all-cause dementia (n=3,507) in the UK Biobank were assessed using Cox proportional hazards models. Interactions of time spent watching television with each PRS were also assessed. **Results:** The genetic correlation between LOAD and all-cause dementia was relatively strong (r^2=0.65). In Cox proportional hazards models, both the PRS for LOAD and the PRS for all-cause dementia were strongly associated with all-cause dementia. We found an association of time spent watching television with incident all-cause dementia, however this association did not differ on the basis of genetic risk level. **Conclusion:** We find that both the PRS of LOAD and the PRS of all-cause dementia PRS are strongly associated with all-cause dementia in the UK Biobank, and that time spent watching television is associated with greater risk of all-cause dementia regardless of an individual's level of genetic risk.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1822 The rs4728142 of *IRF5* as potential novel predictor of radiographic joint damage in rheumatoid arthritis.

Authors:

S. Ali[1], M. Hanif[1], F. Younas[1], A. Usmani[1], L. Hayat[1], W. Mateen[1], A. Hamza[1], J. M. Malik[2]; [1]COMSATS Univ. Islamabad, Islamabad, Pakistan, [2]Rehmat Noor Clinic, Rawalpindi, Pakistan

Rheumatoid arthritis (RA) is a chronic autoimmune condition causing progressive joint degradation and functional disability. The disease has a complex etiology with involvement of different genetic and environmental risk factors. We conducted a case-control association study of RA including 850 participants (450 RA cases and 400 controls). Cases were clinically diagnosed RA patients; while controls were healthy individuals who were matched with cases in age, sex and ethnicity. All participants gave informed written consent. Genotyping was carried out through PCR-based methods for 15 genetic variants from five genes. Multivariate logistic regression analysis was used to determine the association of genetic variants with RA. A protective effect of rs6920220 on overall RA susceptibility was observed [A vs. G allele odds ratio (OR) = 0.23, 95% confidence interval (CI) = 0.18-0.29]. The rs4728142 exhibited association with increased overall RA susceptibility (G vs. A allele OR = 2.3, 95%CI = 1.89-2.8). Stratified analysis by clinical characteristics showed a significant association of rs4728142 with greater risk of radiographic joint damage among RA patients (G vs. A allele OR = 4.77, 95%CI = 3.15-7.22). In conclusion, we report rs4728142 of *IRF5* gene as a novel putative predictor of radiographic damage of joints in RA.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1824 The shared and divergent target genes, pathways and relevant cell types for 15 autoimmune diseases from genome-wide association studies—using IL12/IL23 pathway genes as an example

Authors:

W. Yang[1], X. Dang[1], Y-F. Wang[2]; [1]Univ. of Hong Kong, Hong Kong, China, [2]Chinese Univ. of Hong Kong (Shenzhen), Shenzhen, China

Genome-wide association studies (GWAS) have revealed more than 600 associated loci for various autoimmune diseases. However, for most of these loci, the association mechanism(s) are unclear, including the target gene(s). Lack of functional understanding is the major challenge for translating genetic findings into clinical applications. In this study, we compared the association signals for 15 most common autoimmune diseases, using linkage disequilibrium and conditional analyses to identify the shared and specific association signals. We annotated these signals using genomic data such as eQTL, histone marks, and promoter interactions. Relevant cell types and signaling networks are constructed and compared among these diseases. We identified 620 genomic loci, 353 of which are shared by at least two autoimmune diseases, suggesting shared target genes in most of these cases. However, we also identified widespread independent association signals in most of these loci. There are a total 1586 association signals for the 15 autoimmune diseases, and only a small minority of these signals (325, 20.5%) are shared by at least two diseases, emphasizing different molecular mechanisms for the associations among the autoimmune diseases (differences in target genes, regulatory mechanisms, cell types and other cellular contexts). To use the locus of *IL23R* and *IL12RB2* and other genes in the IL12 and IL23 pathway as an example, it is clear that *IL23R* is associated with psoriasis (PS), inflammatory bowel disease (IBD), ankylosing spondylitis (AS), while *IL12RB2* is associated with rheumatoid arthritis (RA), systemic lupus erythematosus (SLE), and primary biliary cirrhosis (PBC). Analysis of eQTL data on the associated genetic variants reveals that Tfh, Th17 and memory CD4 T cells are the most relevant cell types where *IL23R* expression is regulated, and NK cells and Th1 cells are the most relevant cell types for regulating *IL12RB2* expression. Similarly, *IL12A* (encoding for P35) is associated with PBC, SLE and multiple sclerosis (MS) and the eQTLs are mostly B cell-specific, while *IL12B* (encoding for P40) is associated with PS, IBD and AS. *IL12RB1* is found to be associated with Systemic Sclerosis (SSc), whose expression eQTLs are ubiquitous in various immune cells. The different significance of IL12 signaling and IL23 signaling for different autoimmune diseases may help with designing new clinical trials and precision treatment of these diseases.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1825 The shared genetic architecture of inflammatory bowel disease with clinical, psychiatric, diet, and behavioral traits

Authors:

V. Shaw, J. Byun, R. Pettit, Y. Han, C. Amos; Baylor Coll. of Med., Houston, TX

Background: Inflammatory bowel disease (IBD) development is a complex, multifactorial process that involves extrinsic and intrinsic factors, such as host genetics, the immune system, the gut microbiome, and environmental risks. To help understand the genetic contribution of clinical, behavioral, psychiatric, and diet-related traits, we aim to provide a deeper and comprehensive characterization of the shared genetic architecture between IBD and hundreds of potentially related traits by using GWAS summary statistic data and a regression statistical framework known as cross-trait linkage disequilibrium score regression (LDSR).
Methods: Utilizing publicly available summary statistics from previously published IBD GWAS7 and hundreds of traits from the United Kingdom BioBank (UKBB), we perform LDSR analysis to estimate cross-trait genetic correlations between Crohn's disease (CD), ulcerative colitis (UC), and IBD summary statistics and the UKBB traits of interest.
Results: Nominally significant ($p < 0.05$) genetic correlations were observed for 181 traits in overall IBD, 239 traits in CD, and 94 traits in UC. We replicate the known association between smoking behavior and CD/UC, namely that current tobacco smoking has a positive genetic correlation with CD ($rg = 0.12$, $p < 0.01$) while "ever smoking" has a negative genetic correlation with UC ($rg = -0.07$, $p = 0.042$). Education-related traits, such as professional qualifications (e.g., nursing, teaching) or obtaining a college/university degree, were negatively genetically correlated with IBD, respectively ($rg = -0.14$, $p < 0.01$; $rg = -0.11$, $p < 0.01$). Globally, all three strata (IBD, CD, and UC) demonstrated increased genetic correlations for psychiatric-related traits related to anxiety and depression.
Conclusions: The present analysis reveals the shared genetic architecture between dozens of traits and IBD, CD, and UC. Future work should focus on establishing directionality of the traits and utilize the traits to help decompose the genetic heterogeneity of IBD.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1826 Tissue-specific alterations in alternative splicing associated with aging.

Authors:

K. Zimmerman[1], S. Puppala[1], J. Chan[1], P. Nathanielsz[2], L. Cox[1], M. Olivier[1]; [1]Wake Forest Sch. of Med., Winston-Salem, NC, [2]Univ. of Wyoming, Laramie, WY

Alternative splicing is a complex cellular process that enables production of a diverse set of RNA products from a single gene. Using comprehensive RNA-Seq data on four different tissues from a cohort of healthy baboons that span the adult age range, differential expression analysis and weighted gene co-expression network analysis (WGCNA) was computed to identify genes significantly associated with age. In the liver, we identified a large age-associated WGCNA module ($r=0.46$, $p=0.002$) that was significantly enriched for mRNA-splicing ($p=1.7x10^{-14}$). With the differentially expressed genes, we also identified networks enriched for the spliceosome in the brain and heart. In the liver, brain, and heart we also identified significant associations between the numbers of genes expressing the maximum number of isoforms and age. In the liver, isoform numbers decreased with age ($\beta=-4.9$, $p=6.7x10^{-4}$), while numbers were not significantly different in muscle ($\beta=-2.3$, $p=0.75$) and significantly increased with age in the brain ($\beta=17.1$, $p<2.0x10^{-16}$) and heart ($\beta=5.8$, $p=1.9x10^{-3}$). These results reveal splicing-related changes associated with age, even in a model of healthy aging. This work demonstrates that splicing alterations are tissue-specific and appear to be particularly amplified in the brain with aging. Further work is needed to establish the associations identified here and whether they are tissue-specific. It will also be important to explore how these changes interact with the environment, how they occur at the cellular level, and whether RNA-splicing is actually an adaptive mechanism that cells deploy in response to a challenge.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1827 *TMEM106B* and *APOE* increase risk for TDP-43 and HS pathologies underlying Alzheimer's Disease Related Dementias

Authors:

D. Godrich[1], J. Pasteris[1], T. Phongpreecha[2], E. Martin[1], A. Naj[3], R. Mayeux[4], P. Crane[5], D. Bennett[6], L. Molina-Porcel[7], G. Schellenberg[3], M. Pericak-Vance[1], M. Cuccaro[8], W. Scott[1], T. Rundek[1], W. Kukull[5], T. Montine[2], G. Beecham[1], the Alzheimer's Disease Genetics Consortium (ADGC); [1]Univ. of Miami Miller Sch. of Med., Miami, FL, [2]Stanford Univ., Palo Alto, CA, [3]Univ. of Pennsylvania, Philadelphia, PA, [4]Columbia Univ, New York, NY, [5]Univ. of Washington, Seattle, WA, [6]Rush Univ. Med. Ctr., Chicago, IL, [7]Inst. de Investigaciones Biomédicas August Pi i Sunyer, Barcelona, Spain, [8]John P. Hussman Inst. for Human Genomics, Miami, FL

Alzheimer disease (AD) is pathologically characterized by plaques and tangles. However, pathologies from other AD-related dementias, such as hippocampal sclerosis (HS) and Transactive response DNA-binding protein of 43 kDa (TDP-43) proteinopathy, commonly co-occur. Studying the genetics of pathologies underlying ADRD can provide a better understanding of disease. The role of *APOE* in HS and TDP-43 proteinopathy remains unclear, and previous studies suggested association of *TMEM106B* (rs1990622) with TDP-43 proteinopathy and HS. Here, we investigate the interplay between TDP-43 proteinopathy, HS, cognitive decline, *APOE,* and *TMEM106B*. Data come from 898 autopsied individuals from the National Alzheimer's Coordinating Centers (NACC), 511 of which also have genotype data. Logistic regression was used to model and compare the independent and interactive effects of HS and TDP-43 proteinopathy on cognitive decline (change in MMSE), adjusting for sex and age at death. We also examined the relationship between genetic factors (*APOE* and *TMEM106B*) and TDP-43 proteinopathy and HS, comparing models with individual and joint effects. Mediation analyses were performed to explore direct and indirect effects. The majority (N = 468, 52%) of participants had neither TDP-43 proteinopathy nor HS, while a subset had TDP-43 proteinopathy only (N=230, 26%), HS only (N=40, 4%), or both TDP-43 proteinopathy and HS (N=160, 18%). Regression models showed significant independent effects of both TDP-43 proteinopathy and HS on cognitive decline, but the interaction was not significant. Both loci showed incomplete dominance (*TMEM106B*: heterozygote OR = 2.1, p = 5E-3; alternate homozygote OR = 2.8, p = 2E-4; *APOE* ***ε4*** heterozygotes OR = 2.4, p = 7E-6; ***ε4*** homozygotes OR = 2.6, p = 4E-3). HS showed similar effects for *TMEM106B* (het OR = 3.2, p = 4E-3, hom OR = 3.6, p = 2E-3). *APOE4* was not associated with HS. Mediation analyses revealed direct and mediation effects of TDP-43 proteinopathy on cognitive decline, while HS did not. *APOE* had a direct effect on TDP-43 proteinopathy, and its effect on HS was significant only when mediated by TDP-43 proteinopathy. The same was seen with *TMEM106B*. The associations with *APOE* suggest a common risk factor for the AD and ADRD pathologic changes. However, mediation models suggested that TDP-43 proteinopathy may be the main driver of effect on cognitive decline both independently and together with HS. These findings also suggest that *TMEM106B*, while significantly associated with both TDP-43 proteinopathy and HS, may only affect HS via TDP-43 proteinopathy mediation. Our study also suggests that TMEM106B may follow an incomplete dominance mode of inheritance.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1830 Tradeoff between prediction accuracy and transferability in the design of polygenic risk scores of bipolar disorder.

Authors:

A. Dominguez[1], Y. Zhang[1], C. Pato[2], S. McCarroll[3], M. McInnis[4], M. Boehnke[1], L. Scott[1], S. Zollner[1,4]; [1]Dept. of Biostatistics, Univ. of Michigan, Ann Arbor, MI, [2]Dept. of Psychiatry, Rutgers BioMed. and Hlth.Sci., Newark, NJ, [3]Harvard Med. Sch., Boston, MA, [4]Dept. of Psychiatry, Univ. of Michigan, Ann Arbor, MI

Polygenic risk scores (PRS) use evidence from genome-wide association studies (GWAS) to predict genetic risk for complex traits. However, most GWAS participants are European (EUR), making prediction for non-EUR samples less accurate. Factors like allele frequency differences and linkage disequilibrium can lead to differences in mean PRS across samples. Our goal is to investigate how varying the number of markers in PRS of bipolar disorder (BPD) impact prediction and mean PRS across populations.

Using clumping & thresholding method and majority EUR summary statistics from the Psychiatric Genomics Consortium (PGC), we calculate and assess predictive accuracy of PRS at 13 p-value thresholds for 3,334 African (AFR) Americans from the InPSYght study and 5,670 individuals with majority EUR ancestry from Prechter Bipolar Study and Michigan Genome Initiative (MGI). Additionally, we calculate PRS for 2,504 individuals from 1000 Genomes Project (1KGP) to assess differences in mean PRS across populations. We use principal component analysis to obtain measures of ancestry and control for population stratification.

To assess predictive performance of PRS for case-control samples (2,010 cases/5,994 controls), we use logistic regression to model case status adjusting for sex and five principal components (PCs). Predictive accuracy increases with more markers in PRS but plateaus at $p < 10^{-3}$ (R^2=6.4%). Moreover, the difference in mean PRS between AFR and EUR samples increases with more markers. To assess ancestry impact on PRS, we use regression with PRS as outcome and PCs as the predictor adjusting for sex and affection status. For thresholds $p < 10^{-3}$ and lower, R^2 stays below 8%. At larger thresholds, PCs explain more variance in PRS with a maximum at $p < 0.9$ (R^2 =57.6%). Although more markers in PRS improve prediction, PCs become more associated with PRS.

To compare results, we compute PRS for four additional traits: asthma, height, schizophrenia, and type 2 diabetes. Across traits, as we include more markers, differences in mean PRS across populations and correlations between PRS and PC1 increase. Using simulations, we assess whether correlation between PRS and PC1 is expected under population differentiation. The correlations for our simulated PRS and PC1 were centered at zero regardless of how many markers used, suggesting correlations between our observed PRS and PC1 are not expected under population differentiation.

In this study, including more markers into PRS improves prediction, but also accrues subtle confounding from the original GWAS. There is a tension between having complex, informative PRS with more markers and their susceptibility to population structure.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1831 Trans-ancestry GWAS meta-analysis of peripheral artery disease uncover 89 loci and post-GWAS analyses map the signals to candidate effector genes and disease mechanisms.

Authors:

H. Kim, E. B. Fauman, J. Chiou, FinnGen, UKB-PPP, M. R. Miller; Pfizer, Cambridge, MA

Peripheral artery disease (PAD) is a common cardiovascular morbidity that affects more than 230 million adults worldwide and has limited treatment option leading to high unmet medical need. Compared to other major cardiovascular diseases, existing GWAS of PAD have been relatively small in sample size and identified a limited number of associated loci. In this study, we conducted the largest meta-analysis of PAD GWAS to date, increasing the number of cases by more than 2-fold and maximizing statistical power for discovery.Our study incorporated published GWAS from Million Veteran Program (MVP), Genetics of Lower Extremity Arterial Disease (GoLEAD), and Biobank Japan (BBJ) and a new GWAS from FinnGen, including a total of 64,726 cases and 1,216,879 controls. We first performed fixed-effect meta-analysis among the European ancestry, followed by fixed- and random-effect meta-analyses across different ancestries. This trans-ancestry meta-analysis uncovered 89 loci, of which 69 have not been reported before. Finemapping of the loci revealed a total of 106 independent signals.To assess whether the PAD genetic signals are mediated by the known risk factors of PAD, we ran colocalization analyses against GWAS of lipids, blood pressure, body mass index, waist-to-hip ratio adjusted for BMI, type 2 diabetes, smoking, and C-reactive protein. Of the 106 signals, 50 were colocalized with at least one risk factor with consistent effect direction, suggesting the underlying biological mechanisms for those signals. Many signals did not colocalize with any of the known risk factors, suggesting that there may be additional unknown mechanisms that mediate PAD pathogenesis.To find the genes that likely mediate the signals, we used the following heuristics: 1) whether there is only a single protein-coding gene near the signal, 2) colocalization with protein QTLs generated by UKB-PPP (pharma proteomics project), 3) presence of protein sequence altering variants in close LD with the lead variant, and 4) manual review by experts. Using this approach, 64 signals could be mapped to the likely effector gene(s).Future work will focus on integrating PAD GWAS result with omics datasets generated from tissues relevant to PAD to 1) prioritize genes based on their transcriptomic profile, 2) refine signal-to-gene mapping using multi-omics signature, 3) inform cell and tissue type of action, and 4) identify cell and tissue types that are enriched for PAD genetic signals. In summary, this work extends our knowledge on the genetic underpinnings of PAD and suggests biological mechanisms that may underlie the PAD genetic signals.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1832 Trans-ancestry PRSs of lipid traits: quantification and contextualization in large cohorts and biobanks.

Authors:

Y. Cai[1], Z. Wang[2], M. Kim[1], R. Smit[2], S. Graham[3], Y. Hu[1], S-A. Love[4], J. Lundin[1], S. Buyske[5], K. Young[4], C. Willer[6], K. North[4], C. Kooperberg[1], R. J. F. Loos[7], U. Peters[1], M. Graff[4], The Global Lipids Genetics Consortium, Population Architecture through Genomics and Environment; [1]Fred Hutchinson Cancer Ctr., Seattle, WA, [2]Icahn Sch. of Med. at Mount Sinai, New York, NY, [3]Univ. of Michigan, Ann Arbor, MI, [4]The Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, [5]Rutgers Univ, Piscataway, NJ, [6]Regeneron, Tarrytown, NY, [7]Univ. of Copenhagen, Copenhagen, Denmark

Lipid profiles are heritable risk factors of cardiovascular disease (CVD), a leading cause of global mortality. Polygenic risk scores (PRSs) for lipids can help predict early CVD risk, however, their portability in non-European ancestry individuals may be limited. In addition, effectiveness of lipid PRSs can be influenced by various factors including participant demographics, medication usage and disease status. Therefore, comprehensive evaluation of lipid PRSs in diverse populations, accounting for demographic and clinical characteristics, is crucial for accurate and equitable clinical application.

We applied trans-ancestry PRSs for lipid traits developed by the Global Lipids Genetics Consortium (GLGC) to the Population Architecture through Genomics and Environment (PAGE) study with 71K diverse participants. We evaluated the PRSs for lipid levels using incremental R^2 stratified by self-reported race and ethnicity groups. As expected, the PRSs' performed best in the European population. PRS performed poorly in populations with low presence in GLGC GWAS.

We also conducted stratified analyses of PRS performance on various factors (e.g., age, sex, blood pressure, lipids-lowering medication usage, smoking status, diabetes status), while adjusting for self-reported race and ethnic groups. A z-test method was applied to determine if R^2 was significantly different between strata. Taking LDL as an example, the stratified analyses showed significantly better PRS performance in women. Furthermore, R^2 was significantly higher in participants not using lipids-lowering medication (statins, fibrates, etc.) or not having type-II diabetes. No significant difference in PRS performance was observed based on age and smoking status. We further tested these factors separately within each race and ethnic group to explore potential differences, and overall, these PRS performance characteristics were consistent across populations.

Preliminary replication of our PAGE analyses in six additional diverse studies, which included 69k individuals of non-EUR ancestry, show consistency with our PAGE results, although the current sample size in some race and ethnic groups may impose certain limitations. Overall, we found that PRS performance varied across populations. Stratified analyses revealed differences based on demographic factors, medication usage and disease status, such as higher PRS performance in women and those without lipid-lowering medication or type-II diabetes. Collaborative efforts with other studies will enhance the robustness and reproducibility of our results.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1833 Transcriptional and epigenetic profiles of CD8+ T lymphocytes in multiple sclerosis.

Authors:

E. Paraboschi[1], F. Airi[2], G. Liberatore[3], V. Rimoldi[4], E. Nobile-Orazio[3], R. Asselta[4]; [1]Humanitas Univ., Humanitas Res. Hosp., Pieve Emanuele, Italy, [2]Humanitas Univ., Pieve Emanuele, Italy, [3]IRCCS Humanitas Res. Hosp., Rozzano, Italy, [4]Humanitas Univ., IRCCS Humanitas Res. Hosp., Pieve Emanuele, Italy

Alternative splicing (AS) is a post-transcriptional mechanism that increases the information content of the transcriptome. AS is strictly connected to the backsplicing (BS) process, a mechanism that leads to the generation of circular RNA products (circRNAs). AS and BS rely on the same spliceosomal machinery and trans-acting factors and may be tuned by genetic and epigenetic determinants, as DNA methylation. A dysregulation of AS/BS has been implicated in human disorders, including multiple sclerosis (MS), a chronic autoimmune neurodegenerative disease in which autoreactive T lymphocytes attack antigens of the central nervous system.
The aim of the project is to unravel the mechanisms involved in the dysregulation of AS/BS processes in MS CD8+ T lymphocytes by integrating different omics data.
We collected data on gene expression, AS profile, circRNA landscape, and methylation status of CD8+ T lymphocytes of 20 MS patients and 20 controls, using massive sequencing and array approaches. A first analysis evidenced the presence of several differentially expressed genes (210, FDR<0.1), circRNAs (200, P<0.05) and AS events (600 genes carrying at least one significantly dysregulated AS event, FDR<0.05), as well as of differentially methylated positions (DMPs, 1200 genes, FDR<0.05) in MS cases. Among the top genes, we observed an upregulation of *MALAT1* in MS patients (1.3-fold) that was exacerbated in males. This result was particularly interesting, since *MALAT1* is involved in transcriptional/post-transcriptional regulation of gene expression, including AS and BS. Moreover, we also observed a strong upregulation of *MALAT1*-derived circRNAs (4-fold increase) in males. Regarding the DMPs, we again noted that *MALAT1* was among the top differentially methylated genes, with patients showing a significant demethylation compared to controls (logFC=-0.34). This result suggests a mechanism guiding *MALAT1* upregulation which, in turn, could be the cause of the global dysregulation of the AS and BS landscapes.
Next, all the omics were integrated in the unified framework of MOFA2 that infers a low-dimensional data representation in terms of factors which can be analyzed. We identified a factor that was associated with the disease status ($P<10^{-5}$). The top features of each omic layer were then used to run an enrichment analysis that highlighted, among the top results, “T cell receptor signaling pathway”, “cellular response to stress”, and “TNF signaling pathway”. We plan to implement in the framework also the genetic information on MS-associated SNPs.
Acknowledgments. This work is supported by Fondazione Regionale per la Ricerca Biomedica (FRRB), Early Career Award.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1834 Transcriptome Study of Metabolically Healthy Obesity in Two Cohorts of African American

Authors:

A. Gaye; NIH/NIHGRI, Bethesda, MD

Background: Some obese individuals, named metabolically healthy obese (MHO), do not develop known metabolic consequences of obesity such as insulin resistance, type 2 diabetes and dyslipidemia. Previously, we defined a cluster of co-expressed mRNAs downregulated in MHO subjects that may underlie the activation of endoplasmic reticulum (ER) and related-stress pathways that lead to a chronic inflammatory state. This follow-up project aims to investigate microRNAs (miRs) that modulate mRNAs related to MHO and establish their relationship with key pathways linked to ER-stress revealed in our initial work. **Methods:** The data consist of whole blood microRNA sequencing data of respectively 75 and 210 samples from the MHGRID and GENE-FORECAST cohorts. In both datasets, Random Forests (RF) was used to identify miRs that predict MHO and other cardiometabolic traits and WGCNA was applied to define clusters of co-expressed miRs associated with MHO and related traits. An in-vitro model was designed with high fat (palmitic acid) treated HepG2 cells to determine if the miRs lead to an alteration of biological functions related to ER stress, inflammation and autophagy. **Results:** RF identified 96 miRs predictive of MHO and 11 cardiometabolic traits in MHGRID and GENE-FORECAST. 5 clusters of co-expressed miRs were correlated with MHO status, fasting glucose, BMI, hypertension status and TG/HDL. 129 pathways relevant for obesity, ER-stress and insulin resistance were significantly enriched in the RF and WGCNA results. 48 hours after transfecting the HepG2 cells with a set of miRs that have experimentally validated mRNA targets, significant changes in expression were observed in a number of genes involved in ER-stress and autophagy related pathways. We also observed significant changes in the level of Bip, XBP1, ATF6 and CHOP proteins, compared with controls. **Conclusions:** We undertook an unbiased transcriptome-wide analysis of metabolically healthy obesity in 2 independent datasets of African American subjects and the results confirmed our earlier findings implicating ER-stress. The results of the in-vitro model suggest that the miRs associated with MHO we identified induce changes in biological functions related to ER-Stress in HepG2 Cells.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1835 Transcriptome-wide association study for frequent cocaine use identifies gene dysregulation in CD4+ T cells linked to immune response and cancer pathways.

Authors:

B. Quach[1], S. N. Giamberardino[1], C. Willis[1], K. Weitzel[1], D. Brannock[1], D. B. Hancock[1], D. Konkle-Parker[2], H. H. Bolivar[3], C. D. Lahiri[4], E. F. Topper[5], M. H. Cohen[6], S. G. Kassaye[7], J. DeHovitz[8], M. H. Kuniholm[9], N. M. Archin[10], P. C. Tien[11,12], K. Xu[13,14], B. E. Aouizerat[15], E. O. Johnson[1]; [1]RTI Intl., Durham, NC, [2]Univ. of Mississippi Med. Ctr., Jackson, MS, [3]Univ. of Miami-ACRU, Miami, FL, [4]Emory Univ., Atlanta, GA, [5]Johns Hopkins Univ., Baltimore, MD, [6]Cook County Hlth.System, Chicago, IL, [7]Georgetown Univ., Washington, DC, [8]Downstate Hlth.Sci. Univ., Brooklyn, NY, [9]Univ. at Albany, State Univ. of New York, Rensselaer, NY, [10]Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, [11]Univ. of California, San Francisco, San Francisco, CA, [12]Dept. of VA Med. Ctr., San Francisco, CA, [13]Yale Univ., New Haven, CT, [14]Veteran's Affairs Connecticut Hlth.care System, West Haven, CT, [15]New York Univ., New York, NY

Background: In 2021, 4.8 million (1.7%) people in the U.S. reported using cocaine. Cocaine use affects immune cells and increases susceptibility to infectious diseases, such as HIV, and contributes to the development of health conditions such as cardiovascular disease and chronic kidney disease. Existing transcriptomics research that investigates the effects of cocaine has focused primarily on addiction and brain tissue. There is a lack of transcriptome-wide studies evaluating the modulation of immune cell gene expression by cocaine in blood and how it relates to disease pathogenesis.
Methods: We measured gene expression using paired-end RNA-seq in CD4+ T cells isolated from peripheral blood samples, among individuals with and without cocaine use, to characterize transcriptomic changes by cocaine use status. Samples were from participants enrolled in the Women's Interagency HIV Study with self-reported frequent cocaine use (at least once/week or greater, N=64) or no cocaine use (N=195) within 6 months prior to blood draw. All samples were from women living with HIV who had an undetectable plasma HIV RNA level for at least 2 consecutive study visits. Individuals in the frequent cocaine use and no cocaine use groups had similar age distributions (median ages of 47.3 and 47.6, respectively). Transcriptome-wide differential expression testing was conducted using negative binomial regression with gene expression as the outcome variable and cocaine use status and self-reported race as independent variables. Gene set overrepresentation analysis was conducted for the significantly differentially expressed genes (DEGs) using a hypergeometric test.
Results and discussion: We identified 824 DEGs (FDR<0.05). Of these, a significantly higher proportion showed increased expression in the frequent cocaine use group (624 DEGs; Binomial test p<2.2e-16). DEGs were enriched in Gene Ontology and Molecular Signatures Database biological pathway terms related to immune cell migration, T cell adhesion, T cell activation, and vascular injury and repair. DEGs were also overrepresented in cancer-related signaling pathways, including EGFR signaling, PI3K signaling, and oncogenic MAPK signaling (FDR<0.05). Further, 382 DEGs have an eQTL based on queries against CD4+ T cell eQTL maps in the Database of Immune Cell eQTLs, Expression, and Epigenomics, suggesting that genetic variation may modulate gene expression for almost half of the frequent cocaine use dysregulated genes. Together, these results highlight potential genes and pathways that may lead to altered immune function and health outcomes in women as a result of cocaine exposure.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1836 Transcriptomic analysis of multiple tissues in obesity in Africans

Authors:

G. Chen[1], A. Doumatey[2], J. Zhou[2], L. Lei[2], A. Adeyemo[1], C. Rotimi[3]; [1]NIH, Bethesda, MD, [2]NIH, Bethersda, MD, [3]NIH, Woodstock, MD

Obesity is a complex disease associated with an increased risk of cardiometabolic disorders. Its prevalence has increased globally making it a major public health problem. While many genetic variants associated with obesity risk have been identified, studies of associated tissue and related molecular changes are scarce outside Europe, North America, and Asia. The present study was designed to identify differentially expressed genes (DEG) in obesity in a study of Nigerians enrolled in the Africa America Diabetes Mellitus (AADM) Study. Primary tissues were obtained from participants with obesity (BMI>=30 kg/m^2) and without obesity (BMI<30 kg/m^2) for RNA-seq analysis and included whole blood samples (n=72), subcutaneous adipose tissue (n=50), and skeletal muscle (n=57) - obtained via percutaneous needle biopsy of the vastus lateralis muscle. Samples were sequenced on the Illumina Novaseq 6000 platform and after quality control, 57,821 transcripts mapping to the human reference genome build-37 (hg19) were analyzed. DEG, defined as absolute fold-change >=2 and FDR < 0.05, were identified using a negative binomial generalized linear model (*DESeq2*) and adjusting for age, sex and type 2 diabetes. Enrichment analysis was conducted using Clarivate *Metacore*™. We identified 195, 104, and 8 DEGs in blood, adipose tissue, and skeletal muscle, respectively. While there was little overlap between DEGs identified in the three tissues, the pathways enriched in those of blood and adipose tissue showed more overlap. The *phosphatidylinositol-4,5-diphosphate pathway*, and *signal transduction_Angiotensin II/AGTR1 signaling pathway* were both enriched in blood and adipose tissue. DEGs in adipose tissue were uniquely enriched for several immune response (e.g. IFN-gamma in macrophage activation, *IL-12 signaling pathway*), signal transduction (e.g. *Soluble CXCL-16 signaling*, *FGFR2 signaling*) and cell cycle processes (e.g. *The metaphase checkpoint*, *Initiation of mitosis*). DEGs in blood were enriched for a more heterogenous range of pathways that included *Beta-adrenergic receptors signaling via Cyclic AMP* and *NOTCH signaling*. These findings indicate that obesity is associated with different transcriptomic changes that is tissue dependent, with adipose tissue showing signatures of cellular expansion and differentiation (indicative of adipocyte hypertrophy and differentiation) as well as an immune cell response (consistent with the well-known role of adipose tissue macrophages in obesity-related inflammation). Our findings add to the slowly growing literature on omics profiling to inform cardiometabolic disease pathophysiology in diverse populations.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1837 Transcriptomic analysis reveals the dynamic changes of immune responses during the development of creeping fat in Crohn's disease.

Authors:

S. Park[1], K. Kim[2], Y. Lee[1], J. Baek[1], Y. Kim[1], S. Hwang[2,3], J. Lee[3,4], S. Park[2,3], S-K. Yang[2,3], B. Han[5], K. Song[1], Y. Yoon[3,4], B. Ye[2,3], H-S. Lee[1]; [1]Dept. of Biochemistry and Molecular Biology, Univ. of Ulsan Coll. of Med., Asan Med. Ctr., Seoul, Korea, Republic of, [2]Dept. of Gastroenterology, Univ. of Ulasn Coll. of Med., Asan Med. Ctr., Seoul, Korea, Republic of, [3]Inflammatory Bowel Disease Ctr., Seoul, Korea, Republic of, [4]Dept. of Surgery, Univ. of Ulsan Coll. of Med., Asan Med. Ctr., Seoul, Korea, Republic of, [5]BK21 Plus BioMed. Sci. Project, Seoul Natiopnal Univ. Ceollege of Med., Seoul, Korea, Republic of

Creeping fat (CF) is a poorly understood feature of Crohn's disease (CD), characterized by the wrapping of mesenteric adipose tissue (MAT) around the inflamed intestine. The aim of this study was to investigate the transcriptional profile and compositional features of CF. We collected 59 MAT samples: 23 paired samples from patients with CD (CF [CD-CF] and MAT around the uninflamed intestine [CD-MAT]) and 13 MAT samples from non-CD patients (Con-MAT). Differentially expressed gene (DEG), functional pathway, cell deconvolution, and gene co-expression network analyses were performed. We identified a total of 529 DEGs ($|\log_2 \text{FoldChange}|>1.5$; false discovery rate <0.05), of which 428 exhibited transitional transcriptomic changes. Of these, 323 genes showed an incremental pattern from Con-MAT to CD-MAT, and to CD-CF, while 105 genes displayed a decremental pattern. Genes with an incremental pattern were related to immune cell responses, including B-cell and T-cell activation, while genes with a decremental pattern were involved in cell trafficking and migration. Cell deconvolution analysis revealed significant changes in cellular composition between the CD-CF and Con-MAT groups, with increased proportions of B-cells/plasma cells (p-value=1.16×10^{-4}), T-cells (p-value=3.66×10^{-3}), and mononuclear phagocytes (p-value=3.53×10^{-2}) in the CD-CF group. In contrast, only the B-cell/plasma cell component showed a significant increase (p-value=1.62×10^{-2}) in the CD-MAT group compared to Con-MAT. To conclude, dynamic transcriptional changes and altered cellular components may be crucial for the shift from Con-MAT to CD-MAT and to CD-CF, providing insight into the mechanisms behind CF and highlighting its possible role in CD pathogenesis.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1838 Transcriptomic Dysfunction Disparities: Greater Burden in Female Brainshighlights Immune and Synaptic Pathways for psychiatric disorders

Authors:

Y. Xia[1], C. Xia[2], Y. Jiang[3], Y. Chen[4], R. Dai[5], C. Liu[5], C. Chen[2]; [1]Yale Univ., New Heavon, CT, [2]Central South Univ., Changsha, China, [3]Vanderbilt Univ., Nashville, TN, [4]Central South Univ., Cambridge, MA, [5]SUNY Upstate Med. Univ., Syracuse, NY

- One Sentence Summary: Sex differences in psychiatric disorders:greater burden of transcriptomic dysfunction in females highlights immune and synaptic pathways.
- Many psychiatric disorders exhibit strong sex differences, but the underlying mechanisms remain poorly understood. We analyzed transcriptome data from 2160 postmortem brain samples from the PsychENCODE project in a sex-stratified design, comparing schizophrenia (SCZ), bipolar disorder (BD), and autism spectrum disorder (ASD) with controls. We found that females had a higher burden of transcriptomic dysfunction than males, evidenced by a larger number of differentially expressed genes(based on the same sample size)and a greater magnitude of expression level changes. Additionally, female patients had greater overall connectivity dysfunction, a higher proportion of modules with significant connectivity changes, and a higher proportion of modules with higher connectivity burden, indicating more severe organization disruption of co-expression in females. We identified several modules enriched with DEGs, burden genes, and genome-wide association study genes involved in immune and synaptic functions across different brain cell types. Hub genes, including SCN2A, FGF14, and C3, were highlighted in those modules. Our results support the hypothesis that the burden of transcriptomic dysfunction differs between sexes in psychiatric disorders and highlight the importance of immune and synaptic-related pathways in sex differences.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1839 Transcriptomic profiles of corneal epithelium reflect irregularities in epithelial thickness maps of post-refractive ectasia patients

Authors:

K. Jaskiewicz[1], M. Maleszka-Kurpiel[2,3], M. Kabza[4], R. Ploski[5], M. Rydzanicz[5], M. Gajecka[4,1]; [1]Inst. of Human Genetics, Polish Academy of Sci., Poznan, Poland, [2]Chair of Ophthalmology and Optometry, Poznan Univ. of Med. Sci., Poznan, Poland, [3]Optegra Eye Hlth.Care Clinic, Poznan, Poland, [4]Chair and Dept. of Genetics and Pharmaceutical Microbiol., Poznan Univ. of Med. Sci., Poznan, Poland, [5]Dept. of Med. Genetics, Med. Univ. of Warsaw, Warsaw, Poland

Background: Post-refractive ectasia is characterized by progressive corneal steepening and thinning, leading to severe progressive irregular astigmatism and deterioration of visual acuity. Such ectasia complicates 0.66% of the laser-assisted in situ keratomileusis (LASIK) procedures, as well as the small incision lenticule extraction and photorefractive keratectomy. Post-refractive ectasia, like keratoconus, is classified under “ectatic diseases”. The exact pathophysiology of post-LASIK ectasia remains unknown, however observed histological abnormalities are similar to keratoconus. Here, for the first time in the KTCN research, we present a transcriptomic profile of corneal epithelium (CE) in patients with post-refractive ectasia.
Materials and Methods: In the preliminary study 4 post-refractive ectasia patients (3 males and 1 female) undergoing cross-linking procedure and 5 mild myopia patients (3 males and 2 females) undergoing refractive error correction (controls) were ascertained. The *central*, *middle*, and *peripheral* CE topographic regions were separated from each obtained CE, based on CE thickness mapping. The RNA samples from 27 experimental samples were extracted using RNA/DNA/Protein PurificationPlus MicroKit (Norgen Biotek). The NGS libraries were prepared using TruSeq Stranded TotalRNA LibraryPrep Gold (Illumina) and sequenced on Novaseq platform (80mln read pairs per sample). RNA-seq data was analyzed implementing the previously established pipelines (PMID 30994860).
Results: We observed a characteristic doughnut pattern (thin cone center surrounded by thickened annulus) on each ectatic epithelial thickness map. We recognized characteristic transcriptomic profiles reflecting irregularities in epithelial thickness across the three analyzed regions. We identified differentially expressed genes involved in the hallmarks of DNA repair, E2F targets, mTORC1 signaling, Myc targets, unfolded protein responses, and reactive oxygen species, with a characteristic decrease in *middle* CE region and upregulation in *central* and *peripheral* CE regions. Moreover, the hallmark of interferon-gamma, but not inflammatory responses, was evenly upregulated in all regions.
Conclusions: The phenotypic abnormalities in distinct CE regions are related to the identified transcriptomic alterations. Further detailed studies would help identify patients at high risk of developing ectasia, and thus disqualify them from refractive corneal surgery and prevent postoperative complications.
Support: The National Science Centre grant no.2018/31/B/NZ5/03280.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1840 Trans-Ethnic Analysis of PCOS Subtype Genomewide Association Signals Reveals 3 Shared Subtype-Specific Loci.

Authors:

K. Brewer[1], R. Sisk[2], H. Lee[3], L. Moolhuijsen[4], K. van der Ham[4], Y. Louwers[4], J. Visser[4], M. Dapas[2], S. Franks[5], J. Laven[4], C. Li[6], A. Dunaif[1]; [1]Icahn Sch. of Med. at Mount Sinai, New York, NY, [2]Northwestern Univ., Chicago, IL, [3]Ewha Womans Univ. Sch. of Med., Seoul, Korea, Republic of, [4]Erasmus Med. Ctr., Rotterdam, Netherlands, [5]Imperial Coll. London, London, United Kingdom, [6]Univ. of Southern California, Los Angeles, CA

The phenotypic variation in PCOS suggests underlying genetic heterogeneity, but a recent genomewide association study (GWAS) meta-analysis of European ancestry (EA) PCOS cases found similar genetic architecture of PCOS defined by NIH and Rotterdam diagnostic criteria, suggesting these criteria do not identify biologically distinct disease phenotypes. Using hierarchical clustering (HC) in EA NIH PCOS cases, we identified stable PCOS subtypes, designated reproductive, metabolic, and background (Dapas. PLoS Med, 2020). These subtypes appeared to capture biologically meaningful differences as they were associated with novel and distinct genomewide significant loci. We performed an updated GWAS meta-analysis with additional EA NIH PCOS cases, and a trans-ethnic meta-analysis including a Korean ancestry (KA) NIH PCOS cohort (Lee. Hum Reprod, 2015). PCOS subtypes were identified by HC. Genotyping was as follows: EA discovery cohort (620 cases, 2951 controls), Illumina OmniExpress; EA replication cohort (371 cases, 926 controls), Metabochip; and KA cohort (417 cases, 926 controls), HumanOmni1-Quad v1. All data were imputed using the TOPMed (R2) panel. Eight loci were genomewide significant in the EA meta-analysis: 4 novel loci (reproductive subtype); 2 loci (metabolic subtype), including *C9orf3*; 2 loci (background subtype), including *FSHB/ARL14P*. Five of 6 other lead SNPs in the original analysis remained nominally significant. Three loci were genomewide significant in the trans-ethnic meta-analysis: a novel locus on chr 3, *EPHA6*, a receptor tyrosine kinase that binds GPI-anchored ephrin-A family ligands (rs140766105, p=1.08x10-8, reproductive subtype); a previously identified locus on chr 9, *C9orf3* (*AOPEP*), an aminopeptidase that catalyzes the hydrolysis of amino acid residues (rs10761370, p=4.12x10-9, metabolic subtype); a previously identified locus on chr 11, *FSHB/ARL14P*, which encodes the β-subunit of FSH and is associated with fertility (rs10835649, 1.23x10-11, background subtype). We have replicated the 8 subtype loci in the EA meta-analysis suggesting these subtypes have distinct genetic architecture. For the first time, a trans-ethnic meta-analysis demonstrates that 3 subtype-specific loci play a role in PCOS pathogenesis across diverse populations.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1842 Uncontrolled hypertension is associated with AGT proteomics and risk for clinical sequelae.

Authors:

L. Krohn, R. Hoffing, A. Holleman, P. LoGerfo, M. Plekan, S. Akle, L. Ward, D. Zappe, H-C. Tu, P. Nioi, A. Deaton; Alnylam Pharmaceuticals, Cambridge, MA

Background: Hypertension is a major preventable risk factors for cardiovascular disease. Despite the availability of effective anti-hypertensive agents, hypertension remains a widespread and largely uncontrolled condition. In this study, we aimed to quantify the unmet need in hypertension management in the UK and investigate the impact of uncontrolled hypertension on risk for clinical sequelae. We also assessed angiotensinogen (AGT) as a potential therapeutic target for blood pressure control using a genetic instrument for AGT lowering. **Methods:** Clinical data were collected from UK Biobank intake, hospital, and general practitioner (GP) records. We gathered per-visit systolic blood pressure (SBP) readings and anti-hypertensive prescriptions for those with GP data (N=159,055). Visit-to-visit SBP variability was analyzed using variation independent of the mean (VIM; Yano Y, *AJH*, 2017). As a proxy for AGT-lowering, we selected the strongest pQTL from GWAS of array genotypes on normalized proteomic data from 46,554 participants (rs699-A, beta= -0.33, p < 4e-323). For those with genetic data (N=431,567), we extracted rs699 array genotypes for risk analyses. We assessed the influence of both uncontrolled hypertension and AGT on hypertension-related risks with logistic and Cox regression, each adjusted for age, sex, medication-adjusted LDL, BMI, alcohol intake, smoking status, and type 2 diabetes. **Results:** Of those with GP data, 70% qualified for hypertension, 74% of whom were prescribed one or more anti-hypertensive medication(s) within the covered time frame. Only 23% of the medicated population achieved consistent SBP control (all reads < 140 mm/Hg), while 28% had severely uncontrolled SBP after treatment (2 consecutive reads >=155 mm/Hg) and the highest risk for major adverse cardiac events (MACE, HR=1.2, p=4.6e-11). High visit-to-visit SBP variability is associated with increased risk for- and faster time to- MACE (VIM quartile HR=1.92, p=7.5e-53). The *AGT* rs699-A variant is associated with reduced risk and time to hypertension diagnosis (HR=0.90, p=9.6e-33) and stroke after hypertension diagnosis (HR=0.88, p=0.02), as well as less risk for uncontrolled hypertension after medication (OR=0.98, p=0.005). **Conclusions:** Consistently controlled blood pressure and lower visit-to-visit blood pressure variability are associated with reduced risk for clinical sequelae from hypertension. Genetically lower angiotensinogen is associated with reduced risk for hypertension and sequelae, suggesting that lowering angiotensinogen via *AGT* silencing may improve blood pressure control and reduce MACE.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1843 Uncovering Novel Genetic Loci and Biological Pathways Associated with Age-Related Cataracts through GWAS Meta-Analysis.

Authors:

S. Diaz Torres[1], S. Lee[2], L. Garcia Marin[3], A. Campos[4], G. Lingham[2], J. Ong[5], D. Mackey[6], K. Burdon[7], X. Dong[8], S. MacGregor[1], P. Gharahkhani[1], M. Renteria[1]; [1]QIMR Berghofer Med. Res. Inst., Brisbane, Australia, [2]The Univ. of Western Australia, Perth, Australia, [3]Queensland Inst. of Med. Res., Brisbane, Australia, [4]Inst. for Molecular BioSci., Brisbane, Australia, [5]QIMR Berghofer Med. Res. Inst., Herston, Australia, [6]Lions Eye Inst., Nedlands, Australia, [7]Univ. of Tasmania, Hobart, Australia, [8]Brigham & Women's Hosp, Cambridge, MA

Age-related cataract is a highly prevalent eye disorder that results in the clouding of the crystalline lens and is one of the leading causes of visual impairment and blindness. The disease is influenced by multiple factors including genetics, prolonged exposure to ultraviolet radiation (UVR), and a history of diabetes. However, the extent to which each of these factors contributes to the development of cataracts remains unclear.This study presents the largest genome-wide association study of cataracts to date, using data from 127,985 cases and 837,371 controls. We performed gene enrichment analysis to identify genes and biological pathways associated with cataracts. We integrated our results with gene expression reference datasets to identify genetic variants modifying risk for cataracts through changes in the expression of specific genes. We further explored drug-gene interactions to better understand the potential impact of pharmacological interventions on cataract development. Finally, we explored whether a causal relationship underlies the known comorbidity between type 1 diabetes and cataracts using a mendelian randomization framework, and the association between UV exposure and cataract risk in adults using a polygenic risk scoring approach.Our study identified 85 independent genome-wide significant loci, 37 of which are novel. Gene-based association tests identified 126 genes associated with cataracts, hinting at a potential relationship between negative regulation of lipid biosynthesis and the development of cataracts. Four of the genes identified GNL3, JAG1, METTL21A, and CREB1 are involved in drug-gene interactions. Moreover, Mendelian Randomisation analysis identified a putative causal relationship between genetic predisposition to type 1 diabetes and an increased risk of cataracts. Lastly, we found evidence indicating that early-life exposure to UVR may have an impact on the later development of cataracts.Our findings advance our understanding of the genetic basis of cataract and provide new insights into its etiology. We identified multiple genes and biological pathways associated with the condition, including associations with four genes from which drug repurposing could be proposed. Our results suggest a causal association between type 1 diabetes and cataracts. Also, we highlighted a surrogate measure of UV light exposure as a marker of cataract risk in adults and drug-genes interactions that has the potential of informing novel therapies.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1844 Uncovering phenotypic abnormalities associated with metabolic genes using *All of Us.*

Authors:

K. Bates, L. Howell, B. Bates, D. Stone, J. Bryan, J. Grose, M. Bailey, M. Davis; Brigham Young Univ., Provo, UT

The occurrence of metabolic disease has grown to epidemic proportions. These diseases can result from impaired regulation of metabolic pathways. Stroke, heart disease, and diabetes are a few examples of the health complications that can occur when cells excessively expend essential nutrients, such as glucose, toward one pathway at the expense of other crucial pathways. Understanding the complex genetic regulation of these pathways can increase our ability to treat and prevent metabolic disease and improve the quality of life of millions of individuals. PAS kinase (PASK) is a nutrient sensing protein kinase that senses the availability of glucose and allocates it towards different pathways. Additionally, PASK has two substrates, upstream stimulatory factor-1 (USF1) and ataxin-2 (ATXN2), that have been shown to regulate lipid metabolism and glucose homeostasis in mammalian systems. We seek to reveal associations between PASK, USF1, and ATXN2 variants and a variety of human phenotypes and diseases using the All of Us Research Program dataset. With more than 180,000 individuals that have both genetic and electronic health record (EHR) data, the All of Us Research Program dataset is an incredible asset for uncovering novel relationships between these genes and disease. After completing basic quality control procedures, our final cohort includes 171,882 individuals, with more than 75% of the cohort representing minority or traditionally underrepresented populations in biomedical research. Almost 70,000 of these individuals are of African/African American, Indigenous American/Latino, South Asian, East Asian, or Middle Eastern ancestries. Using the PheWAS package, we identified phenotypes (represented as phecodes) based on the insurance billing codes found in their EHRs. We ran logistic regression to identify associations between PASK, USF1, and ATXN2 variants with phecodes and metabolic lab values. Our preliminary findings suggest single-nucleotide polymorphisms (SNPs) in USF1 associated with infectious disease ($p < 0.0001$), as well as SNPs in ATXN2 associated with increased hemoglobin A1C levels ($p < 0.0001$) and SNPs in PASK associated with increased triglyceride levels ($p < 0.001$). We anticipate discovering more associations between these genes and metabolic phenotypes, in addition to other phenotypes that may be unexpected, including cancer, mental illness, or neurodegenerative disease. These associations may provide valuable insight into several disorders and suggest novel molecular pathways for future study.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1845 Uncovering Potential Susceptibility Genes for Knee Osteoarthritis through Whole Exome Sequencing.

Authors:

A. BISWAS, B. Garg, R. MALHOTRA; All India Inst. of Med. Sci., New Delhi, India

Title: Whole-Exome Sequencing Reveals Potential Susceptibility Genes for Knee Osteoarthritis. **Introduction:** Knee osteoarthritis (kOA) is a major cause of disability in the elderly with the global prevalence of 16% and 22-39% in India. Major risk factors for the kOA is aging, diet, lifestyle and most importantly genetic modality. Till date, genetic predisposition is not fully understood and also there is lack of genetic studies from India. **Aims & Objectives:** The primary objective of this study was to identify casual genes that predispose individuals to kOA in Indian population.**Materials and methods:** In this study, we have sequenced 50 Knee OA patients with confirmed diagnosis through radio-graphical intervention and 20 age and sex matched healthy controls through whole exome sequencing. Clinical history, demographical information & blood sample were collected for serological and molecular work. DNA was isolated and subjected to whole exome sequencing (100X) after preparing libraries using nextera exome enrichment kit. All the variants were filtered out for the candidate's genes reported in previous literature and also, in an independent filtering, all non-synonymous and pathogenic variants were listed as per Clinvar which were associated with osteoarthritis or any related disorders. The expression profiles of mRNAs in blood were assessed using whole transcriptome sequencing focussing on all the prioritised genes from exome study.**Results:** Germline variants were identified in previously known kOA susceptibility genes such as *BTNL2* and *HLA DPB1*, and several other genes with potentially deleterious variants for kOA. In an independent filtering for casual variants which were associated with kOA or related bone disorders based on clinvar database, we found previously reported variants in genes *NACA2, TLR10, GLE1 and MATN3* having high MAF as compared to 1000genomes and SAS data frequency. Interestingly, whole exome analysis demonstrated that most samples shows multiple modifier variants whereas rare variants were specific to the individuals which were evident when 191 associated genes evaluation. All the prioritised genes reported were differentially expressed as compared to controls.**Conclusion:** Our findings indicate that kOA patients harbour potentially deleterious causative germline variants in genes that function in several candidate causative pathways including immune responses, inflammatory and cartilage degradation. Genetic testing for a wider variety of kOA-predisposition genes could provide better screening approach for high-risk groups of osteoarthritis.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1846 Uncovering the nuclear genetic basis of mtDNA heteroplasmy.

Authors:

R. Alhariri[1], W. Shi[1], D. Puiu[2], N. Pankratz[3], N. Lake[4], M. Lek[5], Y. Huang[6], E. Tsai[6], K. G. Meilleur[6], J. Shi[6], H. Runz[6], D. Arking[1]; [1]McKusick-Nathans Inst., Dept. of Genetic Med., Johns Hopkins Univ. Sch. of Med., Baltimore, MD, [2]Dept. of BioMed. Engineering, Johns Hopkins Univ., Baltimore, MD, [3]Univ. of Minnesota, Minneapolis, MN, [4]Yale Univ., New Haven, CT, [5]Yale Univ. Sch. of Med., New Haven, CT, [6]Biogen, Cambridge, MA

Background: Mitochondrial heteroplasmy, the coexistence of mutated and wildtype mitochondrial DNA within a cell, is associated with overall mortality and aging-related phenotypes. Mitochondrial mutations can be inherited maternally or acquired somatically, with previous work estimating that about 70% of heteroplasmies are somatic and increase in an age-dependent manner.
Methods: To study the relationship between nuclear encoded variation and mitochondrial heteroplasmy, we conducted genome-wide association studies on heteroplasmy count as well as a functional approximation of the deleteriousness of the heteroplasmic mutations, using the mitochondrial local constraint score sum (MSS). MSS utilizes a conservation-based metric that accounts for class of mutation as well as constraint at a nucleotide and gene-level. We included 162,853 participants of European Ancestry from the UK Biobank. Whole-genome sequencing data from whole blood DNA samples were used to call heteroplasmy through the MitoHPC pipeline. 11,392,286 SNPs with a minor allele frequency (MAF) >0.005 were included in the GWAS after imputation.
Results: Genes were identified through positionally mapping variant hits in genes or up to 10 KB to the nearest gene. rs2906685 in *DNAH11* was identified as a lead SNP (Beta=-0.017; p= 1.83x10^-10) in the GWAS associated with heteroplasmy count, while 14:96204037_CA_C in *TCL1A* (Beta=0.0061; p=1.20x10^-09) and rs2736100 in *TERT* (Beta=0.0046;p=3.71x10^-10) were identified in our GWAS on MSS. Encoding the axonemal arm of a ciliary dynein, *DNAH11* may be biologically relevant to the development of heteroplasmy given dynein's relation to mitochondrial motility which regulates fusion and fission dynamics. Comparatively, *TERT* has been reported to serve a protective role against mitochondrial DNA from damage and reduce the production of reactive oxygen species. Less is known about *TCL1A* which has been described as a potential tumor suppressor gene.
Conclusion: *DNAH11* and *TERT* were identified as two genes that may be associated with mitochondrial processes of motility and DNA damage repair. Future directions involve validating findings in a cross-ancestry cohort through the inclusion of participants from TOPMed and AllofUs.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1847 Understanding dynamic molecular regulations in brains with regard to Tuberous Sclerosis diseases.

Authors:

S. Yoo[1], Y. liu[1], E. Lee[1], Q. Chen[1], S. McGee[2], L. Wang[1], J. Zhu[1]; [1]GeneDx, Stamford, CT, [2]GeneDx, Gaithersburg, MD

Tuberous Sclerosis Complex (TSC) is an autosomal dominant disorder affecting 1 in 6000 newborns in the US driven by variants in *TSC1* or *TSC2*. These variants can affect diverse organ systems including the brain, eye, heart, kidney, skin, and lung. The primary affected tissue/organ is the brain which results in several neurologic disorders such as epilepsy, developmental delay, and autism.

Proteins encoded by *TSC1* and *TSC2* form a complex that inhibits the mTOR pathway, in turn regulating many pathways essential for cell growth in all cells. To understand the relationship between TSC and mTOR pathway in the brain and other organ systems, we first investigated cell type-specific expression of *TSC1*, *TSC2*, and *mTOR* and inferred mTOR complex 1 pathway activity (mTORC1 activity). Across 79 human cell types, cells in the brain and eye had the lowest mTORC1 activity while these cells had the highest gene expression levels of *TSC1*, *TSC2*, and *mTOR. TSC1* and *mTOR* expression levels and mTORC1 activity were significantly anti-correlated across cell types. The results suggest that brain- and eye-specific cells are more likely to be affected by TSC-mTOR pathway dysregulation. In brain and non-brain human bulk tissue profiles in the GTEx database, the three genes were tightly correlated in brain tissues but less correlated in other tissues (t-test $p=1.9x10^{-19}$). To understand the co-regulation in the brain, we identified *ARNT2*, *ETV5*, and *BHLHE41* as key transcription factors contributing to the brain-specific co-regulation. *ARNT2* plays a role in the development of brain, visual, and renal function, consistent with the top tissues affected in tuberous sclerosis diseases.

The tightly co-transcriptional regulation of *mTOR, TSC1,* and *TSC2* results in stable mTORC1 activity in the brain. Variants in *TSC1* or *TSC2* will disturb the stable state among mTORC1 activity and result in lower or higher mTORC1 activity. To test the hypothesis, we checked microcephaly (likely due to lower mTORC1 activity) and macrocephaly (likely due to higher mTORC1 activity) phenotypes in individuals with *TSC2* variants. In the GeneDx database, among patients with TSC2 variants, 291 patients were with microcephaly, and 185 were with macrocephaly. The corresponding variants in *TSC2* were not enriched in any TSC2 structure domain, suggesting that the lower or higher mTORC1 activity was determined by other genetic or environmental factors. For example, rs1042602 in tyrosinase was significantly associated with macrocephaly among patients with TSC2 variants, suggesting whole genome sequencing or whole exome sequencing is necessary to understand the disease heterogeneity.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1848 Understanding the genetic complexity of puberty timing across the allele frequency spectrum.

Authors:

K. Kentistou, ReproGen Consortium; Univ. of Cambridge, Cambridge, United Kingdom

Pubertal timing varies considerably and has been associated with a range of health outcomes in later life. To elucidate the underlying biological mechanisms, we performed multi-ancestry genetic analyses in ~800,000 women, identifying 1,080 independent signals associated with age at menarche. Collectively these loci explained 11% of the trait variance in an independent sample, with women at the top and bottom 1% of polygenic risk exhibiting a ~11 and ~14-fold higher risk of delayed and precocious pubertal development, respectively. These common variant analyses were supported by exome sequence analysis of ~220,000 women, identifying several genes, including rare loss of function variants in *ZNF483* which abolished the impact of polygenic risk. Next, we implicated 660 genes in pubertal development using a combination of *in silico* variant-to-gene mapping approaches and integration with dynamic gene expression data from mouse embryonic GnRH neurons. This included an uncharacterised G-protein coupled receptor *GPR83*, which we demonstrate amplifies signalling of *MC3R*, a key sensor of nutritional status. Finally, we identified several genes, including ovary-expressed genes involved in DNA damage response that co-localise with signals associated with menopause timing, leading us to hypothesise that the ovarian reserve might signal centrally to trigger puberty. Collectively these findings extend our understanding of the biological complexity of puberty timing and highlight body size dependent and independent mechanisms that potentially link reproductive timing to later life disease.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1849 Unleashing the Power of Telomere Length Genetics: Unravelling Disease Associations and Rare Variant Impacts from 483,778 UK Biobank Genomes

Authors:

O. Burren[1], R. Dhindsa[2], S. Deevi[1], **S. MacArthur**[1], J. Mitchell[1], S. Wen[1], A. Nag[1], N. Razdan[3], H. Olsson[4], K. Smith[1], S. Wasilewski[1], A. Platt[5], M. Fabre[1], K. Carss[1], Q. Wang[1], S. Petrovski[1]; [1]Ctr. for Genomics Res., Discovery Sci., BioPharmaceuticals R&D, AstraZeneca, Cambridge, United Kingdom, [2]Ctr. for Genomics Res., Discovery Sci., BioPharmaceuticals R&D, AstraZeneca, Waltham, MA, [3]BioSci.s COPD & IPF, Res. and Early Dev., Respiratory & Immunology, BioPharmaceuticals R&D, AstraZeneca, Gothenburg, Sweden, [4]Translational Sci. and Experimental Med., Res. and Early Dev., Respiratory & Immunology, BioPharmaceuticals R&D, AstraZeneca, Gothenburg, Sweden, [5]Translational Sci. and Experimental Med., Res. and Early Dev., Respiratory & Immunology, BioPharmaceuticals R&D, AstraZeneca, Cambridge, United Kingdom

Genetic perturbations that influence human telomere length (TL) have also been linked to a spectrum of human diseases, including the discovery of TL associated genes that have been found to possess antagonistic pleiotropy (e.g., reduced risk of cancer and increased risk of interstitial lung diseases). Thus, elucidating the genetic drivers of TL biology will further our understanding of many complex diseases. We compared TL estimates obtained from qPCR[1] and TelSeq[2] (WGS) for 483,778 participants from the UK Biobank (UKB), revealing only a partial correlation (r^2=0.15). To optimize the biological signal and enhance the statistical power of our genetic association analyses, we constructed a composite score (PC1)[2] derived from a principal component analysis (PCA) incorporating both WGS and qPCR measurements. PC1 successfully identified known TL associations, such as with age (β = -0.02, SE ± 0.00, P<$2x10^{-16}$). Applied to common variant analysis, PC1 significantly improved genetic locus discovery yield and narrow-sense heritability (7.9%, SE ±0.9%) compared to independent qPCR (6.8%, SE ±0.8%) or WGS (4.9%, SE ±0.7%) TL measurements alone.
Given the access to 483,778 WGS we conducted analyses to target rare (MAF<0.1%) coding variants individually and in an aggregated gene collapsing framework, performing the largest study of rare variant contribution to TL, to date. We identified 50 rare variants and 22 distinct genes associated with PC1 (p<$1x10^{-8}$). Many of these genes overlapped common variant loci, allowing us to establish allelic series (e.g., *ACD*) and to prioritise causal genes and variants linked with large effects on telomere dysfunction. Three genes were known drivers of clonal haematopoiesis (CH), prompting subsequent targeted somatic variant analyses. We uncovered novel signals, including a positive association between somatic variants in *SRSF2* and longer TL (Effect=0.65, SE ± 0.17, P=1.3 x 10^{-4}), as well as an inverse relationship between TL and *DNMT3A* clone size (Effect=-0.69, SE ± 0.09, P=$3.7x10^{-12}$). Our study highlights the biological insights that arise from integrating orthogonal data modalities within a substantial sample of participants. We reveal a complex interplay between telomere length and clonal haematopoiesis, shedding light on the broader genetic antagonistic pleiotropy existing between cancer and other age-related diseases. Furthermore, our findings emphasise the substantial effects of rare germline and somatic variants on TL, offering valuable implications for both disease pathogenesis and the development of therapeutic strategies. 1. Codd,V. *et al.* (2021).2. Ding, Z. *et al.*(2014).3. Aschard,H. *et al.* (2014).

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1850 Unraveling Oligodendrocyte Heterogeneity in Neurodegenerative Disorders.

Authors:

N. Rahimzadeh, S. Morabito, V. Swarup; Univ. of California, Irvine, CA

Myelin loss and white matter changes are prominent features observed in neurodegenerative disorders, particularly in multiple sclerosis (MS) and Alzheimer's disease (AD). While MS exhibits extensive demyelination, AD also displays pathological signs of white matter lesions and degeneration. Oligodendrocytes (ODC), specialized cells in the nervous system, play a vital role in generating myelin, forming the myelin sheath that envelops axons. The myelin sheath not only supports axon metabolism, but it also enables the rapid and saltatory propagation of action potentials. Understanding the underlying causes of demyelination in MS can provide valuable insights into AD and guide therapeutic interventions aimed at preserving white matter integrity. Single-cell/nucleus transcriptomics holds great promise in advancing our understanding of cellular heterogeneity and its contribution to disease pathology. In this study, our objective is to leverage publicly available single-nuclei RNA-seq datasets for MS, including Absinta et al. (n = 72,959), Schirmer et al. (n = 48,919), Jakel et al. (n = 17,799), as well as AD datasets, including Morabito and Miyoshi et al. (n = 191,890), Mathys et al. (n = 80,660), and Olah et al. (n = 16,242). Using a single-cell variational inference (scVI) model, specifically a variational auto-encoder, we will integrate and harmonize these datasets. By integrating the reported nuclei from each study, we will gain sufficient statistical power to investigate the heterogeneity of ODC populations across MS and AD. Our downstream analysis will include the identification of gene-regulatory networks using weighted correlation network analysis and trajectory analyses to infer cell states across the disease spectrum. This analysis will enable us to uncover both shared and distinct ODC populations associated with the studied traits.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1851 Unraveling sex differences in the complex genetic architecture of blood pressure and insights into cardiovascular disease pleiotropy.

Authors:

M-L. Yang, C. Xu, X. Zhou, S. Ganesh; Univ. of Michigan, Ann Arbor, MI

Background: Blood pressure (BP) plays a vital role in the development of cardiovascular diseases (CVDs). However, significant sex differences exist in terms of prevalence and clinical considerations. The molecular and genetic basis of sex-specific BP patterns remains poorly understood, impeding our understanding of the relationship between BP genetics, specific forms of hypertension, and cardiovascular diseases. Further investigation in this area is necessary to advance our knowledge of sex-specific cardiovascular health.
Methods: We conducted sex-stratified and combined-sex genome-wide association studies (GWAS) on BP traits using UK Biobank (UKB) data. Our analysis aimed to reveal sex differences in genetic associations and identify loci with sex-specific effects. Additionally, we examined the association between sex-specific BP polygenic risk scores (PRS) and hypertension phenotypes, as well as several cardiovascular diseases. Leveraging cross-trait colocalization in a sex-stratified manner, we further identified genetic regions associated with BP that exhibited sex-specific risks for associated CVDs.
Results: Our analysis identified 1,375 loci associated with BP traits, including 29 novel and 1,346 previously known loci. Of these, 412 were specific to females, and 142 were specific to males. Sex-stratified GWAS revealed 90 unique loci (7%) compared to combined-sex GWAS. Gene-by-sex interactions and sexually dimorphic effects (SDE) analysis highlighted four genomic regions, including the chromosome 13q34-*COL4A1*/*COL4A2* locus implicated in both analyses. We identified 18 BP-associated loci with significant SDE, such as chromosomes 17q23.3-*PECAM1*/*TEX2*, 19q13.2-*ACTN4*/*CAPN12*, and 13q34-*COL4A1*/*COL4A2*. In the sex-stratified PRS analysis, we found a strong female-specific association between PRS of pulse pressure and fibromuscular dysplasia (FMD), a hypertension subtype primarily affecting women. This association may stem from robust female-specific signals within the regions encompassing *COL4A1*, *COL4A2*, *CDKN2B-AS*, and *MAP9*. Our study revealed sex-specific genetic effects on BP regulation in established genetic loci linked to CVDs, shedding light on the underlying mechanisms and observed sex disparities.
Conclusion: Our findings emphasize the central role of blood pressure in cardiovascular disease risks and partially explain variations in CVD patterns across different biological sexes.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1852 Unraveling shared genetics across asthma subtypes and 70 biological and physiological asthma-related traits.

Authors:

R. Vernet[1], C. Linhard[1], Y. Suzuki[2], H. Mohamdi[1], J. Sambourg[1], F. Demenais[1], H. Aschard[2], H. Julienne[2,3], E. Bouzigon[1]; [1]INSERM, Université Paris Cité, UMR 1124, Group of Genomic Epidemiology of Multifactorial Diseases, Paris, France, [2]Inst. Pasteur, Université Paris Cité, Dept. of Computational Biology, Paris, France, [3]Inst. Pasteur, Université Paris Cité, Bioinformatics and Biostatistics Hub, Paris, France

Asthma is a heterogeneous disease with a large spectrum of clinical expression over the life span, reflecting the multiple underlying biological mechanisms or endotypes. This clinical and biological heterogeneity might be at least partly due to genetic heterogeneity across patients. To better understand the etiology of asthma heterogeneity, we investigated the patterns of genetic correlations, between various asthma subtypes and multiple biological and physiological traits involved in asthma pathophysiology.

We built a harmonized comprehensive database of 75 GWAS summary statistics of asthma, asthma subtypes (moderate-severe asthma, childhood onset, adult-onset and time-to-asthma onset) and asthma-associated traits (blood cell counts, various cytokines, total serum Immunoglobulin E levels, lung function phenotypes and body-mass index). We applied the JASS (Joint Analysis of Summary Statistics) pipeline on all 75 traits to: i) curate and harmonize GWAS Summary Statistics, ii) impute missing statistics, and iii) compute SNP heritabilities (h^2) and genetic correlations (r_g) based on LD Score regression.

The highest SNP heritabilities were found for moderate-to-severe asthma (h^2=0.18, SE=0.02), childhood-onset asthma (h^2=0.15, SE=0.03), and time-to-onset of asthma (h^2=0.11, SE=0.03) while adult-onset asthma showed lower heritability (h^2=0.02, SE=0.002). Out of the 1,225 pairs of asthma subtypes and biological traits, 45 had significant genetic correlations with $p<4.1x10^{-5}$ (after Bonferroni correction for the number of trait pairs) and seven additional pairs had nominally significant genetic correlations ($p<0.05$).

All asthma subtypes were significantly correlated with lung function traits (FEV_1, FVC and FEV_1/FVC, r_g ranging from -0.31 to -0.45; $p<1.5x10^{-5}$), and eosinophil counts (r_g 0.27 to 0.39, $P<4x10^{-5}$). Interestingly, we also observed genetic correlations that were specific to asthma subtypes. Moderate-to-severe asthma was correlated with obesity and BMI (r_g ~0.20, $P\leq2x10^{-3}$), basophil counts (r_g 0.10, P=0.02) and eosinophil cationic protein levels (r_g=0.68, P=0.04). Adult-onset asthma was significantly correlated to obesity and BMI (r_g=0.15-0.18, $p < 0.01$) and basophil counts (r_g=0.12, p=0.004).

In this study, we observed genetic factors shared by all asthma sub-types and some physiological traits. We also detected specific patterns of genetic correlations of asthma sub-types with biological and physiological traits, which shows a genetic contribution to asthma heterogeneity. (Funding: ANR-20-CE36-0009)

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1853 Unraveling the genetic architechture of autoimmune gastritis and pernicious anmeia.

Authors:

B. M. Morris, A. W. Reynolds; Baylor Univ., Waco, TX

Autoimmune gastritis (AIG) is a condition characterized by inflammation of the stomach lining due to autoimmune destruction of parietal cells that affects between 0.5-2.5% of the US population. Parietal cells are specialized epithelial cells, found in the fundus of the stomach, aid in digestion and nutrient absorption by secreting hydrochloric acid and intrinsic factor. AIG leads to multiple disease end-stage phenotypes, including iron deficiency anemia, pernicious anemia, and increases the risk of developing gastric cancers; but it is unclear why some patients develop different end-stages of the disease and what risk factors influence this. Pernicious anemia (PA), one end-stage of AIG, causes malabsorption of vitamin B12 (cobalamin), an essential micronutrient that helps maintain healthy blood cells, nerves, DNA synthesis and structural stability, methyltransferase activity, and many other important metabolic processes. Quick diagnosis of PA is vital, as left untreated, the condition can cause permanent neurological damage and even death. However, a positive diagnosis of PA can be difficult because of slow progression and nondescript symptoms that can mask the underlying disease, leaving many patients undiagnosed or misdiagnosed for 5 - 10 years. To better understand the genetic risk factors of AIG and PA we used whole-genome sequence data from the All of Us research program to conduct a genome-wide association study (GWAS) on 3761 AIG cases, 979 PA WGS cases, and 245,388 healthy controls. Using a statistical fine-mapping approach, we identified credible sets of putatively causal variants associated with AIG and PA. We compared our findings with previously published GWAS of other autoimmune diseases to shed light on common autoimmune disease risk factors. These findings contribute to ongoing efforts to characterize polygenic and pathway risk scores for AIG, PA, and other autoimmune diseases, laying the foundation for future improvements in clinical guidelines and diagnostic and therapeutic strategies.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1854 Unraveling the genetic enigma of Endometriosis: Novel insights and gene discovery on a deeply characterized cohort.

Authors:

A. Santin[1], B. Spedicati[1,2], A. Morgan[2], S. Lenarduzzi[2], P. Tesolin[1], G. G. Nardone[1], D. Mazzà[2], G. Di Lorenzo[2], F. Romano[2], F. Buonomo[2], A. Mangogna[2], M. P. Concas[2], G. Zito[2], G. Ricci[1,2], G. Girotto[1]; [1]Dept. of Med., Surgery and Hlth.Sci., Univ. of Trieste, Trieste, Italy, [2]Inst. for Maternal and Child Hlth., I.R.C.C.S. “Burlo Garofolo”, Trieste, Italy

Background: Endometriosis (EM) is a multifactorial gynecological disorder. Despite its prevalence, EM is one of the most underdiagnosed and undertreated disorders, with a nine-year diagnostic delay from symptoms onset. Therefore, detangling the genetic mechanisms underlying EM etiopathogenesis could be relevant to identify novel molecular targets as diagnostic and therapeutic biomarkers. Several Genome-Wide Association Studies have been reported in literature, allowing the detection of candidate genes; however, the causal genetic mechanisms underlying EM are still undetermined.
Methods: Eighty patients with confirmed EM diagnosis were enrolled at I.R.C.C.S. "Burlo Garofolo" hospital (Trieste, Italy). All patients underwent a deep clinical assessment, comprising the evaluation of the most common EM-associated symptoms, food preferences, and lifestyle habits. Whole-Exome Sequencing (WES) analysis was firstly focused on a selected list of 46 EM-associated genes and subsequently on novel candidates discovery. Results were then compared with a cohort of 105 healthy women.
Results: WES allowed the detection of over 60 rare predicted damaging heterozygous variants within 24 genes. Results were classified into three main groups: a) recurrent genes shared among many patients, b) private variants within specific genes, c) variants within novel candidates. In group a) we identified rare predicted damaging missense variants within *LAMA5* gene, already associated with EM stage III-IV. Interestingly, 3/5 EM patients carrying variants within this gene were diagnosed with EM stage III-IV.In group b), a rare predicted damaging missense variant in *WT1* gene was detected in a 35-year-old woman. *WT1* is involved in female fertility regulation and linked with EM-associated dysmenorrhea. Notably, this patient was diagnosed with infertility and reports severe dysmenorrhea. Concerning group c), eight EM patients carry rare damaging variants within *ABCA13* gene encoding a gangliosides transporter, a novel candidate involved in pain processing. Finally, to deepen the polygenic architecture of EM, an evaluation of the analyzed genes was conducted, revealing a statistically significant ($p<0.05$) higher burden of genes in EM patients than in controls.
Conclusions: For the first time, a joint approach of a detailed clinical characterization and WES was performed, allowing the detection of rare variants within EM-associated genes and novel candidates, thus identifying intriguing genotype-phenotype correlations. Ongoing analyses will be aimed at deepening the complex relation between EM-associated pain, food preferences, and EM disease severity.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1855 Unraveling the genetics of cellular senescence for target discovery for IPF.

Authors:

R. Mishra[1], L. Benedetti[2], S. Hopkins[3], P. Gormley[4], A. Hutchinson[5], A. Cortes[6], J. Davitte[7], W. Fahy[6], C. Buser-Doepner[1], R. Scott[8], A. Chu[9]; [1]GlaxoSmithKline, Collegeville, PA, [2]GlaxoSmithKline, Stevenage, United Kingdom, [3]GlaxoSmithKline, Steve age, United Kingdom, [4]GSK, Cambridge, MA, [5]GlaxoSmithKline, Cambridge, United Kingdom, [6]GSK, Stevenage, United Kingdom, [7]GSK, Collegeville, PA, [8]GSK, Cambridge, United Kingdom, [9]GlaxoSmithKline, Cambridge, MA

Idiopathic pulmonary fibrosis (IPF) is a complex, rapidly progressive lung disease with significant morbidity, mortality, and unmet medical need. Current antifibrotic treatments slow disease progression but do not halt it and have significant tolerability issues. There is an urgent need to identify new drug targets, with senescence and telomere biology emerging as important mechanisms in the pathobiology of IPF. Given the previously established causal link between telomere biology and IPF, we use leukocyte telomere length (TL), a biomarker for cellular senescence, among individuals of the U.K. Biobank (UKB) to unravel potential novel therapeutic targets for IPF. We used the log-transformed, Z-standardized TL released by UK Biobank. TL was measured in peripheral leukocytes of 411,886 participants of European ancestry represented as the ratio of telomere repeat number relative to a single copy gene. We ran a genome-wide association study using REGENIE, adjusting for age at baseline assessment, genetic sex, genotyping array, and the first 10 genetic principal components. We performed mendelian randomization by constructing genetic risk scores (GRS) for TL and testing the association with the largest public GWAS of IPF (Partanen et al 2021). Longer TL was significantly associated with lower odds of IPF (p-value = $5.68x10^{-54}$). We colocalized all 133 genome-wide significant TL loci with the IPF GWAS and filtered for cellular senescence-related genes. Out of the 133 loci, only one locus demonstrated evidence of a shared signal with IPF and mapped to a cellular senescence related gene, thymidylate synthetase (*TYMS)*. This observation also suggested this locus is associated with longer TL and lower odds IPF. Colocalization using GTEx tissues suggested increased telomere length was also associated with increased expression of *TYMS*. Since the *TYMS* loci does not reach genome-wide significance in IPF (p= 1.75e-06), this target hypothesis for IPF would be missed. Additionally, *TYMS* was shown to be a potential poor prognostic biomarker for IPF patients in a recently published, independent study using a gene expression profile of cellular senescence-related genes. Our findings link the genetics of TL and IPF together and illustrate how telomere length genetics can enable target discovery in IPF.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1856 Unraveling the shared genetic architecture of insulin resistance, type 2 diabetes, and cardiometabolic diseases independent of traditional risk factors.

Authors:

A. Ranta[1], B. Balliu[2,3], D. Sharma[1], M. Räsänen[1], T. Quertermous[1], J. W. Knowles[1], I. Carcamo-Orive[1]; [1]Dept. of Med., Div. of Cardiovascular Med., Cardiovascular Insitute and Stanford Diabetes Res. Ctr., Stanford Univ. Sch. of Med., Stanford, CA, [2]Dept. of Pathology and Lab. Med. & Dept. of Computational Med., UCLA David Geffen Sch. of Med., Los Angeles, CA, [3]Dept. of Biostatistics, UCLA Fielding Sch. of Publ. Hlth., Los Angeles, CA

Cardiometabolic diseases associated with insulin resistance (IR), such as coronary artery disease (CAD), type 2 diabetes (T2D), and non-alcoholic fatty liver (NAFLD), are the leading cause of death worldwide. IR is essential for developing T2D and NAFLD and is a major risk factor for CAD even in the absence of T2D, partly due to shared risk factors (i.e., hypertension and hyperlipidemia). However, recent data suggest that some IR genes have effects on CAD not accounted for by traditional risk factors.

We previously conducted a custom colocalization analysis to identify novel genes with primary effects on insulin sensitivity, leveraging publicly available GWAS from IR-related traits and tissues. Of the loci analyzed, ~25% showed colocalization with a single gene, but many loci lacked a plausible causal gene. Several reasons could explain this, including: i) underpowered QTL datasets, ii) relevant lncRNA, or other QTL types, have not been well characterized, iii) larger GWAS are needed, iv) better partitioning (LD blocks) of disease-associated loci are needed to distinguish independent GWAS signals.

To fill these gaps and study the shared genetic architecture of IR and cardiometabolic diseases we build on our prior work by including the latest large scale multi-ethnic GWAS for cardiometabolic traits and diseases (i.e. T2D, CAD and NAFLD-20 GWAS in total), incorporating millions of additional subjects. We are expanding our QTL tissue coverage (subcutaneous and visceral adipose, skeletal muscle, liver, pancreas) to include the vasculature and generating more extensive QTL maps (eQTL and sQTL) by merging GTEx and STARNET, which has > 600 samples for tissues of interest and includes annotations for lncRNA QTLs. To leverage size and power differences between tissue QTL datasets, we integrate tissue-specific QTLs maps generated from the recently developed FastGxC method. We also apply the recently generated LD blocks from 1000 Genomes phase 3 data to achieve a better partition of disease-associated loci.

We have so far found that a number of loci (~50 genes) are colocalized with IR and cardiometabolic diseases. Approximately 40% of the identified genes are not shared with traditional risk factors. Among the identified shared colocalized genes are *CCDC92*, which has been previously identified as a cardiometabolic gene, and other novel causal gene candidates. In summary, we have created a comprehensive list of prioritized IR/cardiometabolic disease candidate causal genes and assigned actionable tissues for targeted validation.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1857 Untangling the role of heritable individual-level exposures in type 2 diabetes risk: how genetics impacts our unhealthy behaviors.

Authors:

N. Carvalho[1,2], C. J. Patel[2]; [1]Georgia Inst. of Technology, Atlanta, GA, [2]Harvard Med. Sch., Boston, MA

Genotypes, behavior, and environmental exposures form a complex web of associations with each other. Epidemiological research that assumes independence between these factors can therefore confound true associations with disease risk. Expanding on prior work, we built a polyexposure score (PXS) for type 2 diabetes (T2D) risk composed of 25 independently associated individual-level behaviors in the UK Biobank. In the following, we estimated the PXS for T2D to have a modest but significant heritability ($h^2 = 0.119$); TV consumption and alcohol intake were among the most heritable and risk-associated individual-level behaviors. Genes associated with PXS for T2D were expressed in the brain, suggesting that the genetic architecture of the PXS captures cognitive features. We found that the PXS also enhances dissection of diabetes etiology: for example, we found distinct genetic correlations between the PXS and other metabolic traits such as BMI and glucose control. For example, BMI is found to have a higher genetic correlation with the PXS for T2D onset ($r_g = 0.57$) than T2D onset itself ($r_g = 0.42$), suggesting the existence of risk behaviors associated with insulin resistance rather than production. Furthermore, PXS and polygenic risk scores for T2D are largely independent from each other (are weakly correlated) on an individual-level but have a weak association when looking across UK regions. We discuss potential causal pathways for diabetes risk and highlight the importance of studying malleable disease-associated behaviors and their genetic architectures in order to identify at-risk groups and effective preventative treatments.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1858 Untangling the shared genetic architecture between Alzheimer's and cardiometabolic disease.

Authors:

C. Cunha, M. Ureña, T. Kilpeläinen, R. J. F. Loos; Novo Nordisk Fndn. Ctr. for Basic Metabolic Res., København, Denmark

Obesity and type 2 diabetes (T2D) associate with an increased risk of late-onset Alzheimer's Disease (LOAD). However, the genetic relationship between these conditions remains unclear. The increase in sample size of LOAD GWAS' has been largely driven by proxy cases based on parental diagnoses (self-reported impression of off-springs that either parent had dementia) and non-age matched controls (referred to as controls hereafter) in the UK Biobank (UKBB), which may lead to misclassification. We aimed to assess the genetic relationship between obesity, T2D, and LOAD with and without proxies and controls by genetic correlation and cross-phenotype association analyses. We leveraged the largest European-ancestry GWAS summary statistics for BMI, type 2 diabetes (T2D) and LOAD (w/ UKBB proxies and controls), LOAD (w/o UKBB proxies, and LOAD (w/o UKBB proxies and controls). We used LDSC to estimate genetic correlation (GC) between BMI, T2D and LOAD and performed cross-phenotype variant discovery by identifying variants associated with obesity-LOAD, T2D-LOAD, or all three outcomes. The GC between BMI and LOAD weakened when removing proxies for LOAD (-0.07 vs -0.13), while the T2D-LOAD GC estimates also faded when proxies (0 vs -0.09) and controls (-0.03 vs -0.09) were removed. Further, we identified 150 genome-wide significant (GWS) loci for LOAD. Removing proxies and proxies/controls yielded a loss of 17 and 20 loci, respectively, highlighting that proxies create a more heterogenous LOAD phenotype relating more broadly to dementia. The lost loci are likely correlated with dementia, but unspecific to LOAD. Importantly, the *APP* gene - central to the prevailing amyloid cascade hypothesis in LOAD - was not found when removing proxies and controls. We report 73 and 337 independent signals shared between BMI-LOAD and T2D-LOAD, respectively. Removing proxies and controls led to a loss of ≈80% in overlapping GWS hits between our trait-pairs, likely highlighting novel and "truer" variants mediating cardiometabolic disease and LOAD. We establish *INO80E* (DNA and chromatin biology), *DOC2A* (Ca^{2+}-dependent neurotransmission), *ACE* (blood pressure regulation), *NECTIN2* (T-cell signaling) and *SLC39A13* (zinc transport) as strong mediators of all BMI-T2D-LOAD/dementia. Little is known about the biological function of these genes. In conclusion, we show that the use of proxies and controls leads to heterogeneity in cross-cardiometabolic genetic risk assessment and identify several genetic loci that may mediate the relationship between cardiometabolic disease and LOAD. Our findings provide new insights into the cardiometabolic genetic risk towards Alzheimer's Disease.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1859 Unveiling cell-type-specific molecular signatures in autoimmune diseases

Authors:

T-M. Ko; Natl. Yang Ming Chiao Tung Univ., Taiwan, Hsin-chu, Taiwan

Objective: Systemic lupus erythematosus (SLE) presents a challenge in identifying viable treatment targets due to cellular heterogeneity within immune profiling. Overcoming these challenges by addressing cellular heterogeneity can greatly aid in the identification of exploitable targets for SLE treatment. Methods: In this study, we employed single-cell sequencing and integrative analysis of single-cell transcriptomics to explore cell-type-specific molecular targets associated with SLE within the heterogeneous circulating immune cells across different demographic groups. Results: Through comprehensive computational analysis of single-cell data, we successfully identified potential cell-type-specific molecular targets that are expressed in a particular T-cell subpopulation in SLE. By integrating trajectory inference and intercellular communication profiling, we observed that the specific T cells were activated through OX40 signaling interaction with NK cells in SLE patients. Furthermore, we confirmed the significance of these cell-type-specific molecular targets in SLE by demonstrating that the JAK-STAT signaling pathway predominates as the differential pathway involved in T-cell activation. These findings were consistent across single-cell transcriptomics datasets from two independent adult SLE cohorts, encompassing both Asian and non-Asian populations. Conclusions: Our study highlights the utility of integrative analyses of single-cell multi-omics in guiding and monitoring therapy for lupus patients, especially those undergoing treatment with biologics, particularly JAK inhibitors. By addressing cellular heterogeneity and identifying cell-type-specific molecular signatures, this approach holds promise for the advancement of targeted therapies in SLE and potentially other autoimmune diseases.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1860 Unveiling the resilience effect in high-risk Alzheimer's Disease individuals at single-cell resolution

Authors:

Y. Dai[1], X. Li, Female[2], B. S. Fernandes, Female[3], Z. Zhao[4]; [1]UTHlth., Houston, TX, [2]Univ. of Texas HSC at Houston, Houston, TX, [3]UTHlth.at Houston, Houston, TX, [4]Univ Texas HSC Houston, Houston, TX

Background: Alzheimer's disease (AD) is a progressive and debilitating neurological disease that currently affects millions of people in the US and worldwide with enormous societal and economic costs. Despite extensive research efforts, the pathophysiology of AD remains poorly understood, and there is no known cure for the disease. A considerable proportion of elderly people carry a high genetic AD risk, as evidenced by Polygenic Risk Scores (PRS), but evade AD, thus showing a resilience factor to AD. Our previous work also demonstrated that genetic components play their roles in the resilience effect. With the advent of single-cell technology, we have been able to observe the resilient effect of AD in those with cognitively normal (CN) individuals who possess a high PRS load for AD. **Methods:** We adapted single-nucleus RNA-seq data from the Synapse portal (syn2580853), which contains 454 individuals from Religious Orders Study/Memory and Aging Project (ROS/MAP). We obtained the matched whole-genome-sequence (WGS) data from the Synapse portal (syn11724057). LDPred2 was used to estimate the individual's risk by adapting the effect size from Wightman et al., 2022 GWAS summary statistics. We stratified high-risk individuals from the snRNA-seq data and conducted differentially expressed genes (DEGs) analyses for each cell type among high-risk AD and high-risk CN individuals. Lastly, the functional enrichment analysis was applied to identify enriched gene ontology (GO) terms and canonical pathways. **Results:** After we obtained the 407 individuals with matched snRNA-seq and WGS data, we applied LDPred2 to estimate their AD risks and further stratify the high-risk individuals for 20 quantile of top risks, which includes 53 high-risk AD, 15 high-risk CN, as well as 16 mild cognitive impairment (MCI) individuals. We merged the snRNA-seq data and identified distinct distribution differences among cell AD, MCI, and CN in a few cell clusters, such as oligodendrocytes and excitatory neurons. After, we conducted a pair-wise DEG analysis between CN, MCI, and AD. The most significant DEGs with FDR correction are from excitatory neurons (2304 DEGs), oligodendrocytes (1804 DEGs), and microglia (934 DEGs), in AD and CN groups. Interestingly, we found that MCI individuals tend to have expression profiling more similar to AD rather than to CN individuals. The gene set enrichment analyses highlighted a few well-known pathways tau protein binding, and response to oxidative stress, as well as a few new epigenetic modifications and metabolite pathways that related to histidine metabolism and thiolester hydrolase activity.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1861 Using multiple approaches to identify the novel role of trafficking mechanisms underlying epilepsy

Authors:

T. Nagai[1], E. Knapik[2], E. Gamazon[2,3]; [1]Vanderbilt Univ., Nashville, TN, [2]Vanderbilt Univ. Med. Ctr., Nashville, TN, [3]Clare Hall, Univ. of Cambridge, Cambridge, United Kingdom

Epilepsy is a neurological disease that affects millions of people worldwide. The current approach to treating epilepsy focuses on using antiepileptic drugs (AEDs) to control seizure activity. However, despite the availability of various AEDs that have been approved over the past decade, a significant proportion of epilepsy cases remain drug resistant. The development of AEDs has primarily focused on proteins that are direct components excitatory and inhibitory synapses, limiting the targets through which these drugs can reduce seizures. Expanding the repertoire of mechanisms targeted by AEDs would greatly benefit patients with drug-resistant epilepsy. In our study, we employed two different approaches: population genetics and patient data analysis, as well as zebrafish seizure models, to identify and functionally validate novel genes associated with epilepsy. By utilizing a transcriptome-wide association study (TWAS) method called PrediXcan, we identified AP3D1, a component of the AP3 trafficking complex, as being associated with epilepsy. Additionally, analysis of novel rare variants revealed the potential involvement of SEC24C, a protein associated with the COPII protein trafficking complex, in seizure activity. To further investigate these genes we utilized CRISPR-Cas9 to generate zebrafish models with disrupted AP3D1 and SEC24C activity. Our behavioral assay using pentylenetetrazole (PTZ), a seizure-inducing agent, showed that loss of AP3D1 and SEC24C function increased seizure susceptibility in the zebrafish models. Furthermore, analysis using transgenic GCAMP6s zebrafish demonstrated significant alterations in brain activity caused by the disruption of these trafficking proteins, both in a rested state and under PTZ-induced conditions. By combining population genetics, patient data analysis, and zebrafish models, our study provides novel insights into the complex genetic and functional aspects of epilepsy. These findings highlight the crucial role of trafficking mechanisms in seizure susceptibility and epilepsy. Targeting these trafficking proteins could potentially lead to the development of new therapeutic strategies for drug-resistant epilepsy.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1862 † Using plasma proteomics and Mendelian randomization to identify causal proteins for neurological disease

Authors:

L. Belbasis, C. Van Duijn, D. Bennett, R. Walters; Univ. of Oxford, Oxford, United Kingdom

Neurological diseases impose a large economic burden and are a major public health issue. Observational and genome-wide association studies (GWAS) have identified risk factors and genetic variants positively associated with risk of neurological diseases. However, the role of circulating proteins in the development of neurological diseases is under-studied. We have applied a two-sample Mendelian randomization (MR) approach to explore the causal relevance of the plasma proteome to a range of neurological diseases. Non-overlapping *cis* protein quantitative trait loci (pQTLs) were identified from three large proteogenomics studies in populations of European ancestry, for use as instruments for protein abundance in two-sample MR. These were tested for association with neurological diseases in the latest publicly-available GWAS for Alzheimer's disease (AD), Parkinson's disease (PD), Amyotrophic Lateral Sclerosis (ALS) and Multiple Sclerosis (MS). Out of 2,178 unique *cis* pQTLs and 7,788 tested protein-disease associations, 91 associations for 86 proteins were significant at 5% false discovery rate (FDR). Thirty-seven proteins were associated with AD risk, 7 with PD risk, 8 with ALS risk and 39 with MS risk. Five out of the 86 proteins were associated with AD and an additional neurological disease. Enrichment analysis for each outcome for predicted protein-protein interactions (PPIs) in the STRING database identified a greater degree of interactions than expected for proteins associated for both AD and MS. Thirty-four (92%) of the proteins associated with AD and 32 (82%) of the proteins associated with MS were part of a single PPI network. We further tested association of the 86 proteins with brain imaging phenotypes and risk factors of neurological diseases using the two-sample MR approach. Out of 704 protein-brain imaging tests and 371 protein-risk factor tests, 5 and 36 were significant at 5% FDR, respectively. This comprehensive assessment of associations between circulating proteins and neurological diseases identifies proteins that may be potential drug targets for disease management, biomarkers for disease prediction, or mediators of risk factor-disease associations. Ongoing work includes testing for colocalisation of protein and disease GWAS signals, and seeking replication in populations of non-European ancestry.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1863 Using rare genetic mutations to revisit structural brain asymmetry

Authors:

J. Kopal[1], S. Jacquemont[2], C. Moreau[3], D. Bzdok[4], K. Kumar[5], G. Huguet[6]; [1]CHU Saint Justine, Montreal, QC, Canada, [2]CHU Sainte justine Res. Ctr., Montreal, QC, Canada, [3]Canada, [4]McGill Univ., Montreal, QC, Canada, [56]CHU Ste justine, Montréal, QC, Canada

Asymmetry between the left and right brain is a key feature of brain organization. Hemispheric functional specialization underlies some of the most advanced human-defining cognitive operations, such as articulated language, perspective taking, or rapid detection of facial cues. Yet, genetic investigations into brain asymmetry have mostly relied on common variant studies, which typically exert small effects on brain phenotypes. Here, we leverage rare genomic deletions and duplications to study how genetic alterations reverberate in human brain and behavior. We quantitatively dissected the impact of eight high-effect-size copy number variations (CNVs) on brain asymmetry in a multi-site cohort of 552 CNV carriers and 290 non-carriers. Isolated multivariate brain asymmetry patterns spotlighted regions typically thought to subserve lateralized functions, including language, hearing, as well as visual, face and word recognition. Planum temporale asymmetry emerged as especially susceptible to deletions and duplications of specific gene sets. Targeted analysis of common variants through genome-wide association study (GWAS) consolidated partly diverging genetic influences on the right versus left planum temporale structure. In conclusion, our gene-brain-behavior mapping highlights the consequences of genetically controlled brain lateralization on human-defining cognitive traits.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1864 Using RNA-sequencing to investigate the rhythmicity of chronic low back pain

Authors:

A. Zacharias[1], H. Gowdy[1], M. Knezic[1], D. Taccardi[1], E. Bisson[2], Q. Duan[1,3], N. Ghasemlou[1,4,5]; [1]Dept. of BioMed. and Molecular Sci., Queen's Univ., Kingston, ON, Canada, [2]Kingston Hlth.Sci. Ctr., Kingston, ON, Canada, [3]Sch. of Computing, Queen's Univ., Kingston, ON, Canada, [4]Dept. of Anesthesiology & Perioperative Med., Queen's Univ., Kingston, ON, Canada, [5]Ctr. for NeuroSci. Studies, Queen's Univ., Kingston, ON, Canada

Chronic low back pain (CLBP) is highly prevalent and contributes to the most years lived with disability worldwide. CLBP has been shown to vary in intensity throughout the day; however, these effects remain unclear at the molecular level. The aim of this study is to investigate how pain rhythmicity affects outcomes and its underlying mechanisms. Based on self-reported daily pain rhythmicity, 60 participants expressed 4 phenotypic pain patterns: constant low, constant high, rhythmic↑, and mixed. We sought to identify gene expression differences across the 4 clusters using RNA sequencing of peripheral blood cell samples taken at 8:00 and 20:00 (n=116). Sequencing reads were processed with *FastQC*/*MultiQC, Hisat2*, and *StringTie*. Normalization and outlier detection were performed using *edgeR* and *arrayQualityMetrics.* Transcripts were kept if their median absolute deviation was ≥ the 70th quantile (0.054), leaving 104,246 remaining transcripts. *edgeR* identified differentially expressed transcripts (DETs; Bonferroni-adjusted p-values < 0.05) between the rhythmic↑ participants vs. any other phenotypes, adjusted for sequencing batch, sex, and participant ID. Next, we used weighted gene co-expression analysis (*WGCNA* in R) to find co-expressed transcripts associated with pain rhythmicity. Finally, we performed pathway enrichment analysis on DETs and transcripts modules with *gprofiler2*. We identified 214 differentially expressed transcripts between rhythmic↑ and any other phenotype. These transcripts are enriched in the REAC neutrophil degranulation and GO:CC azurophil granule categories (adj. p-value < 0.05). Furthermore, 55 transcripts were differentially expressed in blood collected at day versus night in the rhythmic↑ cohort. These transcripts are enriched for GO categories related to podosomes and regulation of double-strand break repair via homologous recombination (adj. p-value < 0.05). Additionally, we identified 79 and 68 modules of co-expressed transcripts in the day and night networks, respectively. Of these modules, 2 day and 4 night modules were associated with the rhythmic↑ phenotype (GLM p-value < 0.05). These clusters were enriched in T-cell migration, neutrophil degranulation, and mRNA modification pathways (adj. p-value < 0.05). Results from our study suggest that neutrophils, among other immune cells, are important regulators of pain rhythmicity in people with CLBP. This supports previous findings that neutrophil activation contributes to pain chronicity and suggests that neutrophils may be a potential target for intervention in CLBP.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1865 Uterine fibroids causes and consequences: Using mendelian randomization to establish directionality between fibroids and associated genitourinary and neoplasm phecodes

Authors:

J. Piekos[1], H. M. Seagle[1], J. N. Hellwege[2], N. K. Khankari[2], T. L. Edwards[2], D. R. Velez Edwards[2]; [1]Vanderbilt Univ., Nashville, TN, [2]Vanderbilt Univ. Med. Ctr., Nashville, TN

Uterine fibroids (UF) are benign monoclonal tumors of the uterine myometrium that are present in 70 - 80% of biological females. Despite being the most common pelvic tumor in females, research has yet to fully uncover the etiology and the downstream consequences of UF. Previously, we have performed a phenome-wide association study (PheWAS) in Vanderbilt's BioVU to identify phecodes associated with UF in the clinical phenome. We identified 24 and 28 unrelated phecodes in the genitourinary and neoplasm categories respectively to focus on. Here was assign directionality to these relationships using two sample Mendelian Randomization (MR). Genetic instruments for MR were selected from the Finngen Cohort for UF and the genetic instruments for the tested associated phecodes were selected from the UK BioBank. SNPs for analysis were selected by linkage disequilibrium clumping genome-wide significant ($p < 5x10^{-8}$) variants using an r^2 threshold = 0.01. With UF as exposure, 10 genitourinary and 12 neoplasm phecodes were significant consequences of UF (P value < 0.05). The top phenotypes identified as consequences of genetic UF exposure were endometriosis OR = 1.30 (95% CI = 1.22 - 1.38, P value = $3.05x10^{-17}$) and ovarian cysts OR = 1.20 (95% CI = 1.14 - 1.26, P value = $1.81x10^{-12}$) from genitourinary and lipoma OR=1.16 (95% CI = 1.10 - 1.22, P value = $1.01x10^{-7)}$ and benign neoplasm of skin OR = 1.12 (95% CI = 1.07 - 1.16, P value=$1.19x10^{-7}$) from the neoplasm category. Two genitourinary phenotypes were found to be causal towards UF with MR. They were hemangioma and lymphangioma OR = 1.15 (95% CI: 1.10 - 1.21, P value=$1.86x10^{-10}$) and polyp of female genital organs OR = 1.34 (95% CI = 1.11 - 1.60, P value = 0.002). We observed horizontal pleiotropy between UF and breast conditions, congenital or related to hormones OR = 1.40 (95% CI = 1.13 - 1.74, P values = 0.003) using the MR Egger method. The egger intercept of the model was significant (P value = 0.01). Understanding the directionality of the relationships between UF and other phenotypes can be leveraged to help understand the etiology of disease, identify novel risk factors, assist in prioritizing females of UF interventions, and intervene to prevent downstream consequences.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1866 Utilizing genetically determined retinal thickness to link to the risk and Age-at-Onset of Alzheimer's disease.

Authors:

Q. Fan[1], J. La[2], X. Liu[3], E. Martin[4], Y-J. Li[2]; [1]Ctr. for Quantitative Med., Duke-NUS Med. Sch., Singapore, Singapore, [2]Duke Univ., Durham, NC, [3]Dept. of Statistics and Data Sci., Natl. Univ. of Singapore, Singapore, Singapore, [4]Univ. of Miami Miller Sch. of Med., Miami, FL

Objective: Retinal thickness, as measured by optical coherence tomography (OCT), has emerged as a potential non-invasive biomarker for Alzheimer's disease (AD). This study aims to investigate the association between genetically determined individual retinal layer thickness with both risk and age-at-onset (AAO) of AD. **Methods:** Genome-wide association studies were performed for 10 retinal thickness measurements obtained from spectral-domain OCT, including the retinal nerve fiber layer (RNFL), inner nuclear layer (INL), ganglion cell-inner plexiform layer (GCIPL), ISOS-RPE layer, and others from 67,280 European participants with both retinal layer thickness and genotype data from UK Biobank. The pooled AD dataset from 20 ADGC cohorts was used as the testing dataset (9,219 cases and 19,564 controls) (Li et al. 2023). Polygenic scores (PGS) for each retinal layer thickness were constructed using PRSice. The Association of PRS with AD risk was tested by the multivariable generalized linear mixed model. For AAO of AD, a multivariable linear mixed model was performed. Both models adjusted for sex, (age), and population structure (principal components 1-10) and accounted for the cohort with a random intercept. A significance threshold of P< 0.005 was used. **Results:** The PRS for thinner RNFL showed the strongest association with an increased risk of AD (OR(95% CI)=1.05 (1.02, 1.08), P=9.9E-04) and nominal association with decreasing AAO of AD (P=0.028). Interestingly, increasing INL thickness showed an association with early AAO of AD (beta(SE)=-0.24(0.07); P=9.3E -04) but not with the risk of AD (P=0.145). The thinner photoreceptor layer and ISOS-RPE layer were associated later onset of AD (beta(SE)=0.20(0.07), P=0.005) but the association with AD risk was not strong (P=0.018). **Conclusion:** Overall, genetic predictors of thickness in RNFL, INL, and ISOS-RPE layers demonstrated the most significant associations with AD or AAO of AD. These findings support the potential utility of retinal layer thickness as a non-invasive biomarker for AD risk assessment and early detection.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1867 Utilizing Long-read Sequencing to Decipher the Phenotypic Heterogeneity in Motor Neuron Diseases

Authors:

A. Jain[1,2], E. Udine[1,2], N. A. Finch[1], J. Wuu[3], A. Cooley[3], J. P. Hernandez[3], M. DeJesus-Hernandez[1], A. White[1], E. P. Pioro[4], J. Heckmann[5], T. Burns[6], J. Statland[7], J. Caress[8], J. Trivedi[9], J. Ravits[10], C. Jackson[11], L. Elman[12], C. McMillian[12], A. Swenson[13], S. Maiser[14], J. McCauley[3], R. Rademakers[1], G. Wu[15], E. Rampersaud[15], J. Myers[16], J. Taylor[17], M. Benatar[3], M. van Blitterswijk[1]; [1]Mayo Clinic, Jacksonville, FL, [2]Mayo Graduate Sch. of BioMed. Sci., Jacksonville, FL, [3]Univ. of Miami Miller Sch. of Med., Miami, FL, [4]Nothwestern Univ., Cleveland, OH, [5]Univ. of Cape Town, Cape Town, South Africa, [6]Univ. of Virginia, Charlotteville, VA, [7]Kansas Univ. Med. Ctr., Kansas city, KS, [8]Wake Forest Sch. of Med., Winston-Salem, NC, [9]UT Southwestern Med. Ctr., Dallas, TX, [10]Univ. of California San Diego, San Diego, CA, [11]Univ. of Texas Hlth.Sci. Ctr. San Antonio, San Antonio, TX, [12]Univ. of Pennsylvania, Philadelphia, PA, [13]Univ. of Iowa, Iowa City, IA, [14]Hennepin County Med. Ctr., Minneapolis, MN, [15]St Jude Children s Res. Hosp., Memphis, TN, [16]St. Jude Children's Res. Hosp., Memphis, TN, [17]St Jude Children's Res Hosp, Memphis, TN

Identifying determinants of selective upper motor neuron (UMN) and lower motor neuron (LMN) pathology is crucial for understanding the phenotypic variability in motor neuron diseases and predicting the disease course and outcome. While classic amyotrophic lateral sclerosis (ALS) is known for its variable degrees of UMN and/or LMN dysfunction, primary lateral sclerosis (PLS) and progressive muscular atrophy (PMA) represent extreme phenotypes characterized by pure UMN and LMN involvement, respectively. Given the role of splicing defects in motor neuron diseases, we obtain detailed RNA profiles by capturing full-length transcripts through long-read RNA sequencing. In this study, we used blood specimens from age- and sex-matched individuals clinically diagnosed with PLS (n=2), classic ALS (n=20), or PMA (n=17) from the PGB (Phenotype-Genotype-Biomarker, NCT02327845) study of the CReATe (Clinical Research in ALS and Related Disorders for Therapeutic Development) Consortium. RNA was extracted using the Qiagen PAXgene Blood RNA extraction kit, and specimens with an RNA integrity number (RIN) of more than 7 were included. Long-read RNA sequencing (Iso-Seq method) was performed on the Sequel II platform (Pacific Biosciences) at Mayo Clinic's Genome Analysis Core. Transcriptomic data was analyzed using a custom workflow that incorporates various tools (pbmm2, isoseq3, cDNA cupcake, SQANTI3, Integrative Genomics Viewer, etc.). In total, we obtained roughly one million reads per individual. The average read length was about 2.5 kb, demonstrating our ability to produce long reads. We detected over 18,000 genes with more than 300,000 unique transcripts. Transcripts were classified into major structural categories including full-splice matches, incomplete splice matches, and novel transcripts, among others. Our preliminary analysis of over 50 ALS-linked genes showed significant differences (e.g., *C9orf72* and *TARDBP*), providing insight into drivers of UMN and/or LMN dysfunction. Therefore, our novel long-read sequencing approach in a well-characterized clinical cohort of patients with PLS, classic ALS, and PMA should assist in identifying some of the biological underpinnings of the phenotypic heterogeneity of these diseases, which vary not only in the relative degree of UMN versus LMN pathology, but also with respect to rates of progression and survival. This can facilitate the search for much-needed biomarkers and/or disease modifying therapies.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1868 Validation study of the AD8 and Clinical Dementia Rating Scale for a Samoan speaking population.

Authors:

R. Hansen[1], J. Tavana[1], B. Vaitohi[1], C. S. Pakalani[1], T. D. Packard[1], A. T. Perez[1], F. Matthes[1], S. Polaapau[1], S. F. Funaki[1], T. N. G. Tavana[1], J. T. Tschanz[2], J. S. K. Kauwe[3], P. G. Ridge[1]; [1]Brigham Young Univ., Provo, UT, [2]Utah State Univ., Logan, UT, [3]Brigham Young Univ, Provo, UT

Alzheimer's Disease (AD) is the most common form of dementia, yet diagnosing AD is challenging, especially for individuals who are not native English speakers. Approximately 250,000 Samoans live in the USA (including many in American Samoa, an American territory) and are the second fastest-growing race in the USA. Yet, Samoans are underrepresented in virtually every medical dataset, including AD datasets. The Clinical Dementia Rating Scale (CDR) and the AD8 are standard tools for staging dementia. Many Samoans are bilingual but speak Samoan more comfortably than English, especially the elderly. There are no Samoan language AD diagnostic tools, and English-based assessments require a nuanced understanding of English, which is challenging for non-native speakers. This work aimed to develop Samoan versions of the CDR and the AD8.

A professional translator translated the CDR and AD8 into Samoan. We administered the tests to a focus group of 10 Samoans to evaluate cultural relevance and clarity, and measured the semantic equivalence to the validated English versions of the CDR and AD8. Finally, a second professional translator and bilingual psychologist back-translated the tests. Each of the validation steps supported the quality of the translations.

Next, we are determining if the Samoan tests accurately determine dementia status. We recruited 50 Samoans, including bi- and monolingual Samoans. The AD8 is complete for all participants, and the CDR and NACC (National Alzheimer's Coordinating Center) neuropsychological battery assessments are ongoing. A neuropsychologist and neurologist independently evaluate the assessments to determine the presence or absence of dementia. Consistent outcomes from the different tools and the evaluators confirm the accuracy of the Samoan language versions. So far, our data demonstrate that the Samoan versions are effective. We will present the final results at the conference.

Our broad objective is to build more ethnically diverse AD cohorts and reduce the health disparities in Samoans and Pacific Islanders. This study is an essential first step.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1869 Variant-enhancer-gene mapping at the 9p21 locus in smooth muscle cells reveals new enhancer-gene pairs for coronary artery disease

Authors:

M. Ramste[1], C. Weldy[1], D. Li[1], D. Sharma[1], A. Ranta[1], D. Amgalan[2], E. Jagoda[3], G. Qin[1], A. Gschwind[4], T. Nguyen[1], J. Engreitz[5], C. Park[1], T. Quertermous[1]; [1]Div. of Cardiovascular Med., 300 Pasteur Drive, Falk CVRC, Stanford, CA 94305, USA, Stanford, CA, [2]Dept. of Genetics, Stanford Univ. Sch. of Med., Stanford, CA, USA, Stanford, CA, [3]Harvard Univ, Cambridge, MA, [4]Dept. of Genetics, Stanford Univ. Sch. of Med., Stanford, CA, USA, Stanford, Switzerland, [5]3Dept. of Genetics, Stanford Univ. Sch. of Med., Stanford, CA, USA, Stanford, CA

The majority of variants identified by genome-wide association studies (GWAS) that influence coronary artery disease (CAD) risk reside in noncoding regions of the genome. Thus, it is challenging to link them with the genes they regulate and the gene regulatory networks that they form. A deeper understanding of how genetic variation in noncoding regions, and especially in enhancers, can be linked to causal genes will help uncover fundamental mechanisms in coronary artery disease. Enhancers are cell type specific, and vascular smooth muscle cells (SMC) are known to have the highest heritable risk in CAD and play a major role in the atherosclerotic plaque formation. The 9p21.3 locus is the most impactful genetic risk locus for CAD. Although some studies have shown that SMC are the causal cell type for CAD risk at this locus, the causal gene and molecular mechanism of genetic risk are poorly understood.

Here, we report efforts to systematically map SMC specific enhancers to neighboring genes within the 9p21.3 locus in human immortalized SMC by implementing single cell CRISPRi enhancer screens.First, we intersected CAD GWAS loci with human coronary artery SMC (HCASMC) ATAC-seq and H3K27ac ChIP-seq datasets to focus on the disease relevant SNPs that fall within active enhancer regions in SMC. This analysis identified 27 SNPs in 11 enhancers, and single guide RNAs were designed to target the summit of each enhancer using the CRISPRiTiling tool. Pooled lentiviral guides, including negative and positive controls were introduced into animmortalized HCASMC line stably expressing dCas9-KRAB. Perturbed cells were collected at 5- and 10- days post transduction. To enable detection of weak effects and lowly expressed genes, we employed the targeted Perturb-seq (TAP-seq) approach for library generation and sequencing.

We identified several enhancer-gene -pairs, including a strong enhancer-gene connection to both *CDKN2A* and *CDKN2B*. Additionally, we identified multiple enhancer regions that control *MTAP* expression, with smaller but significant effects. We followed up with individual validation of enhancer-gene pairs through qPCR. Furthermore, these results are consistent with chromosomal interaction data obtained from our previous HiChIP.Our results identify new variant to gene links and how the genetic risk in 9p21 is mediated in the vascular wall, provides new understanding of enhancer-gene pairs regulating *CDKN2A* and *CDKN2B*, suggests a novel mechanism of 9p21.3 disease risk through *MTAP* and provides an important discovery platform through which CRISPRi screens can be used to discover causal mechanisms of disease.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1870 Variants Detected in Mitochondrial DNA D-loop in Iranian Patients with Non-alcoholic Fatty Liver Disease

Authors:

S. Kamfar; Pediatric Congenital Hematologic Disorders Res. Ctr., Res. Inst. for Children's Hlth., Shahid Beheshti Univ. of Med. Sci., Tehran, Iran, Tehran, Iran, Islamic Republic of

Background: Non-alcoholic fatty liver disease (NAFLD) is a growing chronic liver disease with unknown etiology which in advanced stages, can be life-threatening. There is a substantial body of evidence that mitochondrial dysfunction due to its impact on hepatic lipid homeostasis, cytokine release, ROS production, and cell death, contributes to the pathogenesis of NAFLD. Therefore, we aimed at comparing mitochondrial D-loop sequence variations in liver tissue samples of Iranian patients with NAFL and NASH. **Methods**: Liver tissue samples from 38 NAFLD patients (16 patients with NAFL and 22 patients with NASH) were collected. Polymerase chain reaction and direct sequencing were performed to detect sequence variation in two hypervariable regions, HVR1 and HVR2 of the mtDNA D-loop region. Sequences of all individuals were compared with the MITOMAP Database. **Results**: Overall, 87 somatic mutations (32 mutations in HVR2 and 55 mutations in HVR1) were identified in mtDNA D-loop, including 85 single-base substitutions, two deletions, and one single nucleotide insertion. The result showed that the variant numbers in NAFLD patients ranged between 1-14. In this case, the comparison between NAFL and NASH groups showed that the numbers were higher in the NASH group (109 mutations) compared with NAFL (101 mutations), but this difference was not significant ($P > 0.05$). C to T and T to C transitions were the most prevalent single-base substitutions among patients (31 and 38 single-base substitutions, respectively, in the NAFL group). C258T, C271T, A335G, C431A, T482C, C497T, C530T, T16093C, C16111T, A16127C, A16160G, A16163G, A16171T, C16186T, C16218T, A16220C, A16227G, C16239T, C16260T, C16291T, T16298C, T16304C, C16344T, T16362C, G16390A were seen only in NASH group. Furthermore, A16180del and G16035C were detected as novel mutations in the NAFL group and 16220-16221ins C in both groups. Also, no important difference was found in mtMSIs, including D310, mt514-523 (CA)n between NAFL and NASH groups. **Conclusions:** Our results indicate that 16220-16221ins C could be a candidate as a potential new biomarker for NAFLD diagnosis.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1871 Variation in mushroom body morphology in cocaine preferring Drosophila Genetic Reference Panel lines

Authors:

A. Ummet, T. Mackay, R. Anholt; Clemson Univ., Greenwood, SC

The mushroom body is a central structure in the Drosophila brain which plays a key role in sensorimotor integration and experience dependent behaviors, including olfactory behavior and memory formation. Kenyon cells and their projections are organized as the alpha, beta, and gamma lobes of the mushroom bodies. Variation in the structure of mushroom body lobes correlates with variation in multiple traits, including sleep, aggression, and lifespan. In this study, I am investigating how naturally occurring genetic variation among lines of the Drosophila Genetic Reference Panel results in variation in mushroom body morphology and whether this variation is correlated with cocaine preference. I use an anti-fasciclin II antibody to visualize the mushroom body and ellipsoid body by confocal microscopy in 20 brains for each genotype, sexes separately. Analysis of three-dimensional reconstructions of mushroom bodies from three cocaine preferring lines and three controls reveal variation in dimensions of the mushroom body lobes. Together with transcriptomic analyses this study will give insights into the relationship between genomic variation, variation in gene expression, brain morphology, and cocaine preference behavior. Based on evolutionary conservation of fundamental cellular processes, insights from the fly model may lead to a deeper understanding of the neurogenetic underpinnings responsible for cocaine addiction in human populations.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1872 Vitamin D genetics: inclusion of residence-based UV exposure reveals novel independent loci in genome-wide association and interaction analyses.

Authors:

R. Shraim[1], J. van Geffen[2], M. van Weele[2], R. McManus[1], L. Zgaga[1]; [1]Trinity Coll. Dublin, Dublin, Ireland, [2]Royal Netherlands Meteorological Inst., De Bilt, Netherlands

Sun exposure has long been established as the major determinant of vitamin D levels in humans. Recent genome-wide studies have identified up to 143 genetic variants linked to vitamin D status, independently and in interaction with sun exposure. The common approach relies on season of blood sampling, dichotomized as summer/winter, as a proxy for UV exposure. In this study, we explore the aetiology of vitamin D status in the UK Biobank cohort, using an exact measure of ambient UV. To capture UV exposure over a period of time, we calculated a weighted UV dose for each participant based on their residence and date of blood sampling. UVB was isolated as the relevant wavelength for vitamin D production, adjusted for cloud cover and the rate of decay of vitamin D in the body. We first carried out a genome-wide association study (GWAS) of 25-hydroxyvitamin-D (25OHD) in 408,820 White British individuals. In addition to age, sex, supplement intake, and population structure (principal components), the regression model was adjusted for the calculated UVB dose. We then carried out a genome-wide gene-environment interaction analysis (GxE) in the same sample where the model additionally included a GxUVB interaction term. The GWAS results identified 54 novel independent variants significantly associated with 25OHD concentration and the GxE analysis identified 37 novel variants. Downstream functional analysis of these variants supports vitamin D involvement in metabolic pathways and hormone and skin phenotypes. Overall, our findings support the importance of accurate environmental exposure measures and accounting for gene-environment interactions in uncovering the genetic architecture of complex traits like vitamin D status.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1873 *VRK2* is implicated in stuttering, beat synchronization, and language decline.

Authors:

D. Pruett[1], A. Scartozzi[2], J. Eissman[3], Y. Mekki[1], H. Polikowsky[2], T-C. Wang[2], X. Bledsoe[4], S. Kraft[5], E. Gamazon[6], L. Dumitrescu[1], T. Hohman[1], R. Gordon[1], J. Below[1]; [1]Vanderbilt Univ. Med. Ctr., Nashville, TN, [2]Vanderbilt Univ., Nashville, TN, [3]Vanderbilt Memory & Alzheimer s Ctr., Nashville, TN, [4]Vanderbilt Univ., Pleasant View, TN, [5]Wayne State Univ., Detroit, MI, [6]VUMC Clare Hall, Univ. of Cambridge, Nashville, TN

Stuttering is a neurological speech condition characterized by syllable repetitions, prolongations, and involuntarily pauses. Although the etiology of stuttering is unknown, familial and population-based studies show a strong genetic influence on stuttering risk. A recent study performed eight ancestry- and sex-specific GWAS of self-reported stuttering (n = 99,776 cases; 1,123,019 controls) and found *VRK2* to span the top two hits in the European-ancestry male cohort. *VRK2* was also recently implicated in independent GWAS investigating musical beat synchronization, functional language connectivity, and language decline in Alzheimer's Disease (AD). These findings are especially compelling considering the role rhythm perception and language play in stuttering: rhythm discrimination is below average in adults and children who stutter and a recent meta-analysis found that children who stutter exhibited significantly lower language scores.

Given the overlapping GWAS findings implicating *VRK2* and the involvement of rhythm and language in the development of stuttering, our study examines how *VRK2* may be related to speech, language, and musicality phenotypes.

Sentinel SNPs mapping to *VRK2* varied across studies: rs11898834 ($p = 9.990 \times 10^{-10}$) in female AD-related language decline GWAS; rs35609938 ($p = 5.84 \times 10^{-12}$) and rs1040225 ($p = 1.82 \times 10^{-11}$) in the male stuttering GWAS; and rs848293 ($p = 9.23 \times 10^{-18}$) in beat synchronization GWAS. We examined effects at significant SNPs across studies and found alleles associated with increased risk of stuttering were associated with poorer beat synchronization and decreased risk of language decline ($p < .003$ Bonferroni correcting for 4 SNPs across 4 studies). To evaluate the neurophysiological consequences of variation in genetically regulated gene expression of *VRK2*, we utilized NeuroimaGENE, a publicly accessible catalog of neuroendophenotypes derived from transcriptome-wide association studies in 33,000 individuals from UK Biobank. NeuroimaGENE findings revealed that the thalamus and somatosensory cortex, areas involved in processing sensory information, are associated with *VRK2* expression ($p <$ Benjamini-Hochberg correction threshold). Phenome-wide association study of genetically regulated VRK2 expression revealed suggestive associations with neurological and hormonal clinical outcomes, including anxiety, and major depression, and a significant association with congenital anomalies of female genital organs (p = .0000235996). Together, findings from this study may highlight biological mechanisms and implications for VRK2 in speech, language, and musicality related phenotypes.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1874 White matter microstructure associations with polygenic risk score for essential tremor in healthy adults

Authors:

A. Pastor[1,2], M. Medeiros[3,4], Z. Schmilovich[2,4], P. A. Dion[4,2], G. A. Rouleau[1,2], A. Dagher[1,2]; [1]Montreal Neurological Inst.-Hosp., Montreal, QC, Canada, [2]McGill Univ., Montreal, QC, Canada, [3]McGill, Montreal, QC, Canada, [4]Dept. of Human Genetics, Montreal, QC, Canada

Intro

Essential tremor (ET) is a common neurological disorder characterized by involuntary upper limb rhythmic movements. ET has a strong genetic basis that may develop via the additive contribution of risk variants of varying frequencies. About 20% of ET liability can be explained by common variants which are incorporated in polygenic risk scores (PRS) that quantify individual risk level (Liao et al. 2022). Separately, white matter tractography, a technique utilizing diffusion-weighted magnetic resonance imaging (DTI) can be used to map the connectivity and integrity of white matter tracts in the brain. In this study we probe the vulnerability of healthy subjects to ET by investigating the association of white matter microstructure with ET PRS

Methods

We paired genetic and imaging data from the UK Biobank to model the associations of ET PRS on brain structure changes. Individual-level genotyping data as well as neuroimaging measures on 31,386 healthy adults were used to correlate genetic risk scores to brain morphometric outcomes. PRS-CS was used to infer SNP posterior effects, which were then combined to generate a PRS for each subject in the cohort using PLINK. Principal component analysis (PCA) was performed to determine ancestry using the 1000 Genomes Project as reference, and those of European descent were retained to limit population stratification. DTI mean diffusion measures (DM) fractional anysotropy, FA, mean diffusivity, MD, and free-water, FW were obtained across 73 anatomically curated tracts in the ORG fiber clustering white matter atlas (Zhang et al., 2018). We performed multiple regression analysis with glmnet (R) for variable selection and cross-validation, and conducted a white matter to grey matter tract crossing and volume analysis

Results

We observed negative associations of ET PRS with FA in motor tracts including the posterior limb of the internal capsule, the corticospinal tract, the motor region of the corpus callosum and the intracerebellar input and Purkinje tracts. We also found converse positive associations with MD and FW for these tracts. Grey matter associations included mainly subcortical structures and the cerebellum. The bilateral putamen, ventral diencephalon, medulla and brainstem were negatively associated with ET PRS after correction by intracraneal volume. Distinct cerebellar regions were also specifically associated with ET PRS (e.g somato-motor, limbic networks)

Summary

These results suggest an involvement of white matter motor tract associations with ET PRS with prevalence of subcortical regions typically involved in motor processing including known cerebellar tracts belonging to distinct motor networks

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1875 † Whole genome burden testing in 333,100 individuals identifies novel rare non-coding aggregate associations with height that impact *HMGA1* and miRNA expression

Authors:

G. Hawkes[1], R. Beaumont[1], Z. Li[2], R. Mandla[3], X. Li[4], J. Locke[1], N. Owens[1], A. Murray[1], TOPMed Anthropometry - Adiposity Working Group, K. Patel[1], T. Frayling[1], C. Wright[1], A. Wood[1], X. Lin[4], A. Manning[5], M. Weedon[1]; [1]Univ. of Exeter Med. Sch., Exeter, United Kingdom, [2]Dept. of Biostatistics and Hlth.Data Sci., Indiana Univ., Indiana, IN, [3]Programs in Metabolism and Med. And Population Genetics, Broad Inst., Cambridge, MA, [4]Harvard T.H. Chan Sch. of Publ. Hlth., Boston, MA, [5]Massachusetts Gen. Hosp., Boston, MA

Background/Objectives: Most sequence-based association studies for common human phenotypes have focussed on rare variants that reside in the coding regions of the genome. However, the recent release of whole-genome-sequence (WGS) data in 100,000s of individuals from several studies provides an unprecedented opportunity to examine rare, non-coding variants and their contribution towards the genetic architecture of common traits.
Methods: We performed the largest WGS-based analysis for height using 333,100 individuals from three studies: UK Biobank (UKB, N=200,003), TOPMed (N=87,652) and All of Us (N=45,445). We developed a generalised analytical pipeline with the aim of finding novel rare (<0.1% minor-allele frequency) non-coding genetic associations. We tested 75,311,546 variants which had at least 20 carriers in the UK Biobank, and performed 52,749,161 genomic aggregates tests split into gene-centric (e.g. proximal-regulatory) and non-gene-centric (e.g. intergenic-regulatory), and grouped by measures of conservation, constraint and deleteriousness. Finally, we performed a hypothesis-free 2kbp sliding window analysis. We then assessed the capacity of novel non-coding regulatory variants to modulate transcription factor binding through in-silico motif analysis and confirmation of binding potential with ChIP-Atlas.
Results: We observed evidence ($p<1e-10$) for non-coding aggregate associations proximal to *HMGA1* and *GH1*, and an association downstream of *C17orf49* overlapping mRNA sequence , all of which showed evidence of replication in TOPMed and UKB. Variants in the *HMGA1* aggregate were predicted to impact *FOX* and *CTCF* binding, with effects up to 5cm. We show that leading miRNA aggregate variants lie in crucial binding sites for miRNA-497 and miRNA-195, which have previously been linked to chondrogenesis and bind to *IHH*, which regulates bone growth. We additionally identified rare genetic variants within pri-miR-497 which may affect miRNA expression by regulating the miRNA host-gene (*MIR497-HG*). We also observed 30 independent novel rare variants associated with height, after conditioning on more than 13,000 previously reported loci. We observed effect sizes range from -7cm to +2cm, and replicated three rare single variant associations.
Conclusions: Our approach found novel non-coding associations for height with evidence of impact on gene regulation. Our work provides a template for the analysis of non-coding rare variants for common human phenotypes, and ushers in the next era of genome association studies.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1876 Whole Genome Sequencing (WGS) in COVID-19 persistent chemosensory dysfunction: assessing the role of Loss of Function variants (LoFs) in a highly characterized Italian cohort.

Authors:

G. G. Nardone[1], B. Spedicati[1,2], A. Pecori[2], A. Santin[1], A. Morgan[2], M. P. Concas[2], P. Gasparini[1,2], P. Boscolo-Rizzo[1], G. Girotto[1,2]; [1]Dept. of Med., Surgery and Hlth.Sci., Univ. of Trieste, Trieste, Italy, [2]Inst. for Maternal and Child Hlth.– I.R.C.C.S. "Burlo Garofolo", Trieste, Italy

Background: Chemosensory dysfunction can be caused by several pathogens, including common respiratory viruses (i.e. rhinovirus, adenovirus), cytomegalovirus, Zika virus, HIV, and, more recently, SARS-CoV-2. Despite all the knowledge gathered in the last years, to date, the biological and genetic mechanisms underlying COVID-19-derived long-lasting chemosensory alteration are still widely unexplored. In this light, WGS analysis of a large cohort of fully characterised Italian individuals previously diagnosed with COVID-19 was carried out to identify candidate genes involved in persistent chemosensory alterations.
Methods: Olfactory and taste function of 201 individuals (164 cases; 37 controls) was assessed with the extended Sniffin Sticks test and validated Taste Strips. Chemosensory alterations were assessed via self-reporting questionnaires and Visual Analog Scale. All individuals underwent WGS and data was filtered for LoFs with two gene lists obtained, respectively, by literature review and using Ingenuity Pathway Analysis Compare tool: 1) 209 genes known to be involved in olfactory perception/smell loss; 2) 26 genes involved both in recessive immunodeficiencies and COVID-19. Linear and logistic regression models for three groups (males, females, sex-combined) and burden tests were performed to investigate the association between LoFs and chemosensory phenotypes.
Results: Our preliminary analyses highlighted the association of several genes involved in immune response with odor discrimination: *CD3G* (p-value: 0.048) in the sex-combined group, *IFNGR2* (p-value: 0.046) and *PPARG* (p-value: 0.046) in the female group and *CFI* in both the sex-combined (p-value: 0.048) and the female (p-value: 0.046) groups. Moreover, complement factor *C3* (p-value: 0.046) and olfactory receptor *OR4E1* (p-value: 0.044) genes were significantly associated, respectively, with olfactory and taste alteration in the male group. Finally, burden test highlighted the association of List 1) genes with taste alteration (p-value: 0.046) and nasal patency (p-value: 0.031) in males, while genes List 2) genes were significantly associated with parosmia in the sex combined (p-value: 0.049) and in the female group (p-value: 0.025).
Conclusions: Here, for the first time, a large cohort of patients, fully characterised through a comprehensive psychophysical evaluation of smell and taste, have been analysed using WGS. The results of this study highlight the unexplored role of LoFs in the determination of COVID-19-derived long-lasting chemosensory alteration, pointing out possible candidates to be targeted in future personalized treatments.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1877 Whole Genome Sequencing (WGS) Meta-analysis identifies Loci Associated with Non-alcoholic Fatty Liver Disease (NAFLD) across ancestries

Authors:

C. Raut[1], Y. Chen[1], A. Oliveri[1], M. F. Feitosa[2], J. R. O'Connell[3], K. A. Ryan[3], J. I. Rotter[4], S. S. Rich[5], K. A. Young[6], A. Hakim[7], P. A. Peyser[8], L. F. Bielak[1], M. T. Long[9], C-T. Liu[10], E. K. Speliotes[11], N. D. Palmer[12], NHLBI Trans-Omics for Precision Medicine (TOPMed)Consortium, Genetics of Obesity and Liver Disease (GOLD) Consortium; [1]Univ. of Michigan, Ann Arbor, MI, [2]Washington Univ Sch Med., Saint Louis, MO, [3]Univ. of Maryland-Baltimore, Baltimore, MD, [4]Lundquist Inst., Harbor-UCLA Med Ctr, Torrance, CA, [5]Univ. of Virginia, Scottsville, VA, [6]Univ. of Colorado Anschutz, Aurora, CO, [7]Brigham and Women's Hosp. and Harvard Med. Sch., Boston, MA, [8]Univ Michigan, Ann Arbor, MI, [9]Boston Univ. Sch. of Med. and Novo Nordisk a/S, Boston, MA, [10]Boston Univ. SPH, Boston, MA, [11]Univ. of Michigan Med. Sch., Ann Arbor, MI, [12]Wake Forest Sch. of Med., Winston-Salem, NC

Background: Non-alcoholic fatty liver disease (NAFLD) affects about 30% of individuals in the United States. There are few effective treatments for the disease due to a lack of understanding on its causes. NAFLD is heritable and previous genome-wide association analyses mostly in individuals of European ancestry have identified multiple common variants that associate with the disease. These common variants however explain about 20% of the heritability of the disease suggesting that more variants that affect the disease exist and remain to be discovered. NAFLD prevalence is also higher in individuals of Hispanic/Asian ancestry than in European or African ancestries. This difference in prevalence is mostly due to genetic differences in these groups. To better understand the genetic underpinnings of NAFLD, we conducted the largest-to-date multi-ancestry whole genome sequencing (WGS) association study of NAFLD.
Methods: We carried out study-, ethnic/race- and sex-stratified association analyses in six cohorts with imaging-measured hepatic steatosis using SAIGEgds adjusted for age, sex (where appropriate), alcoholic drinks per week, and principal component estimates of admixture. Study results were meta-analyzed overall and through male- and female-only using a fixed-effects meta-analysis in METAL. Cochran's Q test and the I^2 metric were used to identify and quantify heterogeneity.
Results: The overall meta-analysis included 16,664 individuals including individuals of European, African American, Hispanic, and Chinese American ancestries with imaging-measured hepatic steatosis. The overall meta-analysis identified three variants significantly associated ($P \leq 5E-08$) with NAFLD, i.e. *PNPLA3, HAPLN4, PPP1R3B.* An additional 12 variants trended toward association ($P \leq 5E-07$). The sex-stratified meta-analysis revealed that the PNPLA3 and PPP1R3B variants were more strongly associated in females. In addition, variants in HAPLN4 and TMTM120B were identified in male- and female-only analyses, respectively.
Conclusions: In a large, multiethnic analysis of imaging-measured hepatic steatosis, we replicated loci previously associated with NAFLD and identified new sex-specific loci. Several variants were trending toward association and will benefit from ongoing analyses to include 6,492 additional samples with imaging-measured steatosis.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1878 Whole Genome Sequencing Analysis of Weight Loss in Chronic Obstructive Pulmonary Disease in Trans-Omics for Precision Medicine

Authors:

A. Rocco[1], J. Chiles[1], W. Labaki[2], E. Wan[3], M. Han[2], R. Bowler[4], S. Christenson[5], L. Lange[6], Y. Gao[7], B. Psaty[8], S. Gharib[9], X. Guo[10], S. Rich[11], G. O'Connor[12], M. Wells[1], R. Patel[1], R. Casaburi[13], H. Rossiter[10], H. Tiwari[14], M-L. Mcdonald[15]; [1]Univ. of Alabama at Birmingham, Birmingham, AL, [2]Univ. of Michigan, Ann Arbor, MI, [3]Brigham & Women's Hosp, Boston, MA, [4]Natl. Jewish Hlth., Denver, CO, [5]Univ. of California, San Francisco, CA, [6]Univ Colorado Denver, Aurora, CO, [7]Univ. of Mississippi Med. Ctr., Jackson, MS, [8]Univ. of Washington, Seattle, VA, [9]Univ. of Washington, Seattle, WA, [10]The Lundquist Inst., Torrance, CA, [11]Univ. of Virginia, Scottsville, VA, [12]Boston Univ., Boston, MA, [13]Univ. of California, Los Angeles, CA, [14]Univ Alabama at Birmingham, Birmingham, IN, [15]Univ. of Alabama of Birmingham, Birmingham, AL

RATIONALE: Weight loss (WL) in chronic obstructive pulmonary disease (COPD) is associated with increased mortality. Identifying genetic variants associated with WL in COPD may help identify causal pathways as possible therapeutic intervention targets. Our prior genome-wide association study of WL in COPD utilized affinity chip data, but whole genome sequencing (WGS) data is now available for analysis, yielding opportunities for increased cohort size and genomic resolution.
METHODS: We utilized data from seven studies within the Trans-Omics for Precision Medicine initiative: the Cardiovascular Health Study (CHS); Genetic Epidemiology of COPD (COPDGene); Evaluation of COPD Longitudinally to Identify Predictive Surrogate End-points (ECLIPSE); Framingham Heart Study (FHS); Jackson Heart Study (JHS); Multi-Ethnicity Study of Atherosclerosis (MESA); and the SubPopulations and InteRmediate Outcome Measures In COPD Study (SPIROMICS). All analyses were conducted in 8,628 participants with COPD, where COPD was characterized using spirometry and smoking history. Of these participants, 29.6% had WL of 5% or more between study visits or a body mass index (BMI) of less than 20 kg/m2. WGS analyses, adjusted for age, sex, and principal components of ancestry, were performed using the GENESIS BioData Catalyst pipeline, with participants of African-American (AA) and non-Hispanic white (NHW) ancestries analyzed separately. Fixed-effects meta-analysis was performed using METAL.
RESULTS: Across nine ancestry- and study-stratified cohorts of research participants with COPD, we identified eight variants that reached significant association with WL in COPD across five distinct loci. In COPDGene NHW participants with COPD, two chromosome 12 variants intergenic to LINC02426 and CCDC59 were associated with increased WL risk (OR = 5.74, p = 2.0 x 10-8 for both). In ECLIPSE, two intronic variants in SEPTIN7 on chromosome 7 were associated with increased WL risk (OR = 5.35, p = 3.3x10-8 for both). In FHS, one 3' untranslated region chromosome 1 variant in NR5A2 was significantly associated with WL (OR = 0.136, p = 3.0x10-8). In JHS, one chromosome 2 variant between DOCK10 and NYAP2 was associated with increased WL risk (OR = 15.5, p = 7.7x10-9). Finally, in CHS, two chromosome 3 variants intronic to HHATL were associated with increased WL risk (OR 4.4, p = 2.7x10-8 and OR = 5.1 and p = 3.2x10-8). None of the identified associations remained significant after meta-analysis.
DISCUSSION: We identified five novel loci associated with unintentional weight loss in COPD. Additional efforts are being directed at fine mapping and replication analyses in larger biobanks with genomic data.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1879 † Whole genome sequencing analysis reveals structural variants in the H1/H2 haplotype region and new susceptibility loci associated with progressive supranuclear palsy.

Authors:

H. Wang[1], T. Chang[2], B. Dombroski[1], U. Kumar[2], Y. Leung[1], C. Dalgard[3], D. Dickson[4], L-S. Wang[1], G. Schellenberg[1], D. Geschwind[2], W-P. Lee[1]; [1]Univ. of Pennsylvania, Philadelphia, PA, [2]Univ. of California, Los Angeles, Los Angeles, CA, [3]Uniformed Services Univ. of the Hlth.Sci., Bethesda, MD, [4]Mayo Clinic, Jacksonville, FL

Background: Progressive supranuclear palsy (PSP) is a neurodegenerative disease characterized by the buildup of tau protein in neural cells. Studies have shown that the H1/H2 haplotype region, including microtubule associated protein tau (MAPT), is associated with an increased genetic risk of PSP. But it is challenging to identify new variants and genes or even causal variants and genes in H1/H2 for PSP. **Method:** We analyzed whole-genome sequencing data from 4878 subjects (N_{case} = 1767, $N_{control}$ = 3111). Single nucleotide variants (SNVs) and small insertions and deletions (INDELs) were identified by GATK while structural variants (SVs) were first called using Manta and Smoove and then joint-genotyped using Graphtyper2. Then, an association analysis was conducted on SNVs, INDELs, and SVs to identify variants associated with PSP. Furthermore, we analyzed complex structure of H1/H2 and examined various variants within the haplotype to gain a deeper understanding of the association of H1/H2 and PSP. **Results:** Our analysis of SNVs and INDELs confirmed the association of *MAPT*, *MOBP*, and *STX6* with PSP and revealed two novel genes, *APOE* and *FCHO1*/*MAP1S* ($P < 5 \times 10^{-8}$). Notably, *APOE E2* and *E4* alleles exhibited opposite effects in PSP compared to Alzheimer's disease. In our analysis, *E2* emerged as a risk allele, while *E4* displayed a protective effect against PSP. Additionally, six loci are suggestive of significance ($P < 1 \times 10^{-6}$) in *SLCO1A2*, *DUSP10*, *SP1*, *KIF13A*, *TRIM24* and *TNXB*. Regarding SV, six high-quality SVs showed a significant association with $P < 5 \times 10^{-8}$, including three deletions tagging the H2 haplotype, two insertions in *UCMA* and *RBMS2*, and one deletion in *PCMT1*. Notably, individuals with a homozygous deletion (chr6:149762615-149763234, $P = 6.62 \times 10^{-11}$) in *PCMT1* show an odds ratio of 8.68. Within the H1/H2 haplotype region, we identified 28 CNVs ranging from 80 bp to 47 kb. Six out of seven common CNVs are covered by transposable elements (SVAs, L1s, or Alus), indicating the crucial role of transposable elements in shaping the landscape of this region. The remaining 21 rare CNVs displayed a significantly higher burden in individuals with PSP ($P = 7.94 \times 10^{-4}$). One singleton deletion (chr17:45993882-45993970), which covers exon 9 of *MAPT*, is identified in an individual with PSP. Overall, we helped resolve the genetic basis of PSP by identifying novel variants associated with the disease and potential causal variants within the H1/H2 haplotype region. These findings provide valuable insights for future research into disease mechanism and therapeutic targets.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1880 Whole genome sequencing and proteomic analysis of mucus plugging in COPDGene

Authors:

K. Kim[1], H. Nath[2], S. Grumley[3], J. Orejas[4], W. Dolliver[4], S. Mettler[1], A. Yen[3], K. Jacobs[3], P. P. Manapragada[2], M. Abozeed[2], M. Aziz[2], M. Zahid[2], S. Sonavane[5], Trans-Omics in Precision Medicine (TOPMed), E. Silverman[1,6], R. Bowler[7], A. Diaz[4], M. Cho[1,6]; [1]Channing Div. of Network Med., Brigham and Women's Hosp., Harvard Med. Sch., Boston, MA, [2]Dept. of Radiology, Univ. of Alabama at Birmingham, Birmingham, AL, [3]Dept. of Radiology, Univ. of California San Diego, San Diego, CA, [4]Div. of Pulmonary and Critical Care Med., Brigham and Women's Hosp., Boston, MA, [5]Dept. of Radiology, Mayo Clinic, Jacksonville, FL, [6]Div. of Pulmonary and Critical Care Med., Brigham and Women's Hosp., Harvard Med. Sch., Boston, MA, [7]Natl. Jewish Hlth., Denver, CO

Introduction:Mucus plugs in smokers with and without COPD are associated with worse lung function and may represent an important treatable trait in airways disease. While genetic risk factors have been identified for symptomatic syndromes associated with mucus plugging (e.g. bronchiectasis), genetic risk factors for CT-identified mucus plugs, which are often asymptomatic, have not been identified. Methods:We performed a whole genome sequencing association study of mucus plugs in 3,433 non-Hispanic white (NHW) and 1,021 African American (AA) individuals in the Genetic Epidemiology of COPD (COPDGene) using data generated from the Trans-Omics in Precision Medicine (TOPMed) study. We calculated lung mucus plugging scores using chest CT scans. We analyzed a dichotomized overall score after adjusting for age, gender, smoking pack-years, current smoking status, sequencing center, and genetic ancestry. We also performed a secondary analysis using the value in each lobe. We investigated genome-wide single variants, loss-of-function rare variants within genes and candidate associations in genes with reported associations for related phenotypes (cough, phlegm, bronchiectasis). We considered a P-value of 5x10-8 for genome-wide significance. We also used SomaScan 1.3k data in a subset of 518 individuals to investigate the association between the trait and proteins. Results:We found no significant associations in the single variant, rare variant, and protein analyses using the binary score for whole mucus plugging. Our top findings in single variant associations were near FAM80B and SLC2A13 (P = < 2 x10-7). In rare variant analysis, the top association for high-confidence loss-of-function rare variants was in HAL (P = 3.84x10-6), while our top protein association was CXCL13 (P = 2.29x10-4). Notably, we did not identify any associations with previously described associations with cough, phlegm, or in other candidate genes associated with bronchiectasis. In a secondary analysis of lobar mucus plugging binary scores, we identified 4 genome-wide significant regions; our top finding was near ZNF407 (8.02x10-9) with other less significant associations near ZNF816, CCSER1, and L3HYPDH. Conclusions: In a whole genome sequencing and proteomic study of mucus plugging, we did not identify genome- or proteome-wide significant associations; nor did we find associations in a set of candidate genes and variants. In a secondary analysis, we did identify associations with lobe-specific values, though their clinical significance remains unclear. Future work will include analysis of quantitative phenotypes and in additional samples.

Session Title: Complex Traits and Polygenic Disorders Poster Session III

PB1881 Whole-exome-widerarevariants analysis detectedgenesassociatedwithpediatric sepsisphenotypewith severe outcome

Authors:

Y. Qin[1], K. Kernan[2], J. Carcillo[2], H. Park[1]; [1]Univ. of Pittsburgh, Pittsburgh, PA, [2]UPMC children Hosp., Pittsburgh, PA

Pediatric sepsis is leading to a global health burden in children. The unavoidable heterogeneity within the patients has hindered the performance of regular treatment. As one of the potential sources contributing to disease heterogeneity, the role of rare variants has not been fully understood. Our previous study identified four phenotypes showing distinct clinical characteristics, cytokine patterns, outcomes, and responses to therapies (PedSep-A, B, C, and D). In this study, we collected whole-exome-wide sequencing data from pediatric sepsis patients from the same cohort and conducted a gene-based analysis to aggregately testassociations between rare variants and pediatric sepsis phenotypes. One whole-exome-wide significant gene (LTBP4) and two suggestive significant genes (PLA2G4E, CCDC157) showed association with PedSep-D, one of the pediatric sepsis phenotypes with the most severe outcomes. Since the genes are implicated for inflammation, immune cell activation, these results implied the influence of rare variants in pediatric sepsis subgroups, enhanced genetics basis of critically ill patients with the disease, and assisted informing more tailored application in precision medicine.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1882 Whole-genome Sequencing Analysis of Body Mass Index in the Trans-Omics for Precision Medicine (TOPMed) Program Identifies Novel African Ancestry-specific Risk Allele

Authors:

X. Zhang[1], J. Brody[2], M. Graff[1], H. Highland[1], N. Chami[3], L. Lange[4], C-T. Liu[5], R. Loos[6,3], K. North[1], A. Justice[7]; [1]Univ North Carolina at Chapel Hill, Chapel Hill, NC, [2]Univ of Washington, Seattle, WA, [3]Icahn Sch. of Med. at Mount Sinai, New York, NY, [4]Univ Colorado Anschutz Med. Campus, Aurora, CO, [5]Boston Univ. SPH, Boston, MA, [6]Univ. of Copenhagen, Copenhagen, Denmark, [7]Geisinger, Danville, PA

Background: Obesity is a global public health crisis associated with high morbidity and mortality rates. Previous genome-wide association studies (GWAS) investigating body mass index (BMI) have primarily relied on imputed data from European individuals. Consequently, most BMI risk variants identified to date are common and in primarily European ancestry populations exhibiting small effect sizes, leaving low and rare variants with potentially large effects largely unexplored.
Methods: This study leveraged whole-genome sequencing (WGS) data from 88,873 participants from the Trans-Omics for Precision Medicine (TOPMed) Program, of which 51% belonged to non-European population groups. We performed GWAS of BMI adjusting for age, age^2, sex, study, population group, principal components, sequencing center, sequencing phase, and project (TOPMed vs. Centers for Common Disease Genomics). We subsequently conducted replication analyses, stepwise conditional analyses, rare variant (MAF $\leq$ 1%) aggregate association analyses, and fine-mapping.
Results: We discovered a total of 18 BMI-associated signals ($P < 5 \times 10^{-9}$), including a novel low-frequency single nucleotide polymorphism (SNP), rs111490516, in *MTMR3* that was common in individuals of African descent (minor allele frequency [MAF] = 13% in African and Barbadian population groups). We successfully replicated this novel SNP and observed directionally consistent associations across replication studies of similar background. In the meta-analysis of 198,621 individuals from both discovery and replication studies, the estimated effect of rs111490516 was 0.037 kg/m^2 with a SE of 0.006 ($P = 4.19 \times 10^{-9}$). Using our diverse study population, we additionally identified two potentially novel secondary signals in known BMI loci (rs2206277 in *TFAP2B* and rs3838785 in *BDNF*) in our conditional analyses and pinpointed two likely causal variants in the *POC5* (rs2307111, posterior probability [PP] = 0.99) and *DMD* loci (rs1379871, PP = 1.00) in our fine-mapping analyses. Finally, we successfully replicated previous gene-based associations with the well-known *MC4R* gene ($P = 8.47 \times 10^{-8}$) by aggregating 111 alleles across 37 sites within coding regions, enhancers, and promoters.
Conclusion: Our work shows the benefits of linking WGS in diverse cohorts for discovering new variants and genes that confer risk for obesity. Ultimately, our study brings us one step closer to understanding the complex genetic underpinnings of obesity, translating these leads into mechanistic insights, and developing targeted preventions and interventions to address this global public health challenge.

Session Title: Complex Traits and Polygenic Disorders Poster Session II

PB1883 Whole-genome sequencing analysis of IgG4-related disease in the Japanese population

Authors:

Y. Zhang, T. Iwasaki, T. Kawaguchi, I. Yamaguchi, M. Takahashi, S. Kawaguchi, F. Matsuda, for the Japanese Research Group of IgG4-Related Disease; Kyoto Univ., Kyoto-city, Japan

Introduction: IgG4-related disease (IgG4-RD) is a rare autoimmune disorder characterized by organ enlargement, tissue infiltration by IgG4+ plasma cells, storiform fibrosis, and elevated serum IgG4 levels. Previously, we identified the *HLA-DRB1* and *FCGR2B* regions as susceptibility loci for IgG4-RD via a genome-wide association study (GWAS) with a single-nucleotide polymorphism (SNP) array. However, the potential involvement of the rare variants and copy number variations (CNVs) in IgG4-RD has not been investigated. **Method:** We performed a GWAS in the Japanese population using whole genome sequencing data (cases, n=732; controls, n=3,148). Associations with the risk of IgG4-RD subtypes, Mikulicz's disease (MD) and autoimmune pancreatitis (AIP), were analyzed. For complement component 4 (C4) CNV analysis, a depth-oriented tool from GATK was utilized to estimate the total copy number (CN) of C4. The estimated CN was further calibrated using known C4 CN data. Next, we determined the ratio of isotypes, C4A and C4B, based on the relative read depths of five C4A/C4B-specific SNPs. We further investigated the association of C4 CNV with IgG4-RD and evaluated its independence from the human leukocyte antigen (HLA) risk alleles. **Result:** We identified a novel locus close to the *PTCH1* region in MD patients. The variant with the lowest *p*-value (the lead variant) was rs66917126, where the presence of the alternative T allele was associated with an increased risk of IgG4-RD (odds ratio (OR), 2.12; $p = 5.18 \times 10^{-9}$). Additionally, the external expression quantitative trait loci (eQTL) data suggested that the T allele of rs35474521, which is in linkage disequilibrium (LD) with the second lead SNP ($r^2 = 0.79$), increased *PTCH1* expression (beta = 0.28, $p = 9.32 \times 10^{-39}$). In patients with AIP, we identified an association of a rare variant (rs544522470) located in the intron of the *DPY19L3* gene. Its alternative A allele increased the risk of IgG4-RD (OR, 7.31; $p = 4.38 \times 10^{-8}$). As for the CNV of C4, we found that individuals with a lower C4A CN and a higher C4B CN had a higher risk of IgG4-RD (p = 0.0047, 0.0145, respectively). We also confirmed their independence from two previously identified HLA risk alleles (*DRB1*04:06* and *DRB1*09:01*) using logistic regression models. **Conclusion:** Based on the whole genome sequencing data, we identified two susceptibility loci and one risk copy number variation for IgG4-related disease. Our findings nominate candidate genes for further exploration and provide novel insights into the etiology of IgG4-RD.

Session Title: Complex Traits and Polygenic Disorders Poster Session I

PB1885 “Age-in-a-dish” model of astrocytes - Can directed differentiation do the trick?

Authors:

M. Carless, U. Bhaskar; Univ. of Texas at San Antonio, San Antonio, TX

Age is undeniably a major risk factor for neurodegenerative diseases, including Alzheimer’s disease (AD), Parkinson’s disease (PD), and other neurological disorders such as epilepsy and stroke. Current estimates predict that at least one in six people will be aged 60 or over by 2030. Astrocytes, the largest group of cells in the brain, and regulators of CNS homeostasis, exhibit dysfunctional synapse regulation, ion dyshomeostasis, neuronal atrophy, synapse loss, and upregulation in neuroinflammatory and oxidative stress response genes with age. Such reactive phenotypes also occur in response to pathological stimuli, and as such, astrocytes are prime candidates in contributing to disease progression. However, few, if any, studies have investigated the relevance of astrocytes to age-related disorders. This is, in part, due to a lack of efficient human model systems that can capture age-associated disease progression. Prior research in neurons has demonstrated that age-associated methylome and transcriptome patterns can be retained following direct conversion from fibroblasts, which bypasses the stem-cell state. Similar strategies in astrocytes have shown inefficient conversion when transforming ***adult*** human fibroblasts. Here, we have developed an efficient direct conversion strategy employing a combination of small molecules to induce trans-differentiation of human adult fibroblasts to astrocytes. We demonstrate that this method produces mature GFAP+/S100B+ cells at high efficiency (40-45%) that are enriched for markers of astrocyte functionality, including ion-channel buffering, gap-junction communication and glutamate uptake; and that these cells exhibit astrocyte-like calcium signaling and neuroinflammatory phenotypes. RNA-Seq analysis indicates an adult rather than fetal astrocytic gene expression signature, with higher relative expression of forebrain-specific markers as compared to hindbrain, ventral, or spinal cord markers. We see evidence of an aged phenotype in our fibroblast-derived astrocytes compared to induced pluripotent stem cell-derived astrocytes via increased reactive oxygen species ($p<0.001$) and increased senescence ($p<0.001$). Future studies will focus on further validation of an aged phenotype through assessing DNA methylation and evaluating "epigenetic age". Our method of generating fibroblast-derived induced astrocytes, provides a useful tool in understanding age-associated disease processes and will be an asset in the study of late-stage brain disorders.

Session Title: Genetic Therapies Poster Session I

PB2001 A case study of a female infant with primary hypertrophic osteoarthropathy demonstrates that early initiation of celecoxib slows but does not prevent symptom progression.

Authors:

K. Zehr[1], M. Buckley[1], J. Owens[2], A. Shillington[3]; [1]Univ. of Cincinnati Coll. of Med., Cincinnati, OH, [2]Cincinnati Children's Hosp., Cincinnati, OH, [3]Cincinnati Children's Hosp., Cincinnatti, OH

Primary Hypertrophic Osteoarthropathy (PHO) is characterized by digital clubbing, periostosis, and pachydermia. HPGD encodes 15-prostaglandin dehydrogenase, which catalyzes the first step of prostaglandin degradation. Loss of function *HPGD* variants cause Primary Hypertrophic Osteoarthropathy Autosomal Recessive 1 (PHOAR1) due to insufficient prostaglandin degradation. Recognizing that PHOAR1 pathology is driven primarily by increased Prostaglandin E2 (PGE2) levels has led to multiple clinical trials exploring the impact of selective COX-2 inhibitors in adults with PHO. Many studies have demonstrated that selective COX-2 inhibitors improve digital clubbing, joint swelling, pachydermia, joint pain/mobility, and inflammatory markers in adults with PHO. In this case study we present a female patient incidentally diagnosed with PHOAR1 due to a likely pathogenic homozygous *HPGD* c.218-1G>A variant at 5 months of age. At presentation, she was noted to have a diffuse erythematous rash secondary to hyperhidrosis, symmetric clubbing of her fingertips, and relatively mild decreased range of motion of bilateral knees and wrists for her age group. By 20 months of age, the bilateral clubbing was more notorious, she had flexion contracture of the knees (~15 degrees), and general signs and symptoms of significant joint pain that limited her ability to play. Six months later, the patient was started on celecoxib, a selective COX-2 inhibitor, and showed improvement but not resolution of her symptoms. This patient's early PHOAR1 diagnosis allowed for a better understanding of the clinical course of PHOAR1 with early initiation of celecoxib, in contrast to previously treated patients who started treatment at an older age with established chronic symptoms. Her case could demonstrate that starting celecoxib early in the course of disease delays but does not stop disease progression, suggesting that fully resolving PHO-related symptoms may go beyond normalizing PGE2 levels through COX-2 inhibition alone.

Session Title: Genetic Therapies Poster Session II

PB2002 A drug repurposing screen reveals dopamine signaling is important for the rare disease DPAGT1-CDG.

Authors:

H. Dalton; Univ. of Utah, Salt Lake City, UT

Glycosylation encompasses a wide class of biological pathways involving co- and post-translational sugar modifications. Mutations in glycosylation genes underlie Congenital Disorders of Glycosylation (CDGs) - ultra-rare disorders that can cause seizures, developmental delay, and early death. There are few treatment options available for CDGs, and small patient populations make clinical trials difficult. Thus, there is a great need for alternative approaches to finding treatments for these rare diseases. One such alternative is the use of drug repurposing screens that utilize libraries of small molecules with established safety profiles in humans, allowing for potentially faster turnaround for patients.

DPAGT1-CDG is caused by mutations in the gene *DPAGT1*, which encodes the essential first enzyme for N-linked glycosylation. We created a model of DPAGT1-CDG in *Drosophila* via *DPAGT1* knockdown in the fly eye to cause a small, rough eye phenotype. This model can be screened for new therapies by quantitatively measuring improvements to its eye size. To find such therapies, we performed a drug repurposing screen using 1,520 small molecules that are 98% FDA/EMA-approved (Prestwick Chemical Library). Drugs were mixed into fly food, flies were exposed until early adulthood, and eye size was compared to control, DMSO-treated flies. The top candidate drugs that rescue, resulting in a larger eye, were validated by dose-response or genetic analyses.

The screen has a 3% hit rate for drugs capable of rescuing the DPAGT1-CDG model eye size ($Z \geq 1.5$). Top drug categories are dopamine receptor D2 (D2R) antagonists, COX inhibitors, and ion transporter-related drugs. Three D2R antagonists were strong hits (Z>1.8, 20%+ eye size). Confirming the role of this pathway, knockdown of the D2R-encoding gene *DRD2*/*Dop2R* could recapitulate this effect (20%+ eye size). This represents a potential new avenue for therapy - especially as seizures and movement disorders are common in DPAGT1-CDG patients. In addition, we genetically validated COX inhibitors through knockdown of the COX-like gene *Pxt* (17%+ eye size), as well as an ion transporter inhibitor through its target encoded by *NKCC1*/*Ncc69* (+12% eye size). All three pathways reveal novel biology related to *DPAGT1* mechanisms, and they may represent new therapeutic, FDA-approved options for DPAGT1-CDG.

Session Title: Genetic Therapies Poster Session III

PB2003 A values typology of underpinning responses to gene editing compared to other means of enhancement

Authors:

c. castelyn[1], W. Van Staden[1], M. Pepper[2]; [1]Univ. of Pretoria, Pretoria, South Africa, [2]Univ Pretoria, Lynnwood, South Africa

Gene editing tools such as CRISPR may be limited to genetic abnormalities that cause disease, however it is inevitable that these technologies will be extended to non-medical applications. This study aimed to uncover what would be perceived as being good, bad, wrong or right about gene editing not only for therapeutic purposes but also for enhancement purposes. The study moves beyond the knee jerk reactions of uneasiness to identify the underpinning value judgements and consequent kinds of values.Data was collected through individual in-depth interviews with eight purposively selected participants. Values-based practice guided the study in recognising the underpinning values. Data was analysed by searching for the value words in the responses, typological coding allowed for the data to be coded according to pre-identified types of values such as good and bad, wrong or right, deontological, societal, cultural and aspirational values.The findings that emerged differentiated between individual and societal perceptions. There was overlap between perceptions about therapeutic and enhancement purposes of gene editing. Gene editing for enhancement could be perceived as being good for personal fulfilment and for humanity's improvement. It would be bad if gene editing were to repeat the past and if some were enhanced at the expense of others. Gene editing would be right if human dignity was upheld. It would be wrong if universal access was unattainable and if it led to people not having the freedom to choose whether to gene edit themselves or their offspring. To account for a wide variety of kinds of values underpinning perceptions of gene editing not only for therapeutic but also enhancement purposes is important so as to inform regulations, frameworks and commissions.

Session Title: Genetic Therapies Poster Session I

PB2004 Advancing Recombinant Adeno-Associated Virus Vector Integration Analysis through Next-Generation Sequencing Optimization

Authors:

B. Niu, H. Latif, Y. Fan, Y. Zhang, C. Goswami, D. Corney, A. O'Hara, L. Turner, G. Zhou; Azenta US, South Plainfield, NJ

Recombinant adeno-associated viral (rAAV) vectors have seen widespread application as a gene therapy delivery platform in preclinical and clinical studies. Though rAAV largely remains episomal after transduced into the nucleus of host cells, data has shown that rAAV does integrate into the host genome in an undefined, non-homologous manner. The frequency of rAAV integration has been estimated at 0.1 to 10% in hepatocytes. In consideration of the theoretical risk of tumorigenesis associated with rAAV integration in humans, assessment of rAAV integration in *in vitro* studies, animal models and patient samples are recommended by FDA for preclinical and clinical studies and long term follow up studies (LTFU). rAAV integration analysis presents several challenges, including the random nature of rAAV integration sites and low expected integration frequency. To address this, Next-Generation Sequencing (NGS) for rAAV integration analysis has been adopted as a highly sensitive assay. Here we conducted a comparative study analyzing multiple NGS-based methodologies to study rAAV integration, including hybrid capture-based target enrichment sequencing, as well as amplification-based sequencing such as Quantitative Shearing Linear Amplification Mediated-PCR (qsLAM-PCR) and Shearing Extension Primer Tag Selection Ligation-Mediated PCR (S-EPTS/LM-PCR). In addition, we assessed multiple approaches for episomal rAAV genome removal to improve the sensitivity of NGS-based methods. The studies described will drive greater insights into rAAV integration and its impact on patient safety, as well as contribute to improvement and standardization of analysis methods.

Session Title: Genetic Therapies Poster Session II

PB2005 AI for Next-Generation Gene Design

Authors:

D. Gaddes[1], D. Cirillo[2], A. Edwards[1], A. Hott[1]; [1]CFD Res. Corp., Huntsville, AL, [2]The Ohio State Univ., Columbus, OH

To support enhanced healthcare in a post-pandemic landscape, it is scientifically and economically imperative to maximize the expression of protein therapeutics or mRNA vaccines. In pursuit of this objective, we present a codon harmonization technique, powered by AI, resulting in a more than 400% increase in protein expression. In modern healthcare, genomic tools like synthetic plasmids or mRNA are employed to synthesize proteins for therapeutics and vaccines. The design of genes for plasmids or mRNA vaccines involves precisely configuring the nucleic acid sequence to achieve desired outcomes. Traditionally, gene and mRNA design has relied on codon optimization, a computational technique aimed at maximizing protein expression and improving translational efficiency. Codon optimization achieves this by integrating the most frequently utilized codons of the host organism into the synthesized gene or mRNA. However, concerns have emerged regarding the efficiency of codon optimization, leading to the proposal of codon harmonization as an alternative approach. The primary goal of codon harmonization is to reduce protein misfolding while simultaneously improving protein yields. This approach achieves its objective by incorporating host organism-specific codons in the synthetic gene in a proportional manner, resulting in a gene design that closely aligns with the codon utilization of the host organism. Towards creating synthetic genes more like the host organism's, we have developed a generative AI tool to design genes based on learning how the host organism typically encodes DNA for protein production. By employing a transformer model, we successfully uncovered the correspondence between genes from *E. Coli* and highly expressed proteins. Our dataset consisted of over 20,000 gene-protein pairs, enabling the model to learn this relationship. Notably, our AI achieved a 73% accuracy in codon prediction and demonstrated 100% accuracy in amino acid prediction, without any mutations or changes in protein length. A comparative analysis with a commercial codon optimization tool revealed remarkable outcomes for our generative AI. Specifically, our tool yielded a striking 428% increase in COVID nucleocapsid protein expression when using instant induction media. These initial findings provide a glimpse into the potential of AI-driven manufacturing, which holds significant promise in reducing development time and treatment costs. Moving forward, our focus will be on further enhancing protein expression and folding across different host organisms, including CHO cells, while also exploring the application of generative AI in mRNA therapeutics.

Session Title: Genetic Therapies Poster Session III

PB2006 An *in vivo* screen identifies small molecule modulators of the endoplasmic reticulum stress response.

Authors:

K. Beebe, H. Evans, K. Hope, A. Berman, C. Chow; Univ. of Utah, Salt Lake City, UT

Misfolded protein accumulation in the endoplasmic reticulum (ER) induces ER stress. Cells respond to ER stress by initiating the unfolded protein response (UPR) that upregulates chaperone protein expression, increases the degradation of misfolded proteins, and inhibits protein translation. Failure to effectively manage ER stress and restore homeostasis results in cellular dysfunction and ultimately apoptosis, a process implicated in numerous human diseases such as retinal degeneration, Alzheimer's disease (AD), Parkinson's disease (PD), Huntington's disease (HD), and amyotrophic lateral sclerosis (ALS), among others. Identifying small molecules that modulate ER stress may be effective therapeutics for human diseases caused by misfolded protein accumulation. Here, we used a *Drosophila* model of retinitis pigmentosa (RP) that expresses misfolded rhodopsin protein, $Rh1^{G69D}$, in the developing eye. $Rh1^{G69D}$ expression induces chronic ER stress and apoptosis, resulting in a degenerative eye phenotype. We took a drug repurposing approach and used the Prestwick Chemical Library, consisting of 1520 small molecules, the majority of which are FDA-approved, to identify compounds that modulate neuron cell death in $Rh1^{G69D}$ expressing flies. We identified multiple classes of drugs that enhance or suppress the degenerative eye phenotype, including compounds acting through monoamine neurotransmitters, folate metabolism, sodium channels, and the renin/angiotensin pathway. Degeneration-enhancing compounds may reveal novel ER stress pathways, while compounds that suppress degeneration are potential therapeutic candidates for RP. We are using an RNAi approach to identify the mechanism of action for the top enhancers and suppressors. Additionally, we will present data showing that compounds that rescue cell death in the RP model can also rescue disease-associated phenotypes in other *Drosophila* models of protein misfolding diseases, such as PD, HD, and ALS. This work identified potential therapeutic drugs for RP and possibly other human diseases that result from misfolded protein accumulation and ER stress.

Session Title: Genetic Therapies Poster Session I

PB2007 † Androgen and mineralocorticoid receptor signaling are drivers of - and therapeutic targets for - pubertal vascular rupture in Vascular Ehlers-Danlos Syndrome (VEDS) mice.

Authors:

E. Juzwiak[1,2], C. Bowen[1,2], A. Zeng[1,2], H. Dietz[1,2]; [1]Johns Hopkins Univ., Baltimore, MD, [2]Howard Hughes Med. Inst., Chevy Chase, MD

Background: In contrast to females, male patients with VEDS are at particularly high risk of spontaneous death due to arterial rupture during puberty. Our VEDS mouse model (*Col3a1*$^{G938D/+}$) recapitulates this sexually dimorphic vulnerability during puberty, with 49% vs. 66% survival of males vs. females by the end of puberty (60 days of age), respectively.
Methods: Given the male bias, we hypothesized that vascular rupture involved androgen-dependent perturbations of cellular function. We treated *Col3a1*$^{G938D/+}$ mice with the potent and selective AR antagonist (ARa) bicalutamide from weaning until 60 days of age. We also utilized a conditional (floxed) allele for the Ar gene in combination with a globally expressed Cre recombinase allele to create a genetic knock out for the AR in *Col3a1*$^{G938D/+}$ mice. *Col3a1*$^{G938D/+}$ mice were also treated with the dual ARa and mineralocorticoid receptor antagonist (MRa) spironolactone from weaning until 60 days of age. Many VEDS patients only come to medical attention post puberty, so we initiated spironolactone treatment of adult *Col3a1*$^{G938D/+}$ mice that had survived puberty and monitored subsequent survival. Finally, we treated *Col3a1*$^{G938D/+}$ mice with a selective MRa (finerenone or eplerenone) from weaning until 60 days of age.
Results: Deletion of the Ar gene affords improved survival at 60 days of age in male *Col3a1*$^{G938D/+}$ mice compared to their unmanipulated counterparts (80% vs. 49%), a protective performance similar to that observed upon treatment with bicalutamide (70% survival). Notably, male *Col3a1*$^{G938D/+}$ mice treated with spironolactone showed the best performance (88% survival at 60 days). Interestingly, use of either finerenone or eplerenone also afforded significant protection (74% male survival at 60 days with either treatment). The residual enhancement of risk for vascular death during puberty seen with isolated pharmacologic AR or MR antagonism in male VEDS mice is only abrogated upon dual AR/MR blockade, as achieved by spironolactone. Female *Col3a1*$^{G938D/+}$ mice are not afforded protection with isolated AR or MR antagonism; however, they do experience near-complete survival through puberty with spironolactone. Postpubertal treatment with spironolactone afforded complete and durable protection in both sexes.
Conclusions: These data document that pubertal vascular catastrophe in *Col3a1*$^{G938D/+}$ mice of both sexes is dependent on both AR and MR signaling with dramatic protection afforded by dual antagonism. There is therapeutic potential for isolated MR antagonism that will maintain normal sexual development in males transitioning through puberty.

Session Title: Genetic Therapies Poster Session II

PB2008 Cell-based functional assays for screening of drug candidates for Free Sialic Acid Storage Disorder.

Authors:

P. Leoyklang[1,2], M. Hackbarth[1,2], M. S. Sabir[2], K. Dobrenis[3], C. Lion[4], V. Rigolot[4], C. Biot[4], L. Wolfe[2], D. R. Adams[5], W. A. Gahl[1,2], M. C. V. Malicdan[1,2], M. Huizing[1]; [1]MGB, NHGRI, NIH, Bethesda, MD, [2]NIH Undiagnosed Diseases Program, NIH, Bethesda, MD, [3]Dept NeuroSci., Albert Einstein Coll. of Med., Bronx, NY, [4]Université de Lille, CNRS, UGSF, Lille, France, [5]OCD, NHGRI, NIH, Bethesda, MD

Background: Free sialic acid storage disorder (FSASD) is a progressive neurodegenerative disease caused by biallelic mutations in *SLC17A5,* which encodes sialin, the lysosomal sialic acid (SA) exporter. Affected cells exhibit lysosomal storage of free SA. FSASD manifests as a spectrum of clinical severity, and biochemical findings include increased urinary excretion and lysosomal accumulation of free SA. Despite extensive research on SA metabolism, membrane transport, and lysosomal biology, the underlying pathobiology remains poorly understood, and no approved therapies are available. **Methods:** We developed FSASD-specific cellular functional assays to identify potential therapeutic strategies. The primary assay involves live incorporation of alkyne-labeled SA into cells. Alkyne-SA is endocytosed to lysosomes, exported through sialin into the cytosol, and incorporated into glycoconjugates in the Golgi; these are detected using fluorescent click-chemistry and imaging. Secondary assays utilize immunofluorescence to examine distribution of lysotracker, LAMP1, LAMP2, and GM2. Tested FSASD cell lines include mouse embryonic fibroblasts, either deficient in sialin or carrying the prevalent p.Arg39Cys variant, as well as human FSASD primary skin fibroblasts with different *SLC17A5* variants. To further validate the assays, disease cells transfected with *GFP-SLC17A5* were also used. Additional testing was performed with candidate small-molecule drugs known to be applicable for other lysosomal storage diseases, including arimoclomol, genestein, and mercaptoethyguanidine. **Results:** Control cells efficiently incorporated alkyne-SA into glycoconjugates, while FSASD cells showed absent or markedly reduced incorporation due to defective sialin-mediated release of SA from lysosomes. Secondary assays showed increased numbers of small endo-lysosomes, elevated LAMP1, LAMP2, and lysotracker signals, as well as lysosomal GM2 accumulation in FSASD cells. Gene replacement with *GFP-SCL17A5* restored the wild type phenotype. Preliminary testing of candidate small molecule drugs yielded encouraging results, warranting further investigation. **Conclusion:** The FSASD-specific functional assays that we developed demonstrated their utility across different FSASD genotypes and cell types. Gene correction and selected small molecule drugs showed promise in rescuing the cellular phenotype and will be further investigated in our p.Arg39Cys knock-in mouse model. These functional assays hold potential for high-throughput drug screening, offering a valuable tool for identifying disease-modifying therapies for FSASD patients.

Session Title: Genetic Therapies Poster Session III

PB2009 Characterization of a novel mouse model with 35 bp deletion in *Champ1* gene and exploration of AAV gene therapy to correct phenotypic consequences arising from *CHAMP1* mutations

Authors:

T. Carneiro, B. Strenkowski, M. Bakay, H. Hakonarson; Children's Hosp. of Philadelphia, Philadelphia, PA

Introduction: Chromosome alignment maintaining phosphoprotein 1 (CHAMP1) is crucial for kinetochore-microtubule attachment during metaphase in mitosis and for chromosome segregation, process known to be important for neurodevelopment (Hempel et al., 2015; Isidor et al., 2016). Mutations in *CHAMP1* were recently associated with global developmental delay, Intellectual disability, hypotonia and dysmorphic features (Hempel et al., 2015; Tanaka et al., 2016). Reported mutations results in a premature stop codon, loss of protein c-terminal region which is essential for CHAMP1 localization and protein dysfunction. Until now, variants are predicted to cause loss of function resulting in haploinsufficiency (Garrity et al., 2021). We propose a gene therapy treatment using animal model to rescue the phenotype developed due to *CHAMP1* mutations. **Methodology:** In order to mimic *CHAMP1* disfunction, an animal model was created through CRISPR/Cas9 gene editing technology in mouse zygotes where a 35bp deletion was made in *CHAMP1* exon 2. The mouse model was developed using the C57BL/6NCrl strain by the mutant mouse resource and research center at UC Davis. Phenotype characterization is executed through behavior tests (mice with 8-10 weeks old), divided in two batteries: Motor and reactivity battery and cognitive battery. In the first one, information about motor delay, muscular strength and overall health of the mice is obtained with the results from open field, rotarod and catwalk gait test. The second battery provides information about learning ability, memory, and social impairment through the performance of Novel object recognition, acoustic startle response (ASR) and social choice. Upon completion of comprehensive phenotyping of the Champ1 mice, a new cohort is treated with AAV9 vector (Penn vector Core, University of Pennsylvania, Philadelphia, PA, USA) carrying the wildtype copy of *CHAMP1* gene. **Results:** To perform the phenotype characterization, 4 cohorts were submitted for behavioral testes, with 2 cohorts (20 mice) per battery. The offspring that composes those cohorts was generated by a breeding trio (a male heterozygous for *Champ1* deletion and two wildtype B6J females obtained from Jax Lab). Champ1 mice present delay in motor development and impairment in memory and learning abilities. Once this phonotype is confirmed by our studies, a gene therapy using AAV9 will be administered in an attempt to rescue the impairments observed, with the behavioral tests repeated in new cohorts. Molecular measurement of CHAMP1 both at gene and protein levels will be conducted for disease confirmation.

Session Title: Genetic Therapies Poster Session I

PB2010 Characterizing novel *in vitro* systems for the investigation of human post-traumatic fibrosis.

Authors:

J. Palmer[1], N. Cubbage[1], E. Rentschler[2], J. Ferreira de Vasconcellos[1]; [1]James Madison Univ., Harrisonburg, VA, [2]Blue Kick, Westerville, OH

Dangers of modern life, such as automobile crashes, severe burns and orthopedic surgery yield large numbers of traumatic neuromuscular injuries closely associated with the development of post-traumatic fibrosis. Modulation of the local inflammatory response following injury is essential for wound healing, however, severe traumatic injuries initiate an exacerbated response that compromises efficient tissue regeneration and leads to the formation of fibrotic tissue. TGFβ plays an essential role in tissue fibrosis and is highly expressed in human muscle after injury. However, robust *in vitro* systems to investigate the molecular mechanisms underlying human post-traumatic fibrosis are still needed. Here we investigated two *in vitro* culture systems used in a novel way in the investigation of the molecular mechanisms associated with human post-traumatic fibrosis. Bone marrow mesenchymal stem cells (BM-MSCs) were cultured in (i) nanofiber plates, comprised of a 700 nm diameter electrospun polycaprolactone nanofiber matrix randomly oriented and (ii) the Petaka™ cell culture system, designed to perform cell culture in an environment protected from external conditions. BM-MSCs were treated with TGFβ [10 ng/mL] for 48 h as a fibrotic model. Gene expression analysis for the fibrotic genes *ACTA2*, *VIM*, *COL1A1* and *COL3A1* was performed by q-RT-PCR. BM-MSCs treated with TGFβ and cultured in the nanofiber plates demonstrated no changes on *ACTA2* (control: 1.0 ± 0.0; TGFβ: 1.1 ± 0.1) and *VIM* (control: 1.0 ± 0.0; TGFβ: 1.0 ± 0.1), and down-regulation of *COL1A1* (control: 1.1 ± 0.1; TGFβ: 0.6 ± 0.4) and *COL3A1* (control: 1.0 ± 0.0; TGFβ: 0.6 ± 0.1). BM-MSCs treated with TGFβ and cultured in the Petaka™ system demonstrated up-regulation of *ACTA2* (control: 1.0 ± 0.0; TGFβ: 1.6 ± 0.4), *VIM* (control: 1.0 ± 0.0; TGFβ: 1.3 ± 0.2) and *COL3A1* (control: 1.2 ± 0.3; TGFβ: 4.7 ± 4.1), while *COL1A1* (control: 1.4 ± 0.5; TGFβ: 0.8 ± 0.4) was down-regulated. For comparison, BM-MSCs treated with TGFβ and cultured in a regular tissue culture plate demonstrated the up-regulation of the fibrotic genes *ACTA2* (control: 1.0 ± 0.0; TGFβ: 2.1 ± 1.4), *COL1A1* (control: 1.0 ± 0.0; TGFβ: 2.4 ± 1.0) and *COL3A1* (control: 1.0 ± 0.0; TGFβ: 1.4 ± 0.5), while *VIM* (control: 1.0 ± 0.0; TGFβ: 0.9 ± 0.1) remained unchanged. Additional studies need to be performed to validate these systems and understand the biological significance of these findings. Altogether, our current results support the notion that the Petaka™ cell culture system may be used for the *in vitro* investigation of human post-traumatic fibrosis and may be a useful culture system for future translational studies.

Session Title: Genetic Therapies Poster Session II

PB2011 Characterizing the role of the *IGF2BP1-3* genes during human osteogenic differentiation.

Authors:

E. Melton, L. Robinson, H. Cole, X. Yin, N. Omandan, N. Cubbage, J. Palmer, J. Ferreira de Vasconcellos; James Madison Univ., Harrisonburg, VA

Hazards of modern life, such as automobile crashes associated with a bone fracture or dislocation yield large numbers of traumatic neuromuscular injuries closely associated with the development of post-traumatic ectopic bone, where extraskeletal bone is formed in muscle and soft tissues. *Let-7* miRNAs are known to be upregulated in the muscle tissues of patients that formed ectopic bone compared to patients who did not develop ectopic bone after injury. Here we characterized the expression of *let-7* miRNAs and the insulin-like growth factor 2 binding protein genes (*IGF2BP1*, *IGF2BP2* and *IGF2BP3*), known targets of the *let-7* miRNAs, during human *in vitro* osteogenic differentiation and hypothesized that the *IGF2BP1-3* genes may play a role in human bone formation. Bone marrow mesenchymal stem cells (BM-MSCs) were induced to osteogenic differentiation for 14 days. BM-MSCs cultured in conventional growth media were used for comparison. RNA was harvested at the timepoints day 0 (before osteogenic differentiation), and days 7 and 14 of osteogenic differentiation. Gene expression was investigated by q-RT-PCR. *Let-7a*, *let-7d* and *let-7f* were upregulated at osteogenic differentiation day 14 to levels that did not reach statistical significance (p=.281, p=.200 and p=.192, respectively). *IGF2BP1* and *IGF2BP2* were significantly downregulated at osteogenic differentiation day 14 (p=.002 and p=.015), while *IGF2BP3* remained unchanged (p=.319). As expected, the osteogenic marker *ALP* was upregulated at days 7 and 14 (p=.007 and p=.012, respectively) during osteogenic differentiation. Knockdown of the *IGF2BP1* (BP1-KD), *IGF2BP2* (BP2-KD) and *IGF2BP3* (BP3-KD) genes were performed in BM-MSCs followed by 14 days of osteogenic differentiation. A scramble control was used for comparison. RNA was harvested at timepoints day 0 (48 h after transfection, but prior to osteogenic differentiation), and days 7 and 14. Gene expression was investigated by q-RT-PCR and alizarin red staining was performed to identify calcium deposits at day 14. Knockdowns were confirmed 48 h after transfection. Following BP1-KD and BP2-KD, the osteogenic gene *ALP* was upregulated at day 7 (p=.03 and p=.49, respectively) and day 14 (p=.09 and p=.10, respectively) compared to controls. Following BP3-KD, no significant effects were observed in the expression of osteogenic genes (osteocalcin, *ALP* and *CBFA1*) at days 7 and 14. Finally, we observed more calcium deposits in the BP1-KD compared to controls, while no differences were observed in the BP2-KD and BP3-KD compared to controls. Overall, these results support the notion that the *IGF2BP1* gene may play a role during human osteogenic differentiation.

Session Title: Genetic Therapies Poster Session III

PB2012 Clinical and translational studies in in-vitro and an *in-vivo* mouse model of a unique HSPB8 associated vacuolar myopathy.

Authors:

A. Shmara[1], L. Weiss[1], P. Pal[1], B. Tedesco[2], H. Zhang[1], P. Farahat[1], V. Boock[1], M. Wu[1], A. Panchagatti[1], S. Metwali[1], A. Villalta[1], X. Xu[1], A. Poletti[2], V. Kimonis[1]; [1]Univ. of California, Irvine, Irvine, CA, [2]Università degli Studi di Milano, Milano, Italy

HSPB8 associated autosomal dominant rimmed vacuolar myopathy is caused mainly by frameshift (fs) mutations. Patients develop a distal myopathy in their thirties, with progressive generalized weakness. Muscle biopsies show fatty replacement, fibrosis, rimmed vacuoles, and aggregates. We have shown that the mutant HSPB8 protein causes increased aggregation and secondary loss of activity of the normal HSPB8. We propose that microRNAs against the toxic Hspb8 mutation in addition to replacement of total endogenous HSPB8 is a logical therapeutic approach. We have generated stem cells from patient skin fibroblasts and transformed them into myoblasts to study the pathogenesis of the disease, and for translational studies. We found reduced HSPB8 protein, increased TDP-43, increased autophagy, as well as aggregations in patient fibroblasts. Using CRISPR technology, a knock-in *Hspb*8 mouse model was made of the c.515dupC fs variant seen in two families. The mice showed muscle weakness starting at 6 months of age. Muscle pathology revealed central nuclei, muscle degeneration, fatty replacement, reduced HSPB8, increased TDP-43, autophagy pathology, and amyloid fibril aggregates recapitulating the clinical phenotype. For our proof of principle gene therapy delivery studies, we have (1) transfected patient iPSC-derived myoblasts with wild-type HSPB8 and found the increase in HSPB8 protein as well as improved TDP-43 pathology; (2) dosed *Hspb*8 c515/+ mice with myotropic AAVmyo-CAG-EGFP by systemic delivery to check the muscle tropism effect. Results showed strong expression of the AAVmyo at 1E+12vg dose in all muscle groups vs. AAV9-CAG-EGFP control. Currently, the *Hspb*8 c515/+ mice are being treated with the AAVmyo-CBH-hHSPB8 construct that incorporates the normal *Hspb8*. Muscle specific promoters (CK8 and MHCK7) are also tested to select the most potent promoter for muscle delivery. Our next step is treating the animals with the best combination of microRNA and/or normal *Hspb8* constructs using AAVmyo with the most effective promoter. This will be followed by evaluating the amelioration of muscle weakness and pathology. Results of our studies holds promise for treatment of HSPB8 associated myopathy in patients.

Session Title: Genetic Therapies Poster Session I

PB2013 Collaborative Research Efforts Drive Therapeutic Advances for Free Sialic Acid Storage Disorder (FSASD)

Authors:

M. Malicdan[1], D. Adams[1], P. Brooks[2], K. Dobrenis[3], W. A. Gahl[4], B. Gasnier[5], M. Hackbarth[6], M. Hossain[7], P. Leoyklang[8], C. Anne-Longin[5], L. Paavola[9], F. Platt[10], L. Pollard[11], R. Reimer[12], M. Sabir[13], S. U. Walkley[3], R. Wang[14], M. Wasserstein[15], L. Wolfe[16], R. Zoncu[17], FSASD Consortium, M. Huizing[13]; [1]NIH, Bethesda, MD, [2]Office of Rare Disease Res., NCATS, NIH, Bethesda, MD, [3]Dept. of NeuroSci., Albert Einstein Coll. of Med., Bronx, NY, [4]NHGRI (NIH), Bethesda, MD, [5]Université Paris, SPPIN, CNRS, Paris, France, [6]NIH Undiagnosed Diseases Program, Bethesda, MD, [7]NHGRI, Bethesda, MD, [8]NIH Undiagnosed Diseases Program, Severn, MD, [9]Neural Ctr. of Neuropsychology, Neural Ltd, Oulu, Finland, [10]Dept. of Pharmacology, Univ. of Oxford, Oxford, United Kingdom, [11]Greenwood Genetic Ctr., Greenwood, SC, [12]Dept. of Neurological Sci., Stanford Univ., Palo Alto, CA, [13]NIH, Bethesda, MD, [14]CHOC Children's Specialists, Orange, CA, [15]3Dept. of Pediatrics and Genetics, The Children's Hosp. at Montefiore,, Bronx, NY, [16]NIH - NHGRI, Bethesda, MD, [17]Dept. of Molecular and Cell Biology, Univ. of California Berkeley, Berkeley, CA

Background: Free sialic acid storage disorder (FSASD) is an extremely rare, autosomal recessive, neurodegenerative, multisystemic disorder caused by defects in the lysosomal sialic acid (SA) exporter SLC17A5 (sialin). FSASD presents a broad clinical spectrum, ranging from mild neurodevelopmental disease to severe infantile forms with fatal outcomes. This study highlights the significant progress achieved in therapeutic development for FSASD through collaborative research efforts. **Methods:** To address the challenges in FSASD research, a highly collaborative multidisciplinary research consortium composed of clinicians, researchers from various specialties, and an FSASD patient advocacy group (S.T.A.R.) was formed. The consortium implemented parallel clinical and preclinical efforts to advance understanding and therapeutic development for FSASD. **Results:** The collaborative research consortium achieved remarkable advances in expanding our understanding of FSASD and identifying potential therapeutic strategies. By leveraging the collective expertise of clinicians, researchers, and patient advocates, a comprehensive approach to FSASD research was established. A pilot Natural History study and a Patient Registry were initiated to gather valuable data and biobank on the disease. A knock-in mouse closely resembling the Finnish *SLC17A5* founder variant (R39C) was generated, faithfully replicating key features of the human FSASD condition and providing valuable insights into disease pathology. Primary cells from FSASD patients and FSASD model mice were utilized to explore various approaches, including gene complementation, base editing, and candidate small drug screenings, which showed promise in ameliorating the cellular FSASD phenotype. Pilot studies investigating adeno-associated virus tropism in mice, a critical step for gene therapy design, yielded encouraging results. **Conclusion:** The collaborative research efforts, driven by the expertise of clinicians, researchers, and patient advocates, have yielded significant progress in data collection and therapeutic development for FSASD. Through the collaborative framework, scientific, clinical, and financial challenges have been overcome, leading to a deeper understanding of the disease and the identification of potential therapeutic strategies. The success of the collaborative model serves as an inspiring example for accelerating therapeutic advancements in rare diseases like FSASD, offering new hope for improved treatments and better outcomes for individuals affected by this condition.

Session Title: Genetic Therapies Poster Session II

PB2014 Developing a framework for precision therapy for rare and genetic diseases

Authors:

S. Kerr[1], J. Zschocke[2], J. Berg[3]; [1]Univ. of Dundee, Dundee, United Kingdom, [2]Med Univ Innsbruck, Innsbruck, Austria, [3]Univ Dundee, Dundee, United Kingdom

Introduction. The identification of the majority of genes that cause monogenic disorders in humans over the past 30 years has led to the possibility of developing highly specific therapies that truly modify disease progression. Multiple different approaches have been developed that are effective in different genetic disorders, and an exponential increase in highly specific therapies (that address a specific gene activity or even a specific pathogenic variant) is expected. A conceptual framework for classifying precision therapies is required to improve understanding amongst all healthcare professionals, and to facilitate design of therapies for new rare indications.
Methods. We identified newly emerging therapies for rare and genetic diseases through a journalistic review and from online clinical trials databases. Common mechanisms of action were identified and mapped to the central dogma of protein production. ***Results.*** We identified 10 specific mechanisms of action.
I. Gene Replacement, II. Gene Editing, III. Epigenetic Modification, IV. RNA Knockdown, V. RNA Editing, VI. Translation (stop codon readthrough), VII. Post Translational Modification, VIII. Protein Interaction, IX. Protein Replacement, X. Environmental Modification
This framework supports a common understanding of the place of action of different therapeutic interventions. It also provides a structure for teaching about emerging new classes of therapeutic agents. Such therapies are currently largely restricted to the precise treatment of rare disease, but application of the same technologies to the treatments of a wide range of common medical disorders is already emerging.

Session Title: Genetic Therapies Poster Session III

PB2015 Developing an approach to screening rare genetic diagnoses for amenability to bespoke genetic therapy development.

Authors:

D. Cheerie[1,2], M. C. Lauffer[3], M. Synofzik[4,5], D. Beijer[4], N=1 Collaborative, M. Szego[2], K. Amburgey[1,2], B. Kalish[1,2], J. J. Dowling[1,2], G. Costain[1,2]; [1]The Hosp. for Sick Children, Toronto, ON, Canada, [2]Univ. of Toronto, Toronto, ON, Canada, [3]Dutch Ctr. for RNA Therapeutics, Leiden, Netherlands, [4]Hertie-Inst. for Clinical Brain Res., Tübingen, Germany, [5]German Ctr. of Neurodegenerative Diseases, Tübingen, Germany

Background: Rare genetic diseases are major contributors to pediatric morbidity and mortality, but there are few genetic therapies that target the underlying pathomechanism(s) at a DNA or RNA level. Proof-of-concept exists for precision genetic therapies like antisense oligonucleotides (ASOs) that are customized for an individual's specific genetic variant and/or ultra-rare condition. A recent analysis of public databases suggested that ~50% of pathogenic variants may have a disease mechanism tractable to amelioration with an ASO. However, this preliminary study did not account for key molecular genetics, clinical, and ethical considerations, and was restricted to one kind of genetic therapy. We hypothesize that less than 5% of genetic diagnoses in real-world patient cohorts offer the realistic prospect for precision therapy development, based on our current understanding of the associated pathomechanisms, natural history, and contemporary treatment methods. **Objective and Methods:** Our study aims to develop and apply a reproducible multi-factorial approach to screening rare genetic diagnoses for amenability to bespoke genetic therapy development. Via modified Delphi techniques, we aim to define clinical, molecular genetics, and bioethical principles that can inform suitability for precision treatment. We will then apply our approach to two well-characterized pediatric cohorts that underwent clinical genome sequencing (total n~1000).**Preliminary Results:** Using a preliminary workflow defining amenability to ASO therapies, 1 of the first 43 diagnosed pediatric patients qualified for exon skipping ASOs on a molecular basis. Despite being the sole ideal candidate based on its molecular principles, the condition is non-progressive and may respond to standard anti-epileptic medication, mitigating the need for developing a precision genetic therapy for this variant. Therefore, as expected, no patients from our preliminary cohort have a realistic prospect for precision therapy development.**Conclusion:** Our initial findings underscore the importance of considering molecular, clinical, and bioethical factors to better reflect the number of genetic diagnoses amendable to genetic therapy. Our proposed methodologies and results promise to help inform large-scale translational research initiatives and healthcare system policies related to the development and provision of precision genetic therapies.

Session Title: Genetic Therapies Poster Session I

PB2016 Development of a human genetics-guided priority score for 19,365 genes and 347 drug indications

Authors:

A. Duffy[1], B. Petrazzini[1], D. Stein[1], J. Park[1], I. Forrest[1], K. Gibson[1], C. Marquez-Luna[1], M. Mort[2], A. Schlessinger[1], Y. Itan[1], D. Cooper[2], D. Jordan[3], G. Rocheleau[1], R. Do[1]; [1]Icahn Sch. of Med. at Mount Sinai, New York, NY, [2]Cardiff Univ, Cardiff, Wales, United Kingdom, [3]Icahn Sch. of Med. at Mt Sinai, New York, NY

Studies have shown that human genetic data can inform clinical trial outcomes. The development of drug prioritization frameworks that leverage the vast array of genetic data is necessary to aid the selection of effective drug targets at the preclinical stage of drug discovery.

In this study, we developed an *in silico* genetic priority score (GPS) for 19,365 protein-coding genes and 347 drug indications to inform drug target prioritization. We constructed an integrated drug genetics dataset using the Open Targets database for drug indications and leveraged eight genetic features which encompassed clinical variants, coding variants and phenome-wide association studies. Using Firth logistic regression with five-fold cross-validation, we built the GPS as the weighted sum of the effects of each of the 8 genetic features with drug indication. We further incorporated the direction and mechanism of the drug indication into a complementary version of the GPS. Hence, this framework prioritizes genes as suitable drug targets based on the number and strength of each predictor with genetic support.

First, we stratified the GPS into bins of 0.3 to confirm that gene-phenotypes with increased genetic support were associated with an increased likelihood of having a drug indication. A GPS in the top 4%, 0.3% and 0.2% conferred a 5.2-, 8.3- and 8.8-fold increased effect of having an indication, respectively. Importantly, the top scores were driven by multiple genetic features (mean=3.1). We validated the results using an independent drug database, SIDER and observed similar results. We demonstrated in a prioritization framework, that gene-phenotype observations with a GPS in the top 0.2% were 4.2-fold more likely to be a therapeutic target compared to a random selection. Furthermore, we showed that targets with a score in the top 0.2% were 1.3, 4.5 and 7.8-fold more likely to advance from phase I to phase II, III and IV, respectively.

Next, we ascertained the GPS for 19,365 genes and 347 phenotypes. We note that the great majority of observations (98.7%) linking a gene to a phenotype with a high GPS score were not found in either the Open Targets or SIDER resources, indicating potential new drug targets. Finally, we highlight both known and novel examples of gene targets supported by a high GPS that are consistent with the direction and mechanism of action of the drug.

In conclusion, we have developed an *in silico* GPS that can be used to inform the prioritization of gene targets for a large number of genes and diseases in drug development.

Session Title: Genetic Therapies Poster Session II

PB2017 Development of an myotropic adeno-associated virus (AAV) for VCP related myopathy

Authors:

L. Weiss, P. Pal, H. Zhang, F. Philip, A. shmara, V. Boock, C. Cheng, S. Metwaly, G. Wu, A. Villalta, X. Xu, V. Kimonis; Univ. of California, Irvine, Irvine, CA

Pathogenic variants in Valosin Containing Protein (VCP) gene cause a unique autosomal dominant disease characterized by inclusion body myopathy, Paget disease of bone and frontotemporal dementia, also known as multisystem proteinopathy (MSP). VCP pathogenic variants lead to hyperactive enzymatic activity, suggesting a gain-of-function. To ameliorate the gain-of-toxicity of VCP mutant proteins, an ideal approach is to silence the mutant gene in an allele-specific manner and to leave the wildtype allele intact which is challenging. We therefore propose co-expression of VCP silencing with rescue vectors to replace endogenous VCP with exogenous functional human VCP. We have designed (1) 3 artificial microRNA against VCP, (2) a codon-optimized wildtype VCP, both tested in HEK293T cells, (3) muscle trophic AAV9 variant: AAVmyo-CAG-EGFP vector, tested in 8 weeks old C57BL/6 and VCP R155H/+ mice. The results show: (1) The best microRNA, mirVCP2 (2µg, ratio of hVCP: mirVCP2 = 1:5, transfection efficacy >70%), achieved the reduction of 75 % VCP mRNA and 55% VCP protein expression in HEK293T cells. (2) The codon-optimal WT VCP was not targeted by mirVCP2 in HEK293T cells. (3) In both WT, VCP mice infected with $4x10^{12}$vg of AAVmyo-CAG-EGFP, the EGFP protein expression is convincingly higher in muscle-specific AAVmyo vector compared to AAV9 in skeletal and cardiac muscles. We then use mirVCP2 and AAVmyo for virus packaging and inject in a well-established VCPR155H/R155H mice at 4 days old. Currently, VCP mice are injected systematically by retro-orbital injection with 1e12 vector genome of virus to determine the safety and efficiency including specific muscle tropism, effective virus dosage, viral immune response, and toxicity. Success in these preclinical studies holds potential for promising therapeutic benefits in patients with VCP disease.

Session Title: Genetic Therapies Poster Session III

PB2018 Dose-ranging pre-clinical studies of systemic AAV9 with codon-optimized reduced size ATP7A (cors-ATP7A) plus subcutaneous Cu-Histidinate in a Menkes disease model

Authors:

S. Kaler[1], L. Venkataraman[1], T. Kaniganti[1], B. Kennedy[1], M. Gelb[2], J. Reese[3], D. Goldstein[4], P. Sullivan[4]; [1]Ctr. for Gene Therapy, Nationwide Children's Hosp., Columbus, OH, [2]Dept. of Chemistry, Univ. of Washington, Seattle, WA, [3]Sentynl Therapeutics, Solana Beach, CA, [4]NINDS/NIH, Bethesda, MD

Menkes Disease (MD) is an infantile onset, X-linked recessive disorder of copper transport characterized by rapidly progressive neurodegeneration, seizures, and failure to thrive caused by mutations in *ATP7A*, a copper transport gene. Prior studies have shown that early intervention with CuHis alone improves survival and clinical outcomes, yet the overall burden of disease remains considerable. Based on the multi-tasking character of ATP7A, MD infants may benefit from viral gene therapy that provides working copies of *ATP7A*. *Mo-br* mice harbor a deletion in *Atp7a*, the murine homolog, and recapitulate salient features of MD, including low serum copper levels, neurodegeneration, and early death (postnatal day 13). We previously documented 23% and 53% respective rescues of this model with CSF-directed AAV5- or AAV9- mediated transfer of human *ATP7A*, in combination with copper. Here, we evaluated a systemic gene therapy approach and performed dose-ranging preclinical studies of intravenous AAV9 delivering a codon-optimized, reduced size version of *ATP7A* (cors*ATP7A*) plus 30µg clinical grade CuHis by subcutaneous injection. AAV biodistribution in 12-day wildtype and *mo-br* mice revealed efficient transduction in various organs except the renal medulla, reflecting inability of AAV particles to effectively traverse renal glomeruli and reach the distal nephron. Safety, efficacy, and toxicological studies identified two doses, $2.6x10^{12}$ and $2.6x10^{13}$ vg/kg body weight + CuHis as highly promising, with combined long-term survival of 89% (n=36) versus 60% in mice receiving CuHis alone (n=10) and 0% for AAV9-cors*ATP7A* alone. Overall outcomes (serum and brain copper, brain neurochemical ratios, growth, neurobehavior) appear superior in the $2.6x10^{13}$ vg/kg + 30µg CuHis cohort survivors (n=19). Administration of a higher dose, $2.6x10^{14}$ vg/kg, was initially well-tolerated, but sudden death occurred in 5 of 6 animals after CuHis administration on P4 or P5. Postmortem analyses showed no evidence of brain, or liver toxicity however acute renal tubular necrosis and high renal copper levels were noted and considered the cause of death. These studies in the *mo-br* mouse represent important steps toward a future pilot gene therapy clinical trial for individuals with this illness. Development of newborn screening for MD, based on detection of metabolites in the catecholamine biosynthetic pathway in conjunction with *ATP7A* variant detection by whole genome sequencing, both from dried blood spots, is also underway. In combination, the potential impact of this progress on clinical practice is high since the largest barriers to good health for patients with MD will be circumvented.

Session Title: Genetic Therapies Poster Session I

PB2019 Endogenous expression of *AGBL5* or reduction of *TTLL5* expression rescues the cellular disease phenotype in *AGBL5*$^{-/-}$ human retinal pigment epithelium cells.

Authors:

S. Villa Vasquez[1], R. Ewing[2], R. Pengelly[1], D. I. Wilson[1], D. Baralle[1], G. Wheway[1]; [1]Univ. of Southampton, Faculty of Med., Southampton, United Kingdom, [2]Univ. of Southampton, Faculty of Environmental and Life Sci., Southampton, United Kingdom

AGBL5 is a dual function deglutamylase that regulates functional tubulin glutamylation levels in cilia. Mutations are associated with a form of autosomal recessive Retinitis Pigmentosa, for which there are no known treatments or cures. This study investigated *AGBL5* CRISPR knockout ARPE19 cells, to further elucidate the mechanism causing retinal dysfunction and investigate opportunities for rescue of the cellular disease phenotype.
AGBL5$^{-/-}$ cells have reduced levels of *AGBL5* mRNA and AGBL5 protein, show hyperglutamylation, significantly shorter cilia, and a lower percentage of ciliated cells than WT cells ($p<0.001$). Whole transcriptome RNA-seq analysis showed that genes differentially expressed in *AGBL5*$^{-/-}$ cells are associated with nervous system development, neuron apoptotic process, cell adhesion, protein localisation to microtubule, photoreceptor cell maintenance, and other ciliopathies.
To counter the glutamylation imbalance due to loss of AGBL5, *TTLL5* which encodes tubulin tyrosine ligase like 5 and functions as an alpha tubulin polyglutamylase, was considered for further experiments. We hypothesised that knockdown of *TTLL5* could rescue the cellular disease phenotype in *AGBL5*$^{-/-}$ cells. Indeed, we found that WT AGBL5-eGFP expression in mutant cells, or CRISPR knockout of *TTLL5* robustly rescued the ciliary phenotype and glutamylation levels in *AGBL5*$^{-/-}$ cells.
This work reports a clear phenotype in *AGBL5*-deficient human retinal cell lines consisting of hyperglutamylation, short cilia and reduced ciliogenesis. It provides functional and transcriptomic data and protein interaction data from this cell model as a resource to the ciliopathy research community. Crucially, the demonstrates the potential effectiveness of therapy to treat *AGBL5*-associated cilia deficiency with endogenous expression of AGBL5 or reduction of *TTLL5* expression.

Session Title: Genetic Therapies Poster Session II

PB2020 Epigenome Editing Durability Varies Widely Across Cardiovascular Disease Target Genes

Authors:

X. Wang, M. N. Whittaker, L. C. Testa, A. Quigley, I. Jindal, S. V. Cortez-Alvarado, P. Qu, Y. Yang, M-G. Alameh, K. Musunuru; Univ. of Pennsylvania, Philadelphia, PA

Background: Hepatic knockdown of the proprotein convertase subtilisin/kexin type 9 (*PCSK9*) gene or the angiopoietin-like 3 (*ANGPTL3*) gene has been demonstrated to reduce blood low-density lipoprotein cholesterol (LDL-C) levels, and hepatic knockdown of the angiotensinogen (*AGT*) gene has been demonstrated to reduce blood pressure. Genome editing can productively target each of these three genes in hepatocytes in the liver, offering the possibility of durable "one-and-done" therapies for hypercholesterolemia and hypertension. However, concerns around making permanent gene sequence changes via DNA strand breaks might hinder acceptance of these therapies. Epigenome editing offers an alternative approach to gene inactivation, via silencing of gene expression by methylation of the promoter region, but the long-term durability of epigenome editing remains to be established.
Objectives and Results: We assessed the ability of epigenome editing to durably reduce the expression of the human *PCSK9*, *ANGPTL3*, and *AGT* genes in HuH-7 hepatoma cells. Cells treated with the CRISPRoff epigenome editor and *PCSK9* guide RNAs were maintained for up to 124 cell doublings and demonstrated durable knockdown of gene expression and increased CpG dinucleotide methylation in the promoter, exon 1, and intron 1 regions. In contrast, cells treated with CRISPRoff and *ANGPTL3* guide RNAs experienced only transient knockdown of gene expression. Cells treated with CRISPRoff and *AGT* guide RNAs also experienced transient knockdown of gene expression; although initially there was increased CpG methylation throughout the early part of the gene, this methylation was geographically heterogeneous—transient in the promoter, and stable in intron 1.
Conclusions: This work demonstrates precise and durable gene regulation via methylation, supporting a new therapeutic approach for protection against cardiovascular disease via knockdown of genes such as *PCSK9*. However, the durability of knockdown with methylation changes is not generalizable across target genes, likely limiting the therapeutic potential of epigenome editing compared to other modalities.

Session Title: Genetic Therapies Poster Session III

PB2021 Establishing predictive models for cystic fibrosis gene editing therapeutic development

Authors:

J. Conte, A. Bell, M. Armstrong, M. Chado, N. Allaire, A. Sivachenko, M. Mense, C. Cotton, H. Valley; Cystic Fibrosis Fndn., Lexington, MA

Gene editing and gene therapy are promising approaches to cure cystic fibrosis (CF) and there is intense, ongoing research and development in this area. Primary human bronchial epithelial (hBE) cells are considered the "gold standard" for demonstration of efficacy of small molecule CFTR modulators. However, hBEs have limited expansion potential and available genotypes; therefore, other cell models are frequently used for early-stage gene editing research and development. HEK-293 cells are a widely used immortalized cell model for gene editing, but do not express CFTR. 16HBE cells, an immortalized cell model, are a valuable tool for CF-specific research. Inducible pluripotent stem cells (iPS), a cell model that can be programmed to differentiate into multiple tissue lineages, are also actively used in early-stage CF research. It is not known whether results obtained in these cell models will translate to cultured primary CF hBEs and ultimately in vivo. The overall goal of this project is to establish if gene editing outcomes in CFTR are correlated between HEK-293, 16HBE, iPS, and primary hBE cells.

We used next generation sequencing to assess indel formation frequency caused by Spy-Cas9:gRNA at eighty unique targets in WT genomic CFTR in each of these four cell models. We designed eight guide RNAs for each of ten genomic CFTR regions: the -20.9 kb regulatory element (RE), 5' UTR, intron 1 RE, the 3' end of intron 1, exon 1, intron 11, exon 12, intron 12 RE, exon 23, and the 3' UTR. Guide RNAs were complexed with Cas9 to form a ribonucleoprotein complex, which was delivered by nucleofection. Following nucleofection, cells were plated and allowed to recover for 72 hours before genomic DNA was harvested. Targeted, next generation sequencing for each of the regions was performed. We compared indel frequency for each of these 80 guides between HEK-293, 16HBE, hBE, and iPS cells.

These data provide robust, quantitative comparisons of the four cell models and inform whether HEK-293, 16HBE, and/or iPS cells are predictive of indel editing outcomes in CFTR in primary hBE cells. There is a strong correlation of indel editing frequency for the guides tested between all 4 cell models tested with R^2 values ranging from 0.39-0.70. The best correlation between hBE cells and another cell model is seen with 16HBE, R^2 = 0.65. The lowest correlation between hBE cells and another cell model is seen with iPS cells, R^2 = 0.48. While there are exceptions, evaluation of guides in HEK, 16HBE, and iPS cells can be used to select guides that are likely to be highly successful in primary hBE cells for CFTR.

Session Title: Genetic Therapies Poster Session I

PB2022 Evaluating anti-diabetic drugs for protection against Alzheimer's disease by integration of genetics, genomics and drug-induced transcriptomic signatures.

Authors:

X. Zhong[1,2], Q. Wei[3,2], Y. Tan[4,2], B. Li[5,2], N. Cox[6,2]; [1]Vanderbilt Univ. Med. Ctr., Nashville, TN, [2]Vanderbilt Genetics Inst., Nashville, TN, [3]Vanderbilt Univ, Nashville, TN, [4]Vanderbilt Univ. - Nashville, TN, Nashville, TN, [5]Vanderbilt, Nashville, TN, [6]Vanderbilt Univ Med Ctr., Nashville, TN

Alzheimer's disease (AD) is the leading cause of dementia with very limited therapeutics. Several anti-diabetic drugs (e.g. Semaglutide) have shown some protection against cognition decline in recent clinical trials. Repurposing existing approved drugs for novel uses can expedite drug development. Drugs with support from human genetics are twice more likely to succeed to the market. In this study, we use our (previously developed) iRIGS framework to infer AD risk genes through integration of genetics, functional genomics and gene-gene networks. We then use these risk genes together with AD-associated direction of perturbation to construct a genetic perturbation signature. By examining the drug-induced perturbation signatures with more than 8700 perturbagens, we identified Sitagliptin, a DPP4 inhibitor, that ranked top (top 0.1%) for reversing the AD disease signature. Analyzing single cell RNAseq data from human AD brains, we identified cell-cell communications among microglia, excitatory neurons, inhibitory neurons, endothelial cells, and pericytes through formation of ligand-receptor pairs between *ADAM17* (a known AD risk gene) and *ITGB1* (a receptor). Given that *ITGB1* directly interacts with *DPP4*, this suggests that Sitagliptin may act to influence AD risk through disruption of the *ADAM17-ITGB1* interaction. In conclusion, we present a strategy for repurposing existing drugs for AD, identify a DPP4 inhibitor as a top candidate and provide further support from single-cell transcriptomics for the potential mechanism of action of the prioritized drug.

Session Title: Genetic Therapies Poster Session II

PB2023 Evaluating antisense oligonucleotides in spinocerebellar ataxia type 15/16.

Authors:

B. Masser, H. S. McLoughlin; Univ. of Michigan, Ann Arbor, MI

Spinocerebellar ataxias (SCAs) encompass a diverse group of dominantly inherited neurodegenerative ataxias, with over 50 SCAs identified to date. Over 40 genes are directly implicated in SCA pathogenesis, underlying their genetic and clinical heterogeneity. However, growing evidence suggests that neuronal calcium (Ca^{2+}) signaling perturbation plays a central role among SCAs. Various polyglutamine SCAs (SCA 1, 2, 3, 6, 7, and 17) have been reported to exhibit cerebellar Ca^{2+} signaling dysfunction, primarily attributed to the inositol 1,4,5-triphosphate receptor, type 1 (IP_3R1). IP_3R1 is a Ca^{2+} channel situated in the membrane of the endoplasmic reticulum. It performs a crucial function in ensuring the survival of Purkinje neurons, the cerebellum's sole output cells. Macrodeletions resulting in haploinsufficiency of the gene encoding IP_3R1 (*ITPR1*) cause SCA15/16, a slowly progressive, pure cerebellar ataxia. To date, no treatment options are available for SCA15/16 patients. Traditional gene replacement strategies are not applicable to SCA15/16 due to the considerable size of *ITPR1* (~10 kb coding sequence), which exceeds viral delivery packaging capacities. Here, we propose utilizing antisense oligonucleotides (ASOs) to obstruct endogenous gene-silencing mechanisms, specifically microRNAs (miRNAs), to increase the expression of IP_3R1 via the remaining functional allele. We hypothesize that ASO binding to active miRNA recognition elements (MREs) within the *ITPR1* 3' untranslated region (UTR) will prevent miRNA-mediated silencing and restore IP_3R1 channel expression and function. Utilizing a dual luciferase reporter construct containing the *ITPR1* 3' UTR, we have validated several predicted *ITPR1*-miRNA interactions. We then designed "miR-masking" ASOs, complementary to these validated MREs and the flanking sequences. These miR-masks showed target site protection in our reporter system. Ongoing studies are evaluating the efficacy of the miR-masking ASOs to protect endogenous *ITPR1* transcripts and safety *in vitro* and *in vivo*. These studies have the potential to provide a promising therapeutic strategy for not only the treatment of SCA15/16 but also for other channelopathies caused by haploinsufficiency, as miR-masking allows for restoration rather than overexpression and upregulation of channel function.

Session Title: Genetic Therapies Poster Session III

PB2024 Experience with maintenance therapy of carglumic acid in patients with organic acidemia

Authors:

S. Hwang[1], J. Yoon[1], H. Bae[1], D. Kim[1], G-H. Kim[2], H. Shin[3], M. Kang[3], I. Choi[4], B. Lee[1,2]; [1]Dept. of Pediatrics, Asan Med. Ctr. Children's Hosp., Univ. of Ulsan Coll. of Med., Seoul, Korea, Republic of, [2]Med. Genetics Ctr., Asan Med. Ctr., Univ. of Ulsan Coll. of Med., Seoul, Korea, Republic of, [3]Asan Inst. for Life Sci., Asan Med. Ctr., Seoul, Korea, Republic of, [4]Dept. of Genetic Counseling, Univ. of Ulsan Coll. of Med., Seoul, Korea, Republic of

Background: Organic acidemias including propionic acidemia (PA) and methylmalonic acidemia (MMA) are associated with secondary hyperammonemia, leading to poor neurological outcomes. Carglumic acid (N-carbamylglutamate) has been considered a therapeutic option for hyperammonemia.
Aim: This study aims to evaluate the efficacy, safety, and impact on hospitalizations of carglumic acid in patients with organic acidemias and concomitant hyperammonemia.
Methods: Three patients with organic acidemias (2 PA, 1 MMA) have received carglumic acid in a dose of 50 mg/kg/day along with standard therapy. The primary outcome was hospitalization days due to hyperammonemia. Secondary outcomes included the number of emergency room (ER) visits, peak concentration of ammonia, duration of hyperammonemia, other biochemical markers, growth, and developmental assessment. Safety was assessed throughout the study.
Results: The patients (median age 0.9 years; range 0.2-4.0) were followed for a median of 1.1 years. Median hospitalization days per year decreased from 35 (range 15-259) to 12 (range 0-25) days compared to before carglumic acid treatment. Annualized rates of admission and emergency room visits were also reduced by 70% and 50%, respectively. Peak ammonia level decreased from 411 (range 232-615) to 69 (range 63-75) μmol/L. Duration of hyperammonemia was reduced by 83%. Other biochemical markers remain stable. No further deteriorations of growth and development were observed. There were no drug-related adverse events throughout the study.
Conclusions: Maintenance therapy with carglumic acid for organic acidemias was efficacious and generally tolerable with reductions in hospitalization days. Longer-term observation with more patients is necessary to validate this efficacy.

Session Title: Genetic Therapies Poster Session I

PB2025 Exploring the potential of readthrough compounds on nonsense mutation, R132* of *MMACHC* gene responsible for cobalamin C defect

Authors:

G. Kaur, R. Kaur, S. V. Attri, I. Verma, A. G. Saini, R. Gupta; Postgraduate Inst. of Med. Ed. & Res. (PGIMER), Chandigarh, India

Background: Cobalamin-C defect is the most common inborn error of intracellular cobalamin metabolism. Nonsense mutations account for majority of inherited genetic diseases. The nonsense-mutation, c.394C>T (R132*) in *MMACHC* gene results in truncated protein and is most common in North Indian population. Based on the location of this mutation, it can be targeted for readthrough drug therapeutics for restoring the MMACHC activity. We evaluated a range of readthrough compounds including aminoglycosides and non-aminoglycosides, to assess their ability to suppress PTC recognition. **Methods:** To check the efficacy of readthrough compounds like Geneticin (G418), PTC124 and Amlexanox, the stable mammalian cell lines expressing wildtype (WT) and mutant (R132*) MMACHC protein were created. WT cDNA clone of *MMACHC* gene was purchased and mutant clone (R132*_MMACHC) was prepared by site-directed mutagenesis. WT and R132*_MMACHC constructs were subcloned into mammalian expression vector (pcDNA5FRT) along with eGFP sequence. Vectors were transfected into Flp-In CHO cell line and selected using hygromycin antibiotic and GFP positive cells. Flp-In CHO cells carrying R132* variant was treated with different readthrough compounds at different concentrations and checked for protein expression by confocal microscopy and western blot. **Results:** Western blot showed good expression in transfected cells carrying WT MMACHC clone; however, no expression of GFP tagged protein was observed in mutant (R132*) cells suggesting no expression of MMACHC with this mutation. On treating R132* Flp-In CHO cells with different concentrations of G418 for 24 hrs, protein expression was restored at 15 µg/ml with relatively higher band intensities at 100-750 µg/ml of G418. PTC124 treatment for 24 hrs also restored the full length expression of MMACHC in the range 3-50 µM. Amlexanox showed restoration of MMACHC protein expression within the range of 50-250 µM. These results were also in concordance with the fluorescence analysed using confocal microscopy. **Conclusion:**The readthrough compounds (G418, PTC124 and Amlexanox) have the potential to restore the full-length functional protein from truncated MMACHC protein with nonsense mutation, R132* in stable cell lines. ***Keywords:*** *Cobalamin-C defect, readthrough compounds, MMACHC, nonsense mutation, stable cell lines*

Session Title: Genetic Therapies Poster Session II

PB2026 Genotype and phenotype analysis of individuals with nano-rare genomic variants considered for experimental antisense oligonucleotide therapy.

Authors:

C. Ngongang Tekendo[1], T. Cole[1], J. Carroll[2,1], J. Douville[1], A. Williford[1], S. Glass[1], J. Gleeson[1,3,4], C. Bennett[5,1], S. Crooke[1]; [1]n-Lorem Fndn., Carlsbad, CA, [2]Dept. of Neurology, Univ. of Washington, Seattle, WA, [3]Dept. of NeuroSci.s, Univ. of California, San Diego, La Jolla, CA, [4]Rady Children's Inst. for Genomic Med., San Diego, CA, [5]Ionis Pharmaceutical, Carlsbad, CA

Recent advances in OMICS technologies allow for the identification of an increasing number of individuals with diseases caused by extremely rare genetic variants. Because nano-rare variants (worldwide prevalence of <30) are so rare, these difficult-to-diagnose individuals are uniquely disadvantaged and pose significant challenges to healthcare systems and the society in large. Despite having diseases caused by actionable single genes variants in many cases, there is no commercial path for treatments for such a small patient population. Since antisense oligonucleotide (ASO) technology has proven to be suited to address the needs of a meaningful portion of these individuals, we have established an industrialized approach that couples detailed genotypic and phenotypic data to the immediate potential for ASO therapy. Here, we leverage our experience in assessing the causality of nano-rare genetic variants and associated proximal molecular pathological events to attempt a correlation between detailed genetic data with patient specific phenotypic observations in 173 nano-rare patients from diverse age groups evaluated for experimental ASO therapy. We found that the time required to achieve a molecular diagnosis varies from 1 month to 36 years, with the mean and median times from symptom onset to diagnosis estimated to be 4.37 years and 2 years respectively. There is a significant bias toward Neurological diseases, with diverse genes and functional families involved and a marked preponderance of variants in ion channels proteins. The variability in phenotypic expression associated with nano-rare variants in genes such as *H3F3A*, *GBE1*, *GNAO1*, *UBTF*, *PACS1* or *PACS2* clearly support previous observations that phenotypes associated with same variants in the same gene can vary. We equally observe that different, but functionally equivalent variants can result in both similar (e.g., *TARDBP*) and different phenotypes (e.g., *CACNA1A*, *KCNQ2*, or *MED13L*). Despite the relatively small size of the patient population investigated, our findings enable a variety of insights into the genotype and phenotype relationships in nano-rare conditions. Moreover, our data show that this unique patient population presents a remarkable opportunity to apply modern OMICS approaches to begin to understand the various homeostatic, compensatory, and secondary effects of these genetic variants on the networks that result in expression of their unique phenotypes.

Session Title: Genetic Therapies Poster Session III

PB2027 Hermansky-Pudlak Syndrome: Exploring the therapeutic effect of gene therapy for HPS-1 pulmonary fibrosis

Authors:

G. Nieto-Alamilla[1], M. Behan[2], S. Abudi[1], M. Hossain[1], W. Gahl[3], M. Malicdan[1]; [1]NIH, Bethesda, MD, [2]NHGRI, Bethesda, MD, [3]NHGRI (NIH), Bethesda, MD

Background: Hermansky-Pudlak Syndrome type 1 (HPS-1) is a rare autosomal recessive multisystemic disease caused by biallelic loss-of-function mutations in the *HPS1* gene. HPS1 is characterized by oculocutaneous albinism, a bleeding diathesis, and progressive restrictive lung disease known as pulmonary fibrosis (HPS-PF). Current treatment options for HPS-PF, such as lung transplantation, have limitations and significant risks, emphasizing the need for alternative therapeutic approaches. Gene therapy, which aims to correct the underlying genetic defect, holds promise for the treatment of HPS-PF. Here, we present a rationale and preliminary findings for HPS1 gene complementation as a potential therapeutic strategy.**Methods:** To assess the feasibility of gene therapy for HPS1 we utilized knockout and a knockin mouse models (Hps1-KO, and Hps1-KI) generated in our laboratory. We employed adeno-associated virus (AAV) vectors to deliver either the enhanced green fluorescent protein (eGFP) open reading frame, as well as the mouse (mHps1) or human *HPS1* (hHPS1) genes, either locally (intranasal) or systemically (intravenous). Tropism analysis was conducted to evaluate the distribution of GFP in multiple tissues. To evaluate if the effect of gene complementation on PF, we induced PF using bleomycin through osmotic subcutaneous delivery.**Results:** Tropism analysis revealed robust AAV expression in the lungs, including alveolar epithelial type 2 cells (AEC2), indicating effective targeting of AAV vectors to HPS1-deficient cells. Overexpression of the mHps1 gene in Hps1-KO mice showed no adverse effects. Notably, Hps1 mice treated with mHps1 exhibited some improvement in pulmonary compliance and stiffness, suggesting a promising therapeutic effect. Human HPS1 expression was present in the lungs of both Hps1 mouse models following hHPS1 delivery. The therapeutic effects of hHPS1 gene therapy are currently under assessment and will be the subject of further investigation.**Conclusion:** Our current results underscore the progress made in gene therapy for HPS1-PF. Successful generation of animal models, including a humanized mouse model and knockout model, has provided valuable tools for studying the disease and evaluating therapeutic interventions. In vivo studies showed encouraging results, highlight the potential of gene therapy as a transformative treatment for HPS1-PF, and pave the way for further research and clinical translation. Continued investigations are necessary to optimize the gene therapy approach and validate its long-term safety and efficacy for the benefit of HPS1-PF patients.

Session Title: Genetic Therapies Poster Session I

PB2028 † High frequency AAV integration into Cas9-induced DSBs: implications as a gene therapy vector in the brain

Authors:

H. Bazick[1], M. J. Zylka[1,2,3], H. Mao[4], J. K. Niehaus[1], J. M. Wolter[5]; [1]UNC NeuroSci. Ctr., UNC Chapel Hill, Chapel Hill, NC, [2]Carolina Inst. for Dev.al Disabilities, UNC Chapel Hill, Chapel Hill, NC, [3]Dept. of Cell Biology and Physiology, UNC Chapel Hill, Chapel Hill, NC, [4]Life Edit, Durham, NC, [5]Univ. of Wisconsin Madison, Madison, WI

Angelman Syndrome (AS), a severe neurodevelopmental disorder caused by disruption of the maternally-inherited ubiquitin protein ligase E3A (UBE3A) allele, is characterized by severe sleep dysfunction, seizure disorders, impaired cognition and speech, motor deficits, and microcephaly. Due to epigenetic silencing of the paternal allele of UBE3A (patUBE3A) by a long non-coding antisense RNA (*UBE3A-ATS*) in neurons, mutation or deletion of maternal *UBE3A* results in near complete loss of UBE3A protein in the brain. Our previous study revealed that adeno-associated virus (AAV)-CRISPR/Cas9 gene therapy targeted to 75-repetitive *SNORD115* sequences along *UBE3A-ATS* leads to integration of the AAV vector into Cas9 target sites, trapping *UBE3A-ATS*, unsilencing patUBE3A, and alleviating AS phenotypes in mice with a dual embryonic/post-natal day 1 injection. Based on this mechanism of unsilencing, we hypothesized that integration of AAV at a single unique site should sufficiently block *UBE3A-ATS* transcription to unsilence patUBE3A. Thus, we designed 25 guide RNAs (gRNAs) single sites within *UBE3A-ATS*. When delivered in an AAV vector with Cas9 to primary neuron cultures, the best gRNAs individually achieved only 46% knockdown (KD) of *UBE3A-ATS*. This contrasted with 80% KD when a single gRNA was used that targeted 78 *SNORD115* genes. Examination of comprehensive editing profiles following AAV/Cas9 treatment in neuron cultures using modified Anchored Multiplex PCR (AMP)-seq revealed that AAV integration accounts for approximately 80% of all editing events, while insertions/deletions (indels) account for only 20%. Using a reporter assay, we found that the most frequent indels failed to block gene transcription, while integration of AAV-derived elements disrupted downstream gene expression. Our study shows that AAV integration is a frequent outcome of AAV/Cas9 therapy in neurons, suggests that AAV-derived elements are capable of disrupting genes, and that AAV integration may be the primary mechanism of *UBE3A-ATS* knockdown in the treatment of AS. Our work has implications for AAV/Cas9 as a gene therapy vector and in considering the risks versus benefits of utilizing an active nuclease and integrating viral genomes in the brain.

Session Title: Genetic Therapies Poster Session II

PB2029 Inhibition of chemokine expression declines cancer cells proliferation and migration in vitro

Authors:

M. Abo-Zeid[1,2], M. Abo-Elfadl[2,3], S. Mostafa[2,3]; [1]Genetics and Cytology Dept., Biotechnology Res. Inst., Natl. Res. Ctr., Dokki 12622, Cairo, Egypt, [2]Cancer Biology and Genetics Lab., Ctr. of Excellence for Advanced Sci., Natl. Res. Ctr., Dokki 12622, Cairo, Egypt, [3]Biochemistry Dept., Biotechnology Res. Inst., Natl. Res. Ctr., Dokki 12622, Cairo, Egypt

Background: The role of silencing chemokines to lose their function promotes initiation of cell cycle arrest and inhibits cell division, proliferation and migration. Therefore, knocking down chemotactic gene markers of metastatic cancer cells obstacle their migration and invasion to new loci. **Methodology**: In the present study, different types of cancer cells were transfected with specific chemokine inhibitor silencing CXCR4 gene. The low observed adverse effect level concentration of the siRNA was used to monitor DNA damage response by comet assay, gene expression by qRT-PCR for CXCR4 gene expression as well as related cell division, invasion and migration genes markers, protein expression by immunofluorescent assay and cell migration by wound healing procedure to record the delay in cell migration process. **Results:** When different cancer cell lines were transfected with the chemokine inhibitor for 48 h by the LOAEL concentration, the percentages of DNA damage were increased remarkably, and the gene expression of CXCR4 was downregulated extensively along with protein expression on the tested cancer cell lines. Also, the gene expression of cell division, proliferation and invasion were down regulated effectively when cancer cells transfected with that CXCR4 inhibitor. Moreover, the gap area of migration between non-transfected cells was smaller than that observed on transfected cells remarkably. **In conclusion**, target gene therapy based on silencing chemokines could contribute to down regulate other gene markers responsible for cell division, proliferation and migration, which consequently could diminish metastasis to new loci leading to secondary tumors.

Session Title: Genetic Therapies Poster Session III

PB2030 Investigation of gene transcription modifications associated with obesity in*Danio rerio* (zebrafish)

Authors:

K. Dobrinski[1], R. Stephenson[1], B. Alston[1], J. Gran[1], S. Desai[2], B. Ward[3], J. Kwon[4], Q. Zhu[4], C. Lee[4]; [1]Univ. of Tampa, Tampa, FL, [2]Coll. of Osteopathic Med. of the Pacific COMP-Northwest, Western Univ. of Hlth.Sci., Pomona, CA, [3]Moorehouse Med. Sch., Atlanta, GA, [4]The Jackson Lab. for Genomic Med., Farmington, CT

“We gratefully acknowledge the contribution the Genome Technologies Service at The Jackson Laboratory for expert assistance with the work described herein.” This study investigates transcriptional differences in the model organism, *Danio rerio* (zebrafish), associated with obesity. Zebrafish are an excellent model organism for obesity as they share 70% genetic identity with humans. Eighty-four percent of genes associated with human diseases are also found in zebrafish. Additionally, zebrafish have similar organ systems and store excess nutrients in their white adipocytes. The objective of this study is to use a zebrafish model to investigate early gene transcription changes associated with obesity. These changes may lead to downstream effects resulting in disease such as diabetes and non-alcoholic fatty liver disease. To create a model of obesity in zebrafish, two separate diets were used to raise obese fish and control fish for liver dissection. Control fish were fed a normal diet of 0.006 grams frozen artemia while obese fish were fed with 0.06 grams frozen artemia per day. Body Mass Indexes between obese and controls were analyzed using ANOVA (p value = 1.04e-11). Furthermore, histological analysis indicated that obese zebrafish had more and greater sized adipocytes which are a marker of increased adipose tissue. After 8 weeks fish were sacrificed, and livers dissected for DNA and RNA extraction. RNA sequencing was used to generate short read sequences and transcription differences were evaluated using the following bioinformatic tools: Trim Galore, STAR, Docker, and Stringtie. Preliminary analyses showed that genes overexpressed in obese fish included: preproinsulin, fatty acid binding proteins, amnionless, and cubilin. These genes along with others found to be overexpressed in obese fish are involved in increased production of insulin, increased receptors for insulin, increased fatty acid transport and increased inflammation. Some genes found to be under expressed in obese fish include pathogen binding proteins, MHC class I genes, and G-protein coupled receptors. Inactivating mutations in G-protein coupled receptors have been described to be associated with endocrinological related diseases. Overall, genes under expressed in obese fish are involved with a reduced immune system and reduced signaling pathways. This work has demonstrated that an obese zebrafish model can be used to investigate obesity related gene expression to uncover genes associated with human disease.

Session Title: Genetic Therapies Poster Session I

PB2031 In-vitro evaluation of CRISPR/Cas9-mediated excision of the Spinocerebellar Ataxia 27b repeat expansion in *FGF14*.

Authors:

G. Spurdens[1], A. Rebelo[1], M. Danzi[2], C. Yanick[1], D. Pellerin[3], B. Brais[3], M. Saporta[1], S. Zuchner[1]; [1]Univ. of Miami Miller Sch. of Med., Miami, FL, [2]Univ. of Miami, Dakota Dunes, SD, [3]McGill Univ., Montreal, QC, Canada

Spinocerebellar ataxia 27b (SCA27b) is a rare, autosomal dominant disorder caused by an expanded GAA repeat located deep in intron 1 of *FGF14*. SCA27b has been recently identified as one of the most common inherited late-onset forms of cerebellar ataxia, exhibiting episodic features, downbeat nystagmus, and cerebellar degeneration. In patients with repeat expansions exceeding 250 repeats, we see a significant decrease in the amount of FGF14 protein. FGF14 is highly expressed in the brain and is an important regulator of sodium channels and neuronal excitability. Due to locational restraints, neurodegenerative disorders have proven difficult to treat. However, with the continuous advancements in gene therapy, targeted approaches to treat neurodegenerative disorders are now a possibility. Here, we explore a corrective approach using CRISPR/Cas9 to excise the SCA27b repeat expansion. A dual-guide RNA (gRNA) strategy was used to precisely target the 5' and 3' regions flanking the repeats. Guide RNA validation and evaluation of cutting efficiency was performed in HEK293 Cas9-stable cells to screen for the best-performing gRNA pairings. Additional in-vitro testing was executed in patient fibroblasts and patient-derived iPSC motor neurons containing the repeat expansion. An adeno-associated virus (AAV) dual-vector approach was utilized to deliver all CRISPR/Cas9 components. Our results demonstrate efficient excision of the SCA27b repeat expansion in-vitro. The objective of our study is to successfully rescue FGF14 protein expression with minimal off-target effects to provide a promising therapeutic approach for SCA27b.

Session Title: Genetic Therapies Poster Session II

PB2032 JB2 therapy for a mouse MEF2C haploinsufficiency model of human autism in Mali

Authors:

M. kouyate; ACE-USTTB, Street 171/ Door 85, Mali

The MEF2C transcription factor regulates many genes related to autism spectrum disorders (ASD) and human MEF2C haploinsufficiency leads to ASD, intellectual disabilities, and epilepsy. However, the molecular mechanisms underlying the MEF2C haploinsufficiency syndrome remain poorly understood. We report here that Mef2c+/-(Mef2c-het) mice show behavioral deficits resembling those of human mice. behavioral deficits resembling those of human patients. Gene expression analyses of these mouse brains show changes in genes associated with neurogenesis, synapse formation and neuronal cell death. For example, Mef2c-het mice show decreased neurogenesis, increased neuronal apoptosis, and increased numbers of dead neurons. Neuronal apoptosis and an increased ratio of excitatory to inhibitory neurotransmission (E/I) Importantly, neurobehavioral deficits, E/I imbalance, and histological lesions are all ameliorated by treatment with JB2 . by treatment with JB2, a novel dual-acting compound related to memantine, JB2 regulates the phosphorylation of postsynaptic autism risk factors an FDA-approved drug. JB2 is an FDA-approved, non-competitive/rapidly deactivating antagonist of NMDA-type glutamate receptors. These results suggest that MEF2C haploinsufficiency leads to abnormal brain development, E/I imbalance, and neurobehavioral dysfunction, which can be mitigated by pharmacological intervention.

Session Title: Genetic Therapies Poster Session III

PB2033 Literature-based predictions of treatments for genetic disease pathology

Authors:

C. Deisseroth[1], W-S. Lee[1], J-Y. Kim[1], H-H. Jeong[1], J. Wang[2], H. Zoghbi[3], Z. Liu[1]; [1]Baylor Coll. of Med., Houston, TX, [2]Univ. of Pennsylvania, Philadelphia, PA, [3]Baylor Coll. of Med./HHMI/NRI, Houston, TX

Most debilitating genetic disorders caused by deficiency or excess of a specific protein are currently treated only at the symptomatic level—such as anti-epileptics for seizures and laxatives for constipation. Treatments that target the underlying molecular cause would provide greater therapeutic benefit, and reduce the risk of polypharmacy. Candidate drugs that might accomplish this goal must be tested individually, but a systematic analysis of molecular pathways is a promising method to prioritize these drugs for testing. Using known drug-gene and gene-gene interactions, one can predict drugs that are likely to modulate a protein of interest. These can treat haploinsufficiency conditions caused by loss-of-function genetic variants, as well as diseases caused by protein excess or toxic gain-of-function variants.

Existing pathway analyses rely on manually curated pathway databases to make predictions. With the exponentially growing body of medical literature, keeping these databases up to date with the latest discoveries is not sustainable. However, natural language processing (NLP) methods have made major advances in recent years, and enable us to extract this information automatically from the literature itself.

To find candidate drugs that correct protein imbalances without relying on manual pathway curation, we developed an artificial intelligence tool called PARsing ModifiErS via Abstract aNnotations (PARMESAN). This tool uses NLP to automatically identify gene-gene and drug-gene relationships from PubMed and PubMed Central. It then assembles a pathway database, and uses known relationships to predict undiscovered ones. It assigns a score to each hypothesis based on the amount of evidence supporting and opposing it.

We compare the drug-gene relationship predictions to the relationships displayed by the Drug-Gene Interaction Database (DGIdb), and the gene-gene predictions to the functional interactions in Reactome. PARMESAN's drug-gene predictions scoring above 7 (of which there are 264,980) are 23 times more likely to be supported by DGIdb than contradicted by it. Likewise, the gene-gene predictions scoring above 10, of which there are 325,379, are 21 times more likely to be confirmed than contradicted by Reactome.

PARMESAN is a free, open-source tool, and its source code. curated knowledgebase, and predictions will be made publicly available upon publication. This tool sustainably updates its knowledgebase, and allows researchers to prioritize tests for drugs that correct protein imbalances. This will save a tremendous amount of time and resources in drug screens, and pave the way to making many severe genetic disorders treatable.

Session Title: Genetic Therapies Poster Session I

PB2034 Monthly intravitreal enzyme replacement in classic late-infantile CLN2 disease slows decline in visual acuity: a case report.

Authors:

C. Courage[1], E. M. Maier[2,3], O. Ehrt[4], H. Pudritz[5], G. Rudolph[4], M. Gerhardt[4], C. Priglinger[4]; [1]Med. Genetics Ctr. (MGZ), Munich, Germany, [2]Dr. von Hauner Children's Hosp., Ludwig-Maximilians-University, Munich, Germany, [3]Labor Becker MVZ GbR, Newborn Screening Unit, Munich, Germany, [4]Dept. of Ophthalmology, Ludwig-Maximilians-Univ., Munich, Germany, [5]Pharmacy, Ludwig-Maximilians-Univ., Munich, Germany

Late infantile neuronal ceroid lipofuscinosis type 2 (CLN2) is a rare fatal neurodegenerative disorder characterized by deficiency of the lysosomal enzyme tripeptidyl peptidase-1 (TPP1). The autosomal-recessively inherited disorder is caused by pathogenic variants in the *TPP1* gene. The initial symptoms typically manifest between 2-4 years of age with epilepsy and speech delay, followed by rapid loss of motor and language function, vision loss, and early death. Intracerebroventricular enzyme replacement therapy (ERT) with recombinant human (rh) TPP1 (Cerliponase alfa©) has shown to slow the neurologic decline but has no effect on ongoing retinal degeneration in affected children. Intravitreal administration of rhTPP1 was shown to slow retinal degeneration in canine model of CLN2. A first-in-man use of intravitreal rhTPP1 has been started in the UK in 8 individuals in 2021 undergoing 8-weekly intravitreal injections of rhTPP1. We report on a 6-year-old male patient with classical CLN2 receiving monthly intravitreal rhTPP1 injections. The patient started ERT with 3 years and 4 months, being symptomatic with epilepsy. The patient was genetically confirmed with the known homozygous pathogenic variant c.509-1G>C in the *TPP1* gene. Intravitreal injections of rhTPP1 (0.2mg in 0.05ml BSA for the first 8 injections, following 0.4mg in 0.05ml) into the right eye (RE) have been started at the age of 5 years. The left eye (LE) was left untreated as a paired control. Until June 2023 the patient has undergone 10 intravitreal injections. At baseline best corrected visual acuity (BCVA) was 0.5 decimal in both eyes. Upon fundoscopy beginning bull´s eye maculopathy was evident. Optical coherence tomography (OCT) revealed perifoveal photoreceptor outer segment (OS) reduction with subfoveal sparing and diffuse thinning of the outer nuclear layer (ONL). The average of VA over the last 6 months was 0.3 in the treated RE and 0.2 in the control LE. In OCT volume scans a thin subfoveal OS band was still preserved in the RE. In the LE no volume scans were possible due to poor fixation. Line scans demonstrated perifoveal loss of outer nuclear layer and OS. No adverse reactions have been observed. Monthly general anaesthesia required for the intravitreal injections was well tolerated. In the observation period intravitreal ERT slightly delayed but did not stop retinal degeneration. BCVA in the treated eye remained better preserved during the first ten months when compared to the untreated eye, but slowly declined. Dose finding studies and clinical trials exceeding one year and are required to determine whether intravitreal ERT is able to significantly attenuate CLN2-associated retinopathy.

Session Title: Genetic Therapies Poster Session II

PB2035 Observational registry study of treatment practices and long-term outcomes of children with neurofibromatosis type 1 and plexiform neurofibromas initiating selumetinib in real-world practice in the United States: Study design and methodology

Authors:

R. de la Rosa Rodriguez[1], A. J. Waanders[2], K. Yohay[3], S. Fang[1], S. Wahlstrom[4], M. Bornhorst[5], J. Meade[6,7], K. Vallance[8]; [1]Alexion, AstraZeneca Rare Disease, Boston, MA, [2]Ann & Robert H. Lurie Children's Hosp. of Chicago, Chicago, IL, [3]Comprehensive Neurofibromatosis Ctr. at NYU Langone Hlth., New York, NY, [4]AstraZeneca, Wilmington, DE, [5]Children s Natl. Hosp., Washington, DC, [6]UPMC Children's Hosp. of Pittsburgh, Pittsburgh, PA, [7]Univ. of Pittsburgh Sch. of Med., Pittsburgh, PA, [8]Cook Children's Med. Ctr., Fort Worth, TX

Selumetinib (ARRY-142886, AZD6244) is an oral, allosteric MEK1/2 inhibitor approved in the United States (US) for children aged ≥2 years with neurofibromatosis type 1 (NF1) and symptomatic, inoperable plexiform neurofibromas (PN), based on results from the pivotal SPRINT study. Aims of the US Selumetinib Registry study [NCT05683678] include understanding treatment practices and assessing short- and long-term safety and effectiveness outcomes of selumetinib treatment in children with NF1-PN in real-world US practice. Clinical and non-clinical factors affecting outcomes will also be explored.This observational registry study of pediatric patients with NF1-PN will be conducted in up to 22 US centers. Eligible patients will be 2-18 years old at the time of selumetinib initiation (on/after April 10, 2020), and not currently participating in a clinical trial. Patients will be divided into three cohorts: Cohort 1 - treatment discontinued before enrollment, Cohort 2 - treatment initiated before enrollment and currently on treatment, Cohort 3 - treatment initiation intended within 3 months of enrollment. Patients will be followed for ≥36 and up to 60 months. Primary objectives are to describe patient demographics and disease characteristics (including diagnosis criteria and related manifestations, diagnostic tests and results, and PN-related morbidities), selumetinib treatment course, short- and long-term effectiveness and safety, and disease course and treatment following discontinuation. Key secondary objectives include measures of quality of life, pain and physical functioning before, during, and after selumetinib. Target enrollment is 200 patients with a 24-month enrollment period and will begin in 2023.The US Registry study will facilitate understanding of treatment practices and assess short and long-term outcomes of selumetinib for NF1-PN in a real-world setting.

Session Title: Genetic Therapies Poster Session III

PB2036 PEG-PLGA Polymersomes as a Delivery Mechanism for CRISPR-Cas9 in HEK293-GFP Cells.

Authors:

A. Harley[1], V. Kanduri[1], C. Argenti[1], D. LaVigne[1], C. Ennis[1], C. Forenzo[1], S. Edgecomb[2], D. Larsen[1]; [1]Clemson Univ., Clemson, SC, [2]Univ. of Mississippi, University, MS

The increasing prevalence of genetic diseases is a global concern. However, current treatments for many genetic diseases are limited to managing symptoms instead of correcting the cause. The gene-editing tool CRISPR-Cas9 (clustered regularly interspaced short palindromic repeats-associated protein 9) can be used to target specific DNA sequences and has shown potential as a possible therapeutic treatment for genetic diseases by enabling gene correction. The delivery of CRISPR-Cas9 into specific cells can be difficult to achieve due to immune system responses and specific biological barriers in place to prevent foreign materials from infiltration. To enable CRISPR-Cas9 delivery into cells, we investigate the use of self-assembling polymer-based nanoparticles (polymersomes) as an effective encapsulation and delivery method for the Cas9 ribonucleoprotein (RNP) to support non-viral gene editing. Film rehydration was used to synthesize polyethylene glycol and poly(lactic-co-glycolic acid) (PEG-PLGA) polymersomes, which were then optimized using sonication and filtration and characterized through dynamic light scattering (DLS). A TAT peptide was added to the polymersomes to provide a positive charge and improve endocytosis and endosomal escape. The human embryonic kidney cell line (HEK293-GFP) was selected for study because it contains a green fluorescence protein (GFP) which can be targeted for knockdown. First, sgRNA was designed to target GFP and made into an RNP that was tested via electroporation. Findings supported that the RNP was effective in GFP knockdown and was confirmed via fluorescence microscopy and flow cytometry. Treatment of HEK293-GFP cells with the RNP using sgRNA (CCATGGCACGGGCAGCTTGC) resulted in an average fluorescent knockdown of 64 ± 4.9% following a 72-hour incubation resulting in the selection of the RNP for encapsulation within polymersomes**.** The average size of polymersomes created via film rehydration was 84 nm ± 9.6 nm with an average PDI of 0.19 ± 0.03. Bicinchoninic acid (BCA) assay determined the average loaded content to be 5µg ± 1µg RNP for a total concentration of 2.5 µg/mg of Cas9 per polymer. Initial knockdown studies are ongoing using polymersomes, but preliminary fluorescent images are promising, demonstrating a decrease in GFP expression 72 hours after incubation. In conclusion, RNP was found to be effective in fluorescent knockdown. Future directions include in depth *in vitro* analysis in which encapsulated RNP is delivered to HEK293-GFP cells and then progressing to *in vivo* testing in Zebrafish.

Session Title: Genetic Therapies Poster Session I

PB2037 Quantitative characterization of on-target and off-target variation induced by CRISPR+Cas9 systems at the single-cell resolution.

Authors:

I. Salas-Gonzalez, V. Weigman; BioSkryb Genomics, Durham, NC

Introduction: Genome editing using CRISPR-Cas9 has revolutionized biomedical research by enabling the modification of DNA at targeted locations. The capacity of these systems to modify specific loci is dependent on the ability of Cas9 to induce DNA cleavage at genomic sites that exhibit precise base pairing with designed single guide RNAs (sgRNA). CRISPR/Cas systems are, conceptually, targeted editing tools, but the prevalence DNA edits that occur beyond the targeted gene (off-target effects) are increasingly observed. *In silico* tools have been developed to identify sgRNA guide-specific genome-wide potential off-target sites. Existing limitations of single cell genome enrichment have limited ability to confirm these in vitro, urging the need of unbiased methods of detection and validation. The goal of our study was to develop a bioinformatics workflow that could identify and prioritize on- and off-target gene editing events from single-cell whole genome sequencing data. Methods: We leveraged previously published data in which CD34+ cord blood cells and U2OS sarcoma cells were transfected with by two previously described sgRNAs (EMX1 and VEGFA) and Cas9. Mock treated and Cas9-only transfected cells served as a control. Single cells were isolated and primary template-directed amplification was used to amplify genomes from single cells. Libraries were created and deep sequencing (~20X coverage) was performed. Secondary bioinformatics analysis was performed using the WGS pipeline available in BaseJumper (BioSkryb Genomics). We called CNV events using Ginkgo, translocations were called using an ensemble approach overlapping events called by Manta and GRIDSS. *In silico* prediction of off-target events was done with Cas-OFFinder allowing up to 7 mismatches and 1 indel. Results: The EMX1 sgRNA induced more on-target mutations versus VEGFA sgRNA (5/5 vs 3/5 CD34+ cells and 5/5 vs 1/5 U2OS cells). Both guides exhibited heterogeneity in their on-target induced changes. Further, we characterized ~3200 reproducible off-target indels and 18 translocations from which only ~15% could be predicted in silico. Additionally, we identified ~150 mega bases across both cell types that exhibited significant fluctuation in their ploidy when exposed to the CRISPR+Guide treatments. Conclusions: We have developed a single-cell whole genome sequencing workflow for detecting indels, CNV, and SV alterations gene-edited cells. Overall, the data presented showcase the hidden heterogeneity of on-target modifications induced by CRISPR-Cas9 systems and the emergence of prevalent genome-wide off-target events impossible to determine with *in silico* approaches.

Session Title: Genetic Therapies Poster Session II

PB2038 † Rationale synthetic design of a broadly expressed NRF1-array mini-promoter for application to disease modeling and gene therapy.

Authors:

R. Nicholls[1], T. Zhang[1], J. Cheng-Zhang[2], W. Zhu[1], E. Hundertmark[2], M. Johnson[1], E. Koppes[1], D. Kemaladewi[3], G. Gittes[1]; [1]UPMC Children's Hosp, Pittsburgh, Pittsburgh, PA, [2]Univ. of Pittsburgh, Pittsburgh, PA, [3]Univ. of Pittsburgh Sch. of Med., Pittsburgh, PA

Adeno-associated virus (AAV) has strict size limitations for packing into a viral particle, which previously limited the size of encoded therapeutic proteins or genome editing applications. To overcome this size limitation of AAV vectors and to provide a modular component for gene regulation, rather than the commonly used CMV promoter-enhancer (613-nt) or many cell-type specific promoters for gene expression, we rationally designed mini-promoters of length 34-nt to 210-nt. These mini-promoters are tandem arrays of 2-10 units in length of motifs (consensus is 5'-YGCGCANGCGCR-3') for a transcription factor (TF), NRF1, with 10-11 nucleotides spacing equal to one-turn of the DNA α-helix between binding sites to allow each NRF1-dimer to interact for potential cooperative binding (while an additional set of the vectors includes a TATA-box motif). Indeed, use of NRF1-arrays of 1, 2, 4, 6, 8 and 10 units displayed cooperative increases in luciferase expression within transfected SK-N-SH neuroblastoma cells. Subsequent characterization focused on the 4xNRF1 and 10xNRF1 promoters, with high levels of mNeonGreen expression from mammalian expression plasmids in cell lines including INS-1 β-cells, αTC1-6 α-cells, C2C12 mouse myoblasts, HEI-193 schwannoma, and mHippoE-2 hippocampal neurons. Next, robust expression was also demonstrated by fluorescence microscopy and luciferase assays using each of these two promoters in an AAV6-luc-GFP vector following viral transduction of rodent β- and α-cell lines, a human α-cell line, and by ductal infusion of each virus in wildtype mice with *in vivo* imaging of bioluminescence clear expression was observed in the pancreas. Critically, the 4xNRF1 mini-promoter was also utilized in a miniaturized CRISPR-activation (CRISPRa) AAV9 vector for therapeutic *Lama1* gene upregulation in a *Lama2*-deficient muscular dystrophy mouse model, resulting in reduced hindlimb paralysis, improved mobility, body weight, and survival. Previously generated synthetic promoters are based on random DNA motifs or utilize endogenous promoter fragments, but do not have the broad expression profile of the NRF1-array mini-promoters. Miniaturization of promoter and encoded functions will aid *in vitro* and *in vivo* disease modeling and gene therapy AAV applications.

Session Title: Genetic Therapies Poster Session III

PB2039 Retroelement analyses identify diagnostic and therapeutic opportunities in genetic disorders

Authors:

B. Zhao[1,2,3], M. A. Nguyen[1,2], S. Woo[4], J. Kim[4], T. W. Yu[1,2,3], E. A. Lee[1,2,3]; [1]Boston Children's Hosp., Boston, MA, [2]Harvard Med. Sch., Boston, MA, [3]The Broad Inst. of MIT and Harvard, Cambridge, MA, [4]KAIST, Daejeon, Korea, Republic of

Advances in clinical sequencing have dramatically accelerated the discovery of causal variants in genetic diseases; however, a substantial proportion of genetic cases remain unresolved as certain classes of genetic variation still escape detection in conventional analysis. Retroelement insertion is one such class, which has been increasingly reported as a cause of Mendelian diseases and may offer unique therapeutic implications.

To understand the contribution of retroelements to genetic diseases, we analyzed retroelement insertions in whole genome sequencing (WGS) data from 237 individuals with ataxia telangiectasia (A-T), a representative recessive Mendelian disorder. We identified 15 patients bearing retroelement insertions in the causative *ATM* gene, all but one of which landed in noncoding regions, highlighting the advantages of WGS over exome sequencing. Notably, 13 (~5.5%) patients carried one of five distinct pathogenic insertions. We found one exonic Alu insertion, as well as three intronic Alus that landed in close proximity (<50 nt) to exon-intron boundaries; the latter events led to different levels of exon skipping as validated by RNA sequencing, RT-PCR, and/or minigene splicing assays. Beyond Alu insertions, we also resolved a deep intronic *DUSP16* pseudogene insertion, which resulted in loss of ATM function by activating cryptic splice sites.

The discovery of splice-altering insertions may represent therapeutic opportunities with splice-switching antisense oligonucleotide (ASO). Previous work has shown that aberrant splicing caused by an SVA insertion could be corrected using ASOs—short, chemically modified RNAs—tailored to the aberrant splicing defects (Kim *et al.*, 2019, *NEJM*). For the *DUSP16* insertion, we developed proof-of-concept ASOs that suppress cryptic exonization, supporting the experimental amenability of some of splice-altering insertions with RNA-based ASO intervention.

Leveraging a well-phenotyped cohort of patients with A-T as a model, we provide a first systematic estimate of the contribution of retroelements to the genetic architecture of recessive Mendelian disorders as 5.5% (13/237 cases). Our study underscores the importance of retroelement insertions as an underexplored source of pathogenic genetic variation and therapeutic opportunities.

Session Title: Genetic Therapies Poster Session I

PB2040 Safety evaluation for CRISPR/Cas9-based PD-1 gene-editing products

Authors:

H. Kim; Natl. Inst. of Food and Drug Safety Evaluation, Cheongju-si, Korea, Republic of

CRISPR/Cas9 system is a versatile and specific gene-editing technique for applying to treat genetic diseases as well as intractable diseases. Despite the remarkable advances in CRISPR/Cas9 over the past decade, the assessment methods for CRISPR/Cas9-based human gene-editing products remain to be verified from regulatory perspectives. Here, we assessed the efficacy and safety of CRISPR/Cas9-based programmed cell death 1 (PD-1) gene-editing product by current 'on-target' and 'off-target' analysis methods. NGS-based analysis of insertion and deletion (Indel) showed that the Indel events occurred in 91% of transfected HEK293T cells. From the results of ELISA assay, the expression level of PD-1 protein was decreased by 44% in PD-1-targeted CRISPR/Cas-9 transfected Molt-4 cells, demonstrating that the CRISPR/Cas9 system disrupted the PD-1 gene in Molt-4 cells. Mycoplasma was not detected in PD-1 gene-edited Molt-4 cells. Next, the off-target effects were examined by using *in silico* prediction tool (Cas-OFFinder) and *in vitro* method (Digenome-seq). To perform Digenome-seq, we produced synthetic gRNA and Cas9-RNP complex *in vitro* followed by an assessment of DNA digestion efficiency using qPCR. The potential off-target sites of these Cas9-RNP were investigated by Whole Genome Sequencing. The results from targeted sequencing shows that the Indel events were hardly detected (<0.1%) among the highly ranked potential off-target sites obtained by Cas-OFFinder and Digenome-seq analysis. In this study, as a safety concern, we assessed off-target effects in CRISPR/Cas9-based PD-1 gene-editing products using *in silico* and *in vitro* off-target methods with comparative analysis. To increase the predictability of off-target effect, we suggest that integrated analysis would be necessary with orthogonal approaches using currently available and appropriate off-target methods depending on each CRISPR/Cas9-based gene-editing product.

Session Title: Genetic Therapies Poster Session II

PB2041 Safety of a reduced infusion time of Agalsidase Beta in the treatment of Fabry Disease: An Observational Study

Authors:

D. GERMAIN[1,2], A. EL MOUDEN[1], W. TRANG[1]; [1]Div. of Med. Genetics, APHP, GARCHES, France, [2]Div. of Med. Genetics, Univ. of Versailles, Montigny, France

Background: Fabry disease is an X-linked inborn error of metabolism caused by the deficiency of the lysosomal alpha-galactosidase A. Enzyme Replacement Therapy (ERT) with agalsidase beta was approved in 2001 and has demonstrated clinical benefits. Very few studies have explored the optimal duration of agalsidase beta infusions and many patients over the world are still infused during 3 hours every other week, in a lifetime way, significantly impairing their quality of life. In this work, the safety and tolerability of high speed agalsidase beta infusions were investigated.Patients and Methods: The safety of an optimized infusion time of 90 minutes (specified in the Summary of Product Characteristics) for the administration of 70 mg of agalsidase beta (diluted in 250 mL of NaCl 0.9%) was studied in 29 consecutive adult patients (15 males, 14 females) with Fabry disease (23 different *GLA* pathogenic variants) for a total of 53 infusions. The study population included 8 patients naïve to the enzyme and 17/29 patients with significant comorbidities. Physical examination was performed and vital signs were measured before and after each infusion. Results: Agalsidase beta was infused without premedication a constant rate of 0,77 mg/min (46,7mg/h) in all 53 infusions. None of the 29 patients complained of any discomfort. No adverse effect was reported. Vital signs remained stable in all patients during all infusions. Three patients received 105 mg of agalsidase in 90 minutes (1,16 mg/min - 70mg/h) with no side effects. Discussion: The infusion of 70 mg of agalsidase beta in 90 minutes (0,77mg/min) was well-tolerated with no report of discomfort or adverse effect. The optimized regimen reduces the burden of biweekly infusions, and improves patients quality of life in patients with Fabry disease.

Session Title: Genetic Therapies Poster Session III

PB2042 Short peptides targeting MECP2 Rett syndrome mutations through phage display

Authors:

J. Vincent[1], S. Pastore[2], T. I. Sheikh[1], J. Ausio[3], T. T. Madanagopal[1]; [1]Ctr Add/Mental Hlth, Clarke Div, Toronto, ON, Canada, [2]Ctr. for Addiction & Mental Hlth., Toronto, ON, Canada, [3]Univ. of Victoria, Victoria, BC, Canada

Background Mutations in MECP2 (methyl CpG binding protein 2) have been found to cause Rett syndrome (RTT). About 33% of all RTT patients have missense mutations, and mainly in the methyl-CpG-binding domain (MBD) that binds to DNA and forms a chromatin-associated protein complex. These mutations interfere with DNA binding and/or disrupt the clustering of chromatin. A similar proportion of RTT patients have nonsense mutations that truncate the MECP2 protein, leading to protein degradation resulting in insufficient levels in the cell. Studies have suggested this rapid degradation/insufficiency might also be true for missense mutations. Stabilizing the mutant protein could be a key to reverse the symptoms caused by insufficient levels of MECP2 and in-turn RTT. **Methods** T158M and R111G mutations were introduced to a MECP2/MBD vector construct by site directed mutagenesis with a GST tag. Wild type and mutant MECP2 vectors were over expressed in BL21 cells, and the resulting protein was used as the bait for phage display. After four rounds of selection, clones were sequenced by next generation sequencing, and peptides selected for MECP2 mutant binding over WT, or GST. Following this selection, a reverse approach using GST-tagged selected peptide as bait was used to pull down the MECP2-GFP protein using GST pull-down kit, as validation of the peptide-MECP2 binding. The peptides selected from this step were then synthesized with a Rhodamine B-tag. These peptides were then transfected into C2C12 cells using d-Arg8 (cell penetrating peptide). Only the peptides that co-localized with MECP2 were used for further analysis. The selected peptides will be used for cellular studies, to explore ability to bind MECP2 and stabilize and/or increase the bioavailability of mutant MECP2 protein. **Results** From the phage display experiment 11 peptides for wild type MECP2, 13 for T158M mutant and 13 for R111G mutant were selected. These peptides were then used for the reverse prey/bait GST pull down, and only 6 peptides for wild type and 5 peptides for T158M mutant were confirmed for binding to MECP2. No peptides were confirmed for the R111G mutant. These 11 peptides were then co-transfected in to C2C12 cells along with MECP2-GFP. The peptides that co-localized with MECP2 will be further analyzed for MECP2 protein stability and bioavailability with ELISA, cycloheximide chase assay and FRAP. **Conclusion** The selection of peptides that binds and stabilize MECP2 with specific RTT mutations through this approach may help develop personalized therapeutic approaches for RTT, and the approach may also be extended to other genes and diseases.

Session Title: Genetic Therapies Poster Session I

PB2043 Single-cell RNA Sequencing Data-based Drug Repurposing for Alzheimer's Disease Using Gene Expression data of ROSMAP Cohort

Authors:

T. Wang, C. Wang, D. Wu; Univ. of North Carolina at Chapel Hill, Chapel Hill, NC

Alzheimer's disease (AD) and related dementias currently affect 50 million individuals worldwide. Only 6 drugs, including four symptom relief agents and two disease-modifying agents, have been approved for AD. Drug repurposing (DR) offers a cost-effective and readily available strategy to address the need for new AD treatments. In this study, we employed the ASGARD (A Single-cell Guided Pipeline to Aid Repurposing of Drugs) algorithm to define a novel predictive drug score that considers all cell types and evaluates drug efficacy across multiple cell types. We applied ASGARD to publicly available single-cell RNA sequencing (scRNA-seq) data from the Religious Orders Study and Memory and Aging Project (ROSMAP). The ROSMAP is a large longitudinal clinical-pathologic cohort study of aging and AD, and the dataset includes snRNA-seq data obtained from post-mortem brain samples of 48 individuals, including 24 participants with AD and 24 participants without AD as controls. The analyzed tissue originated from the Brodmann area 10 prefrontal cortex. From the AD cases (26,049 cells) and controls (44,281 cells), we collected scRNA-seq data of 27,737 genes from brain tissues, analyzing 6 out of the 8 main cell types: excitatory neurons, inhibitory neurons, astrocytes, oligodendrocytes, microglia, and oligodendrocyte progenitor cells. Pericytes and endocytes data were excluded due to lower cell counts. Differential gene expression analysis using the Wilcoxon rank-sum test in Seurat identified differentially expressed genes between cases and controls. The ASGARD computed a drug score for each drug, quantifying the reverse of expression levels in the L1000 drug response dataset (GSE92742, GSE70138). Applying ASGARD, we predicted 6 drugs (FDR < 0.1 and overall drug score > 0.75 quantiles) to be potentially effective for AD, including dasatinib, vorinostat, sirolimus, dopamine, dexamethasone, and testosterone, ranked by decreasing drug score. Dasatinib, currently approved for chronic myeloid leukemia, is being investigated in phase I and II trials as a DR candidate for AD (SToMP-AD, STAMINA). Vorinostat, the second-best candidate, is an approved drug for T-cell lymphoma and is undergoing a phase I trial as a DR candidate for AD (VostatAD01). Sirolimus, approved for preventing kidney transplant rejection, is being evaluated in a phase II clinical trial as a DR candidate for AD (REACH). The predicted drug candidates will undergo validation through pharmacoepidemiologic analyses using real-world data (US Medicare claims) to assess their effectiveness in preventing incident dementia and slowing the progression of AD.

Session Title: Genetic Therapies Poster Session II

PB2044 Single-cell RNA sequencing reveals novel cellular factors for response to immunosuppressive therapy in aplastic anemia

Authors:

J. Jang[1], H. Kim[1], S-S. Park[2], M. Kim[3], Y. Min[3], H-o. Jeong[1], S. Kim[1], T. Hwang[1], D. W. Choi[1], H-J. Kim[2], S. Song[4], S. Lee[1], C. Lee[3], J. Lee[2]; [1]Ulsan Natl. Inst. of Sci. and Technology, Ulsan, Korea, Republic of, [2]Seoul St. Mary's Hosp., Seoul, Korea, Republic of, [3]Korea Res. Inst. of Chemical Technology, Daejeon, Korea, Republic of, [4]Chungnam Natl. Univ. Sch. of Med., Daejeon, Korea, Republic of

Aplastic anemia (AA) is a lethal hematological disorder; however, its pathogenesis is not fully understood. Although immunosuppressive therapy (IST) is a major treatment option for AA, one-third of patients do not respond to IST and its resistance mechanism remains elusive. To understand AA pathogenesis and IST resistance, we performed single-cell RNA sequencing (scRNA-seq) of bone marrow (BM) from healthy controls and patients with AA at diagnosis. We found that $CD34^+$ early-stage erythroid precursor cells (EPCs) and $PROM1^+$ hematopoietic stem cells (HSCs) were significantly depleted in AA, which suggests that the depletion of $CD34^+$ early-stage EPCs and $PROM1^+$ HSCs might be one of the major mechanisms for AA pathogenesis related with BM cell hypoplasia. More importantly, we observed the significant enrichment of $CD8^+$ T cells and T cells-activating intercellular interactions in IST responders, indicating the association between the expansion and activation of T cells and the positive response of IST in AA. Taken together, our findings represent a valuable resource offering novel insights into the cellular heterogeneity in the BM of AA and reveal potential biomarkers for IST, building the foundation for future precision therapies in AA.

Session Title: Genetic Therapies Poster Session III

PB2045 Small molecule inhibition of *Plasmodium AP2-I*&*BDP1* genes in *P. berghei*-infected mice

Authors:

D. O. Oladejo[1,2], D. O. Anzaku[1,2], C. O. Mamudu[1,2], T. M. Dokunmu[1,2], E. E. J. Iweala[1,2]; [1]Covenant Univ., Ota, Nigeria, [2]Covenant Applied Informatics and Communication Africa Ctr. of Excellence, Ota, Nigeria

Plasmodium transcription factors (TFs), along with other proteins in a complex, control gene expression during the intraerythrocytic development cycle of the parasites. We predicted *in silico*, the binding of a small molecule 7-[(7-methoxy-4,5-dihydro-1H-benzo[g]indazol-3-yl)carbonyl]-2-phenyl-5,6,7,8-tetrahydropyrazolo[1,5-a]pyrido[4,3-d]pyrimidin-9(1H)-one (MCL) to the active sites of *Pf*AP2-I and *Pf*BDP1 thereby inhibiting their actions. This study validated experimentally the inhibitory effects of MCL against *Pb*AP2-I and *Pb*BDP1 *in vivo* after due ethical approval. Forty mice were inoculated with 0.2 mL infected blood suspension having 1.0×10^7 parasitized erythrocytes intraperitoneally. Three days post-infection, mice were treated with 1, 10, 20, 30, 40, and 50 µM/kg body weight of MCL compound, respectively, and 10.00mg/kg chloroquine (control) for 7 days. Blood samples were obtained from the treated mice for analysis. RNA was extracted, and reverse-transcription PCR of AP2-I and BDP1 genes was carried out. Data were compared using ANOVA at $P<0.05$, and the relative gene expression of AP2-I and BDP1 genes was expressed as $2^{-\Delta\Delta Ct}$ of the untreated and treated relative to GAPDH (control). Relative gene expression results after a 7-day single oral dose shows MCL inhibited the expression of *Pf*AP2-I (by 4.52 and 8.08 folds) and *Pf*BDP1 (by 37.92 and 79.07 folds) at 40 and 50µM doses. Our findings validate that MCL inhibits *Pf*AP2-I and *Pf*BDP1 significantly by reducing its expression in treated mice, hence representing potential new targets of antimalarial therapy and for control of malaria parasite invasion.

Session Title: Genetic Therapies Poster Session I

PB2046 Suppression of dominant-negative *SPTAN1* allele in patient-derived cell lines via antisense oligonucleotide therapy

Authors:

C. Wang, S. Li, S. Zhao, J. Sinson, N. T. Ho, B. Vuocolo, G. Clark, S. Nagamani, P. Liu; Baylor Coll. of Med., Houston, TX

Spectrins are a family of cytoskeletal proteins that support positioning and stabilization of ion channels, cell adhesion molecules, and membrane proteins. Pathogenic variants in four out of six spectrin genes in humans have been discovered to cause neurodevelopmental disorders. Pathogenic variants in *SPTAN1* encoding αIIspectrin cause developmental and epileptic encephalopathy 5 (DEE5). SPTAN1, like other spectrins, comprises alpha and beta heterodimers that form tetramers in a head-to-head arrangement. Heterozygous pathogenic variants in the heterodimerization domain of *SPTAN1* (e.g., p.D2303_L2305dup) are suspected to cause dominant-negative aggregation of the thermolabile heterodimers, leading to more severe clinical phenotypes. Previous work has demonstrated accumulation of αII-spectrin aggregates in glutamatergic neurons derived from induced pluripotent stem cells (iPSCs) generated from patient-derived fibroblasts with the p.D2303_L2305dup variant. Targeting such dominant-negative variants from a therapeutic standpoint is made challenging by the presence of in-tandem spectrin repeats in the heterodimerization domain. We hypothesized that using a specific strategy with anti-sense oligonucleotides (ASO) to induce RNase-mediated cleavage of mRNA from the pathogenic allele can be of potential therapeutic value. We have obtained samples from three individuals, two of whom are monozygotic twins, with the p.D2303_L2305dup pathogenic variant. Here, we present a research grade ASO design strategy that successfully targets the p.D2303_L2305dup variant of *SPTAN1* in iPSCs and iPSC-derived neural progenitor cells and neurons generated from the three individuals with this particular variant. Expression analysis using qPCR revealed that the ASO led to a significant knockdown of mutant *SPTAN1* expression in iPSCs. Ongoing work is aimed at exploring off-target effects, as well as confirming reduction in the aggregation of thermolabile heterodimers in neurons and neural progenitor cells using fluorescent microscopy. Our approach would be an important first step in designing ASO and performing validations for DEE5 in patient-derived cell lines, and if successful, could help in the development and validation of therapeutic strategies for other neurodevelopmental disorders.

Session Title: Genetic Therapies Poster Session II

PB2047 Suppressors of TAZ-mediated cellular dysfunction identified through genetic interaction screening

Authors:

S. Masud[1], D. Thomson Taylor[1], N. Mikolajewicz[1], J. Wei[2], P. Mero[2], K. Chan[2], K. Brown[2], A. Habsid[2], O. Sizova[2], A. Quaile[3], J. Chan[3], J. Montenegro Burke[3], C. Boone[3], J. Moffat[1]; [1]Univ. of Toronto, Hosp. for Sick Children, Toronto, ON, Canada, [2]Hosp. for Sick Children, Toronto, ON, Canada, [3]Univ. of Toronto, Toronto, ON, Canada

The genotype-to-phenotype relationship in health and disease is complex, influenced by environmental and genetic factors. Unmasking the non-trivial pairwise genetic interactions on a genome-wide scale remains an active area of investigation in many prominent disease models. The diverse spectrum of phenotypes observed in Barth Syndrome (BTHS), caused by mutations in the gene *TAZ* is consistent with such complex genotype-to-phenotype interplay, however, systematic efforts to understand these interactions remain lacking. Furthermore, numerous mutations in the *TAZ* gene have been documented in BTHS patients, including frameshift, point mutations, as well as mutations that disrupt alternative splicing of *TAZ*. This diverse pattern of *TAZ* mutations suggest that several independent mechanisms contribute to a loss or altered function of Tafazzin, the protein product of *TAZ*, which may account for the pleiotropic nature of BTHS. We address this by leveraging our expertise in genome-wide CRISPR screens, mapping the genetic interactions between mutations in *TAZ* and all other genes in a systematic manner using a HAP1 cellular model. Our work has identified suppressing interactions with *TAZ* loss-of-function, restoring the critical lipid species cardiolipin, and rescuing *TAZ*-mediated mitochondrial dysfunction. Using several cell lines and animal models, we integrate functional genomics, proteomics, and the development of novel biologics that can lead to clinical/commercially relevant outcomes with the potential to greatly impact BTHS and expand on biological pathways critical in physiological and pathological mitochondrial conditions.

Session Title: Genetic Therapies Poster Session III

PB2048 Transforming a pathogenic dominant-negative variant into a benign loss-of-function variant: *ATAD3A* as a paradigm

Authors:

S. Ezer[1,2], S. Yanovsky-Dagan[1], L. Omari[1,2], A. Saada-Reisch[1], T. Harel[1]; [1]Hadassah Med. Ctr., Jerusalem, Israel, [2]Hebrew Univ. of Jerusalem, Jerusalem, Israel

Introduction *ATAD3A* is a nuclear gene that encodes a protein with various roles in the mitochondria. Among the pathogenic variants reported in this gene is a recurrent dominant-negative variant NM_001170535.3 (*ATAD3A*): c.1582C>T (p.Arg528Trp). Information available from families who carry deletion variants in *ATAD3A* confirms that carriers of one deletion with one wild-type allele are healthy. Thus, we assume that nullifying the dominant-negative allele using allele-specific targeting approach with antisense oligonucleotides (ASOs) could provide a therapeutic option. **Methods** We designed ten ASOs spanning the pathogenic variant and transfected these into patient-derived fibroblasts carrying the monoallelic variant. ASO specificity and efficiency was evaluated by several methods: RNA was sequenced followed by Synthego ICE analysis for measuring the relative change in expression of the mutant transcript triggered by each ASO; reverse transcriptase quantitative PCR was performed to measure the effect of ASO on reduction in total *ATAD3A* expression and C>T allele expression, and to test for normalization of the expression of selected genes we have previously shown to be affected by the dominant negative variant (e.g., *MFN2*). Mitochondria function was assessed by reactive oxygen species (ROS) production and cytochrome c oxidase (COX) activity. **Results** Several ASOs were found to improve the c.1582 wild-type/mutant allele ratio at the expression level, and accordingly improved mitochondrial function i.e. COX/citrate synthase ratio, normalize gene expression and decrease ROS production compared to untreated fibroblasts. At least one ASO consistently seemed to be harmful to the variant-carrying cells. Finally, ASOs that reduce the C>T allele drastically are also less specific and cause damage by also reducing the wild-type allele. **Conclusion** Allele-specific ASOs can be used to induce a shift in the relative expression of mutant/wild-type alleles, while improving mitochondrial function in cells with the *ATAD3A* recurrent pathogenic variant. While none of the ASOs studied resulted in a complete knockdown of the C>T allele, it is possible that improvement can be achieved by shifting the stoichiometry towards the wild-type allele. Non-specific targeting of the wild-type allele is harmful, even if there is a strong shift towards knockdown of the other allele. Further research in neuronal cells and animal models may provide further validation of this allele-specific therapeutic approach in monoallelic *ATAD3A*-associated disease. Moreover, this approach is generalizable to other disease genes with both dominant and recessive inheritance patterns.

Session Title: Genetic Therapies Poster Session I

PB2049 Understanding the role of Curcuma caesia Roxb. in *PIN1* degradation using SKOV3 ovarian cells.

Authors:

T. Patel[1], K. Mukunthan[2], P. Bhattacharya[3], N. Hempel[4]; [1]VIT Univ., Vellore, India, [2]Manipal Inst. of Technology, Manipal Academy of Higher Ed., Manipal, India, [3]Scientist, Neuberg Ctr. for Genomic Med., Neuberg-Supratech, Ahmedabad, India, [4]Faculty Member, Cancer Biology Program, UPMC Hillman Cancer Ctr., Pittsburgh, PA

Peptidyl-prolyl isomerase (*PIN1*) is over-expressed in many cancers and is known to contribute to the expansion of cancer stem cells. All trans retinoic acid (ATRA) induced down-regulation of *PIN1* reduces cancer cell proliferation especially using the slow releasing formulation of ATRA. Hence, it is employed in the treatment of ovarian cancer, bladder cancers, Kaposi sarcoma, squamous cell carcinoma, and neuroblastoma. However, its usage is limited by a very short half-life and so search for newer ATRA-like compounds is ongoing. C. caesia, also known as Black Turmeric (BT) is a bitter spice predominantly found in India and other south-east Asian countries. Our study was aimed to explore the protective role of Curcuma caesia Roxb. against cancer cell proliferation and *PIN1* over-expression. Computationally, we show that compounds from the hexane rhizome extract (HRE) of C. caesia have drug-like properties especially, Compound 3 that docks to PIN1 protein with optimum affinity and can be a novel PIN1 inhibitor. We carried out *in vitro* studies using SKOV3 ovarian cancer cells to compare the efficacy of ATRA and hexane extract. The MTS assay determined an IC_{50} value of 2.09 µg/mL and 12.77 µg/mL respectively for ATRA and HRE in SKOV3 cells. Clonogenic assay showed that hexane extract reduced the survival fraction of SKOV3 cells as compared to ATRA albeit not significantly. Furthermore, a significant increase in HRE induced cellular apoptosis was established as compared to ATRA with Annexin V assay. The relative mRNA expression for *PIN1* was significantly ($p<0.001$) reduced with HRE and was comparable to ATRA which also down-regulated *PIN1* drastically. In contrast, Western blot analysis showed a significant increase in PIN1 degradation using hexane extract as compared to ATRA treatment in SKOV3 cells. The mechanisms which may lead to PIN1 degradation by HRE needs to be elucidated in future studies. Our pilot study thus demonstrates that hexane extract from C. caesia may harbor molecules that can target cancer cells and may contain potential PIN1 inhibitors.

Session Title: Genetic Therapies Poster Session II

PB2050 † Weight loss in individuals with 16p11.2 deletion syndrome receiving GLP-1 receptor agonists and/or SGLT-2 inhibitors for type 2 diabetes treatment.

Authors:

H. Shimelis[1], C. Wood[1], L. K. Walsh[1], B. Finucane[1], S. M. Myers[1], J. E. Franceschelli[1], S. D. Christopher[1], M. T. Oetjens[1], L. H. David[2], C. L. Martin[1]; [1]Geisinger, Danville, PA, [2]Univ. of Florida, Jacksonville, Jacksonville, FL

Obesity is a common feature affecting 75-95% of adults with 16p11.2 deletion syndrome (16p11.2DS), a condition caused by a recurrent proximal deletion involving breakpoints 4-5. GLP-1 receptor agonists (GLP1-RA) and SGLT-2 inhibitors (SGLT-2-I) have shown benefit for weight management in patients with or without type 2 diabetes (T2D), however, there has been limited research on the effectiveness of these medications in adults with 16p11.2DS. This study compares weight loss and glycemic control among Geisinger MyCode participants with 16p11.2DS versus matched variant-negative controls who received GLP1-RA, SGLT-2-I, or co-treatment for T2D. Of 160 adults with 16p11.2DS, 19 (12%) were prescribed GLP1-RA (N=6), SGLT-2-I (N=6), or GLP1-RA and SGLT-2-I (N=7) and were compared with matched variant-negative controls (N=149). One-year percent total weight loss (TWL) and hemoglobin A1c (HbA1c) were assessed with linear mixed model. At baseline, the 16p11.2DS and control cohorts were similar for mean age (±SD): 51.9±10.3 years / 53.3±9.2 years; body mass index (44.3/42.9), gender, and T2D diagnosis (100%/100%), respectively. After one year, TWL among those treated with GLP1-RA alone was not significantly different between 16p11.2DS and controls, with each group experiencing ~2% weight loss. By contrast, SGLT-2-I monotreatment was far less effective for weight reduction in adults with 16p11.2DS compared to controls (-0.96% vs 3.23%), a mean difference of 4.19% (p=0.047). Co-treatment with both drugs did not significantly boost the effectiveness of GLP1-RA or SGLT-2-I monotherapy in controls. However, in the 16p11.2DS cohort, the combined use of both drugs resulted in significantly greater weight reduction than SGLT-2-I alone (6.46% vs. -0.93%), with a mean TWL difference of -7.43% (p=0.023); and GLP1-RA alone (6.46% vs. 2.08%), although the mean difference between co-treatment and GLP-RA only trended toward significance (p=0.15). Furthermore, the use of SGLT-2-I in the control group resulted in a significant HbA1c reduction (-1.08%), while the effects of GLP-1RA or co-treatment (-0.86%, -0.42%) were not significant. By contrast, in those with 16p11.2DS, monotherapy with SGLT-2-I or GLP-1RA did not show significant HbA1c reduction (-0.71%, -0.70%), however the effect of co-treatment (-1.82%) trended toward significance (p=0.09). Our findings suggest that adults with 16p11.2DS may derive significantly more benefit from co-treatment with GLP1-RA and SGLT-2-I than from SGLT-2 alone. Additional research is needed to confirm our findings, which may inform medication strategies for obesity treatment in adults with 16p11.2DS.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Session I

PB2051 A comprehensive genetic analysis unravels the impact of *PVRL1* in the risk of nonsyndromic cleft lip with or without cleft palate across ethnically diverse populations.

Authors:

J. Park[1], G. Kang[2], S-H. Baek[3], Y. Kim[4], D-H. Kim[5]; [1]Dept. of Med. Genetics, Hallym Univ. Coll. of Med., Chuncheon, Korea, Republic of, [2]Hallym Univ. Coll. of Med., Chuncheon, Korea, Republic of, [3]Dept. of Orthodontics, Sch. of Dentistry, Seoul Natl. Univ., Seoul, Korea, Republic of, [4]Dept. of Orthodontics, The Inst. of Oral Hlth.Sci., Samsung Med. Ctr. , Sungkyunkwan Univ. Sch. of Med., Seoul, Korea, Republic of, [5]Dept. of Social and Preventive Med., Hallym Univ. Coll. of Med., Chuncheon, Korea, Republic of

Nonsyndromic cleft lip with or without palate (NSCL/P) is a prevalent congenital anomaly, affecting approximately 135,000 infants worldwide each year. However, inconsistent replication of genetic effects, including the poliovirus receptor-related 1 (*PVRL1*) gene, in previous studies, suggests the presence of genetic heterogeneity across different ethnic groups. To address this, we initially performed a transmission disequilibrium test and evaluated parent-of-origin effects using 10 tagging single nucleotide polymorphisms (tSNPs) in 142 Korean families with NSCL/P. To validate our findings, we conducted replication analyses using publicly available data from 245 Asian families and 487 European families (dbGaP accession: phs000774.v2.p1). Additionally, case-control analyses were performed within each ethnic group. Our results demonstrated robust evidence of transmission disequilibrium at rs7940667 (allelic $p = 0.007$), which was consistently replicated in both Asian ($p = 2.96\times10^{-7}$) and European families ($p = 1.82\times10^{-11}$). Furthermore, the initial observation of marginal evidence for maternal over-transmission of the A allele at rs10790330 (PO-LRT $p = 0.044$) in Korean families was exclusively replicated in European families (PO-LRT $p = 0.0009$). Notably, all ten SNPs displayed significant transmission disequilibrium in European families, whereas five SNPs exhibited this effect in Asian families. In terms of parent-of-origin effects, distinct patterns of maternal over-transmission were observed for different SNPs, specifically rs7129848 in Asian and rs10790330, rs3935406, and rs10892434 in European families. Our findings provide robust evidence of transmission disequilibrium and highlight the impact of maternal over-transmission within the *PVRL1* gene on NSCL/P risk. The replication of these effects in both Asian and European families, along with larger family sizes, strengthens the statistical associations and supports the presence of a shared genetic basis across different ethnic groups. These results contribute to our understanding of the role of the *PVRL1* gene in NSCL/P risk and shed light on genetic heterogeneity underlying this condition.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Session II

PB2052 A longitudinal study on the maternal blood transcriptome during the early phase of onset of labor

Authors:

T. Leung, S-C. CHIM, R. Wang, W. Ding, K-W. WONG, E-L. KO; The Chinese Univ. of Hong Kong, Hong Kong, Hong Kong

BACKGROUND: The mechanism that instigates human parturition remains incompletely understood. Hence we investigated the changes in transcriptome of maternal blood before and after onset of labor.METHODS: Healthy singleton-pregnant women who were indicated for induction of labor (IOL) were invited to participate this longitudinal study. The maternal whole blood samples collected just before IOL and after latent phase of labor onset (cervical dilation <3 cm) were subjected to RNA-seq and RT-qPCR validation. To gain insights into molecular functions of the differentially expressed genes, gene functions from the Gene Ontology (GO) Consortium were analyzed.RESULTS: In the RNA-seq dataset, 913 mRNA transcripts were changed by >2-fold after the onset of labor, compared with before onset. The fold-change values showed the same direction of change between the RNA-seq and validation datasets. For instance, the maternal blood levels of CLEC4E and LIG4 mRNAs were increased after labor onset in both datasets. Women with high (>95th percentile) blood levels of CLEC4E mRNA underwent labor earlier (4.8 h vs 30 h, Logrank p =0.04) and delivered sooner (13 h vs 44 h, p =0.01) than their counterparts with low levels. Women with high (>median) blood levels of LIG4 mRNA underwent labor earlier (4.3 h vs 7.3 h, p =0.02) and delivered sooner (11 vs 14 h, p =0.02) than their counterparts with low levels. In the gene list with increased maternal blood RNA levels after labor onset, among other GO terms, regulation of neutrophil degranulation (7.4- fold), leukocyte aggregation (5.4-fold), production of molecular mediators involved in inflammatory response (5.3-fold) were enriched. In contrast, in the list of genes with decreased blood RNA levels after labor onset, no GO terms related to the immune system was found.Conclusion: During the early phase of labor onset, the maternal blood transcriptome is drastically changed, characterized with the upregulation of many genes in the immune system. The roles and potential of these genes in the prediction of a successful IOL (i.e. ending in spontaneous labor) warrants further investigation in a cross-sectional study along with women who failed IOL (i.e. ending in Cesarean delivery).

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Session III

PB2053 A molecular autopsy of fetuses with cerebral abnormalities and description of PLAT as a new putative hydrocephalus gene

Authors:

K. Uguen, M-A. Delrue, E. Lemyre, C. Maftei, A. Karalis, A-M. Laberge, J-F. Soucy, S. Perreault, C. Fallet-Bianco, J. Michaud; CHU Sainte-Justine, Montreal, QC, Canada

Purpose. Central nervous system (CNS) anomalies represent an important group of birth defects, occurring in approximately 0.3% of newborns. We hypothesized that fetal CNS anomalies may involve a distinct genetic landscape than that observed in children with homologous anomalies, as the former group includes a subset of cases that are so severely affected that they cannot survive beyond the perinatal period. In order to characterize the genes associated with these anomalies, we investigated fetuses with CNS anomalies from interrupted pregnancies by performing whole-exome sequencing (WES).
Methods. We recruited 85 fetuses with structural cerebral anomalies. They had negative array-CGH and no history of environmental factors that would explain their condition. All fetuses underwent autopsy with deep neuropathological examination. We also included cases with fetal akinesia as this disorder can be caused by severe CNS impairment. We performed WES in 62 of these trios while the sequencing of the remaining cases (n=23) is underway.
Results. WES revealed a genetic etiology in 31% of fetuses. The diagnostic yield was higher in fetuses with anomalies of the corpus callosum (6/14, 43%) and cortical defects (2/4, 50%), but lower in those with cerebellar hypoplasia (3/8, 38%) and isolated ventriculomegaly (3/10, 33%). Pathogenic variants were de novo in 63% and inherited in 37% cases. Interestingly, we describe the occurrence of a severe form of obstructive tetraventriculomegaly in two siblings from a consanguineous union. WES revealed the presence of a homozygous stopgain variation in the PLAT gene (NM_000930.5:c.85C>T, p.(Arg29Ter)), which encodes the tissue plasminogen activator. This finding and the previous description of hydranencephaly in two siblings with another homozygous truncating variant (NM_000930.5:c.102_103del)1 in PLAT strongly support the involvement of this gene in severe ventriculomegaly.
Conclusion. Our study highlights the diagnostic utility of WES in fetuses with cerebral anomalies. Moreover, the description of a second family with a homozygous truncating variant in PLAT emphasizes the potential role of this gene in severe ventriculomegaly, .
1. Shamseldin HE et al, Hum Genet. 2016;135(10):1209-1211.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Session I

PB2054 A naturally occurring variant of *MBD4* causes maternal germline hypermutation in primates.

Authors:

A. Stendahl[1], R. Sanghvi[2], S. Peterson[1], K. Ray[1], A. Lima[1], R. Rahbari[2], D. Conrad[1]; [1]Oregon Hlth.& Sci. Univ., Beaverton, OR, [2]Wellcome Trust Sanger Inst., Cambridge, United Kingdom

As part of an ongoing genome sequencing project at the Oregon National Primate Research Center, we identified a rhesus macaque with a rare homozygous frameshift mutation in the gene Methyl-CpG binding domain 4 (*MBD4*). MBD4 is responsible for the repair of C>T deamination mutations at CpG dinucleotides and has been linked to somatic hypermutation and cancer predisposition in humans. We show here that MBD4-associated hypermutation also affects the germline: the 6 offspring of the *MBD4*-null dam have a 4-6-fold increase in *de novo* mutation burden. This excess burden was predominantly C>T mutations at CpG dinucleotides consistent with *MBD4* loss-of-function in the dam. There was also a significant excess of C>T at CpA sites, indicating an important, unappreciated role for MBD4 to repair deamination in CpA contexts. The *MBD4*-null dam developed sustained eosinophilia later in life, but we saw no other signs of neoplastic processes associated with *MBD4* loss-of-function in humans, nor any obvious disease in the hypermutated offspring. This work provides the first evidence for a genetic factor causing hypermutation in the maternal germline of a mammal, and adds to the very small list of naturally occurring variants known to modulate germline mutation rates in mammals.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Session II

PB2055 A novel basic carrier test for all ethnicities using long read sequencing and an artificial intelligence platform.

Authors:

C. Kyriakidis[1], M. Garrido Navas[2], G. Madjarov[3], L. Krych[4], A. Schack[5], K. Alapont Celaya[1], M. Chaushevska[1], Z. Velkoski[1], D. Galevski[6], A. Nikov[6], G. Nowicki[7]; [1]gMendel ApS, Copenhagen, Denmark, [2]Genyo Ctr., Alcalá la Real, Jaén, Spain, [3]Univ. Ss Cyril & Methodius, Skopje, Macedonia, The Former Yugoslav Republic of, [4]Univeristy of Copenhagen, Copenhagen, Denmark, [5]Univ. of Copenhagen, Copenhagen, Denmark, [6]Netcetera AG, Zurich, Switzerland, [7]GenXone S.A., Poznan, Poland

The prevalence of genetic disorders in the global population is expected to continue increasing. Through development of technologies for early detection, including for carriers of genetic disorders, advancement can be made in efforts to contribute to better disease management. In this presentation, we will focus on Basic Carrier Test, which, as recommended by the American College Obstetricians Gynecologists (ACOG) and the American College of Medical Genetics (ACMG) identifies adult carriers of all ethnicities of genetic disorders Fragile X, Spinal Muscular Atrophy (SMA), and Cystic Fibrosis (CF). We developed an Artificial Intelligence (AI) platform for real-time and automated simultaneous analysis and detection of genetic diseases. It enables extensive numerical and visualization analysis and accurate, systematic and timely diagnosis significantly improving disease management. This platform can be applied directly on the stream of base-called DNA reads generated by the Nanopore device. It exceeds the limits of the real-time monitoring and analysis per DNA sample, which can significantly reduce the overall cost. The calculated throughput of the analysis pipeline is 4120 reads per sec measured on a referent hardware architecture using thread parallelism of 10. This platform managed to correctly identify CF, SMA and Fragile X. Following testing and validation on a mix of real and synthetically prepared samples, the sensitivity and specificity exceeded 90% in patients from Denmark and Spain. This work is a groundbreaking proof of concept that the nanopore low-cost sequencing platform can serve as a certified diagnostic tool with AI-based disease classification for mass-screening significantly reducing time and cost of future diagnostics for any genetic disorder. At the same time, this work opens an opportunity to substitute older diagnostic tools with more accurate, faster and cost-effective technology, and therefore will contribute to improvement of life for people with genetic disorders by improving the current practices in healthcare.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Session III

PB2056 *Abca2* depletion associated with gastrulation and neurological defects in *Xenopus tropicalis.*

Authors:

F. Aslam[1,2], E. Mis[1], A. Acharya[1], W. Hwang[1], S. Naz[2], M. Khokha[1], S. Lakhani[1]; [1]Yale Univ., New Haven, CT, [2]Univ. of the Punjab, Lahore, Pakistan

ABCA2 is a member of membrane associated transport proteins known to transport molecules across extracellular and intracellular membranes. *ABCA2* encoded protein, highly expressed in the brain tissue, maintains the homoeostasis of lipids, especially sphingolipids, and sterols. *Abca2* knockout mice have neurological defects due to aberrant myelination. We identified recessively inherited frameshift variants of *ABCA2* in individuals of consanguineous families with neurological disease. Patients presented with developmental delay, intellectual disabilities, ataxia and seizures. In order to understand the molecular mechanism of ABCA2, we depleted *Abca2* in *Xenopus tropicalis* (*X. tropicalis*) by targeting the gene with CRISPR Cas9 using three nonoverlapping guide RNAs. Depletion of *Abca2* caused severe gastrulation defects. Embryos that survived past gastrulation had motility problems in later embryonic stages. We assayed motility in *X. tropicalis Abca2* depleted tadpoles by stimulating stage 45 tadpoles near the head and measuring the distance traveled. Spontaneous seizures were also observed in the form of a 'C' shaped curve in the tails of embryonic stage 45^{+} tadpoles. Human wild-type *ABCA2* overexpression rescued the gastrulation, motility and seizure phenotypes in *Abca2* depleted embryos but patient variants failed to do so, verifying the detrimental nature of the variants. Hence, *X. tropicalis* knockout of *Abca2* recapitulated the human phenotypes and also demonstrated an important role of the gene in early development. This work was supported by NIH/NICHD (R01hd102186) and IRSIP funding to FA from HEC, Pakistan.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Session I

PB2057 Aberrant landscapes of maternal meiotic crossovers contribute to aneuploidies in human embryos

Authors:

R. McCoy[1], D. Ariad[2], S. Madjunkova[3], M. Madjunkov[3], S. Chen[3], R. Abramov[3], C. Librach[3]; [1]Johns Hopkins Univ., Baltimore, MD, [2]Johns Hopkins Univ., Balitmore, MD, [3]CReATe Fertility Ctr., Toronto, ON, Canada

Meiotic recombination is a fundamental source of human genetic diversity and is also critical for ensuring the accuracy of chromosome segregation. Understanding the landscape of meiotic recombination, its variation across individuals, and the processes by which it goes awry are long-standing goals in human genetics. Current approaches for inferring the landscape of recombination either rely on population genetic patterns of linkage disequilibrium (LD)—capturing a time-averaged view—or direct detection of crossovers in gametes or multi-generation pedigrees, limiting the scale and availability of relevant datasets. Here, we introduce an approach for inferring sex-specific landscapes of recombination from retrospective analysis of data from preimplantation genetic testing for aneuploidy (PGT-A) based on low-coverage (<0.05×) whole-genome sequencing of biopsies from *in vitro* fertilized (IVF) embryos. To overcome the sparsity of these data, our method exploits its inherent relatedness structure, knowledge of haplotypes from external population reference panels, as well as the frequent occurrence of chromosome loss in embryos, whereby the remaining chromosome is phased by default. Based on simulation, we show that our method retains high accuracy down to coverages as low as 0.02×. Performance is affected by the local density of SNPs and sequencing reads, as well as ancestry matching between the reference panel and the target samples. Applying this method to low-coverage PGT-A data from 18,967 embryos, we mapped 70,660 recombination events at an average resolution of ~150 kbp. The resulting sex-specific maps of crossovers generated by our method were strongly correlated with published sex-specific genetic maps based on prospective whole-genome sequencing of living parent-offspring trios, in support of the accuracy of our method. However, we observed that the total length of the female genetic map is reduced for trisomies compared to disomies, while the genomic distribution of crossovers is also altered in a chromosome-specific manner. These observations include a depletion of crossovers near the centromere for trisomies compared to disomies of chromosome 16. Based on haplotype configurations in regions surrounding the centromeres, our data additionally indicate that individual chromosomes possess unique propensities for different mechanisms of meiotic error ($\chi 2[2, N = 2540] = 489.2$, $p = 1.6 \times 10^{-77}$). Together, our results provide a detailed view of the role of aberrant meiotic recombination in the origins of human aneuploidies, as well as a flexible tool for mapping crossovers in low-coverage sequencing data from multiple siblings.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Session II

PB2058 † Application of single-cell multiomics technology to establish an effective screening method for identifying fetal nucleated red blood cells suitable for non-invasive prenatal genetic testing.

Authors:

N. Ito[1], T. Fujii[1], K. Taniguchi[1,2], K. Hata[1,2], H. Sago[1], K. Nakabayashi[1]; [1]Natl. Ctr. for Child Hlth.and Dev., Tokyo, Japan, [2]Gunma Univ. Graduate Sch. of Med., Gunma, Japan

Non-invasive prenatal genetic testing (NIPT) based on cell-free DNA (cfDNA) in maternal plasma has prevailed worldwide. However, low levels of the fetal DNA fraction may lead to false-negative results. In contrast, since fetal cells in maternal blood represent a pure source of fetal genomic DNA, they are expected to be ideal for NIPT if successfully isolated. We have previously established a protocol to efficiently isolate fetal nucleated red blood cells (fNRBCs) from maternal blood using fluorescence-activated cell sorting with erythrocyte-associated surface antigen markers (Ito N et al. 2021). However, the efficiency of the whole genome amplification in the single isolated cells was not consistent enough for subsequent genomic diagnostics. The quality of genomic DNA may have deteriorated due to enucleation and/or apoptosis in most isolated fNRBCs. In this study, we tested whether multiomics technology effectively overcame this problem. We adopted a single-cell multiomics technique, G&T-seq , which can obtain parallel sequencing data of single-cell genomes and transcriptomes. Using umbilical cord blood as a model, we isolated 12 single lymphocytes and 69 single NRBC candidate cells from two samples and obtained their genomic and transcriptomic data. As expected, we could distinguish lymphocyte and NRBC candidate groups based on gene expression patterns. When plotted using multi-dimensional scaling (MDS), three NRBC-candidate cells were clustered separately and located between the lymphocyte cluster and the other 66 NRBC-candidate clusters along the dimension-1 axis. Based on the expression levels of the marker genes for the primitive, intermediate, and mature stages of NRBC maturation, three NRBC-candidate cells from the separate group were classified as those in the primitive stage and the other 66 NRBC-candidate cells as those in the mature stage. We subsequently assessed the quality of the genomic sequencing data based on their library yield and mapping rate for all cells. Two out of three cells in the primitive stage showed high library yields and mapping rates comparable with those of lymphocyte cells, suggesting the intactness of their nuclear genome when the two cells were isolated. Therefore, G&T-seq effectively identified NRBC cells in the primitive stage among NRBC candidates. Applying single-cell multiomics technology is expected to facilitate the development of fNRBC-based NIPT.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Session III

PB2059 Bioinformatic identification of small X chromosome CNVs in 46, XY pregnancy loss cases suggest novel male lethal candidate genes.

Authors:

J. Cao[1], K. Hatchell[2], S. Aguilar[2], M. Aminbeidokhti[3], C. Tise[4], A. Rajkovic[3,5,6], J. Bernstein[7], S. Yatsenko[8,9,10,11], M. Snyder[1], S. Aradhya[2,12], R. B. Lathi[13]; [1]Stanford Univ. Sch. of Med. Dept. of Genetics, Stanford, CA, [2]Invitae Corp., San Francisco, CA, [3]Univ. of California San Francisco Dept. of Pathology, San Francisco, CA, [4]Lucile Packard Children's Hosp. and Stanford Univ. Dept. of Pediatrics, Stanford, CA, [5]Univ. of California San Francisco Inst. of Human Genetics, San Francisco, CA, [6]Univ. of California San Francisco Dept. of Obstetrics, San Francisco, CA, [7]Stanford Univ. Sch. of Med. Dept. of Pediatrics, Stanford, CA, [8]Univ. of Pittsburgh Dept. of Pathology, Pittsburgh, PA, [9]Magee-Womens Res. Inst., Pittsburgh, PA, [10]Univ. of Pittsburgh Dept. of Obstetrics, Pittsburgh, PA, [11]Univ. of Pittsburgh Dept. of Human Genetics, Pittsburgh, PA, [12]Stanford Univ. Dept. of Pathology, Stanford, CA, [13]Stanford Univ. Dept. of Obstetrics and Gynecology, Stanford, CA

Background: Human reproduction is inefficient, with ~70% of all conceptions not progressing to live birth. While large chromosome abnormalities are commonly seen in pregnancy losses, nearly half of pregnancy losses remain unexplained by standard of care chromosome analysis. An overrepresentation of karyotypically normal 46, XY conceptions among pregnancy losses suggests many X-linked genes may be essential for development and survival of male fetuses. Currently, only ~30% of X-linked genes have been linked to specific genetic conditions, while the clinical importance of the remaining 70% is undetermined. The objective of this study was to identify copy number variations (CNVs) on the X chromosome to assess possible correlations with male pregnancy loss and identify novel genes and their functions that may be essential for embryonic development. **Methods:** We analyzed anonymized high-resolution chromosomal microarray data and reported test results from 12,506 pregnancy loss samples, referred for genetic testing at a single clinical laboratory. We examined the clinically reported chromosomal analysis results and CNVs for each sample. We then performed blinded de novo bioinformatic analysis, independent of clinical laboratory processes, to enrich for and characterize smaller CNVs of unknown significance on the X chromosome of male samples. **Results:** Among the 12,506 pregnancy loss CMA results we examined, 7,428 cases were euploid, including 4,214 (56.7%) males and 3,214 (43.3%) females. Clinically reportable CNVs were present in 1,269 (30.1%) of XY euploid conceptions, including 75 X chromosome CNVs, consistent with results reported to referring physicians. We further identified ~400 samples with CNVs on the X chromosome of uncertain clinical significance, including many <1Mb CNVs that were not reported clinically. Within these abnormal regions, we identified CNVs involving several genes (*NKRF, SEPTIN6, SRPK3, IDH3G, BRCC3*) which have not been reported in live born males and currently have no known OMIM phenotype or disease association. **Conclusion:** In this analysis of >7,000 46,XY pregnancy loss cases, we demonstrate an enrichment of male losses and numerous incidences of small X chromosome CNVs that are not typically sought out or reported during clinical testing. These findings provide critical and new information about specific genes and point to genomic regions on the X chromosome that may be lethal when copy number is disrupted. This work and similar studies of this magnitude have the potential to identify novel gene alterations associated with human lethality, and improve the understanding of embryonic and fetal development.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Session I

PB2060 Can first trimester prenatal samples be used to functionally assess Noonan syndrome variants?

Authors:

N. Zjacic[1], G. Siegel[1], I. Ivanovski[1], A. Bahr[1], M. Mocera[1], P. Krejci[1], B. Oneda[1], T. Burkhardt[2], A. Rauch[1]; [1]Inst. of Med. Genetics, Univ. of Zurich, Schlieren, Switzerland, [2]Univ. Hosp. Zurich, Univ. of Zurich, Zurich, Switzerland

Noonan syndrome (NS) is a congenital disorder that affects 1 in 1000-2500 births and is most commonly caused by gain-of-function variants in genes of the essential Ras-mitogen-activated protein kinase (Ras/MAPK) pathway. This results in pleiotropic developmental problems of varying severity such as cardiac defects and neurocognitive deficits, as well as increased predisposition to childhood malignancies. After exclusion of chromosomal aberrations, NS and other RASopathies are the most common differential diagnosis in fetuses with increased nuchal translucency detected by first trimester screening. There are currently no routine functional tests available for parents expecting a child with a variant of unknown significance (VUS) in a RASopathy gene that would allow them to make informed decisions after genetic screening or prepare for post-pregnancy. Our goal was therefore to create a direct, patient cell-based analysis pipeline that would provide functional information on VUS in NS genes in a short enough time frame for critical pregnancy-related decisions. We tested Ras/MAPK pathway functional readouts from established overexpression studies in HEK293 and Cos-1/7 cells, as well as a novel approach, for direct functional assessment of cultured chorionic villi sample (CVS) cells obtained during invasive prenatal sampling. Five patient CVS lines containing one unclear and four (likely) pathogenic variants in the most common NS disease genes, *PTPN11* and *SOS1*, were compared to five healthy control lines. We found no significant differences between patient and control activation of ERK1/2 via Western blot or activation of RAS via ELISA analysis of whole cell lysates. This contrasts the published HEK293 and Cos-1/7 overexpression studies which showed increased and prolonged activation of pathway components. In addition to the established readouts for individual components of the Ras/MAPK pathway, we tested a commercially available serine/threonine kinase assay to compare the activity of 144 kinases from multiple pathways in patient and control cells. The assay did not produce robust enough results to select biomarkers that distinguish between patients and controls. Overall, our data show that cultured CVS cells from invasive first trimester prenatal testing may not suitably reflect the increased Ras/MAPK pathway activity underlying NS when analyzed using conventional functional assays from the literature or a novel kinase assay approach. A different cell type or an entirely different strategy will therefore be required for creating a functional pipeline for VUS in NS genes.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Session II

PB2061 Cell-type specific methylome-wide significant associations for adverse neonatal outcomes detected in blood.

Authors:

T. Campbell[1], L. Xie[1], C. Hultman[2], E. Van den Oord[1], K. Aberg[1]; [1]Virginia Commonwealth Univ., Richmond, VA, [2]Karolinska Inst., Stockholm, Sweden

Adverse neonatal outcomes such as small birth size, early gestational age, low APGAR score, and neonatal jaundice have been associated to an increased risk of physical and mental health conditions. A potential mechanism by which neonatal outcomes may lead to these conditions, is via epigenetic variation, such as DNA methylation (DNAm). DNAm is often cell-type specific, requiring the need to control for cell-type composition when studying bulk tissue. However, even with such adjustments, standard bulk analyses represent an average value across the cell-types present, often preventing differences in specific cell-types to be detected. Here we use sequencing-based MBD-seq data, from nearly 28 million CpGs, from 332 neonatal blood samples. Next, we use statistical deconvolution to performed cell-type specific methylome-wide association studies (MWAS) of neonatal outcomes. In addition to investigating 9 individual neonatal outcomes, we used an exploratory factor analysis that identified two factors, which represents shared effects mainly from size related (size factor) and disease related (disease factor) outcomes. We also regressed out these factors from the outcomes allowing us to study outcome specific effects. We identified methylome-wide significant ($q<0.1$) associations for the disease and size factors and for APGAR. The significant findings for the size factor were detected for B cells (N=5), granulocytes (N=10), monocytes (N=12), cytotoxic T cells (N=27) and bulk (N=13). Of note, the top findings in granulocytes ($P=3.35x10^{-9}$, $q=0.0694$) and monocytes ($P=1.12x10^{-11}$, $q=0.0003$) were the same CpG, but with different direction of effect, located in the *BLNK* gene, highlighting the importance of cell-type specific analysis. For the disease factor, findings included monocytes (N=2), cytotoxic T cells (N=28), T helper cells (N=216) and bulk (N=2). Here, we detected the same three significant CpGs in cytotoxic T and T helper cells ($P<6.10x10^{-8}$, $q<0.0922$) that overlapped *SLC12A7,* which has been identified as a biomarker for childhood cancers. For APGAR, findings were detected for granulocytes (N=1), natural killer cells (N=1) and bulk (N=10). This included five CpGs ($P<1.08x10^{-8}$, $q<0.0374$) overlapping *ASRGL1,* which has been identified as a biomarker for spontaneous preterm birth. To our knowledge, this is the first time that cell-type specific methylation patterns have been studied in relation to neonatal outcomes. Our results suggest that most outcomes are linked to two distinct processes (i.e., from different factors) which provide valuable insights into the cell-type specific role DNAm plays for these neonatal outcomes.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Session III

PB2062 Clinical significance of cell free DNA screening positive trisomy 15: Cohort study and literature review.

Authors:

R. Choy[1], S. Xue[1], D. Lam[2], S. Cheung[2], Y. Zhao[1], Y. Zheng[1], C. Chung[1], Y. Kwok[1], X. Zhu[3], T. Leung[1], X. Kong[3]; [1]Chinese Univ. of Hong Kong, Hong Kong, China, [2]NIPS Dept., NGS lab, Xcelom Limited, Hong Kong, China, [3]Genetics and Prenatal Diagnosis Ctr., Dept. of Obstetrics and Gynaecology, The First Affiliated Hosp. of Zhengzhou Univ., Zhengzhou, China

Cell free DNA (cfDNA) testing has expaned to include rare autosomal trisomies. Trisomy 15 is one of the most common imprinted chromosome aneuploidies identified in cfDNA test and fetuses with trisomy 15 positive result may be at risk not only of fetal trisomy but also of fetal uniparental disomy (UPD) 15 responsible for Angelman/Prader-Willi syndrome. Currently, the positive predictive value (PPV), risk of fetal UPD 15 and the clinical significance of trisomy 15 positive cases are largely unknown. This study aims to investigate the clinical significance of cfDNA positive for trisomy 15. We prospectively studied the performance of cfDNA for testing trisomy 15 in 36,466 pregnant women in a single center. The incidence, PPV, and risk of UPD 15 were assessed based on amniocentesis results by cytogenetic, molecular analysis and reported clinical outcomes. Eleven of 36,466 cases were reported positive for trisomy 15 in our cohort, yielding a cfDNA screen-positive rate of 0.030% (11/36,466). Invasive diagnostic testing results revealed the PPV for trisomy 15 was 36.4% (4/11) and 3 of 9 (33.3%) cases were confirmed with maternal UPD 15. We further conducted a systematic review of cfDNA cohort studies that have reported the detection of trisomy 15 to generate a large cohort. The incidence, diagnostic accuracy as well as pregnancy outcomes of trisomy 15 positive cases were studied. We generated a large cfDNA cohort of 826,601 pregnancies and identified 130 cases (0.016%) that were screened positive for trisomy 15. In total, 18 of 75 cases with diagnostic results were confirmed in the fetuses, including 8 full trisomy 15 and 10 true fetal mosaicism, with an overall PPV of 21.6% (18/75, 95% CI, 3.9-44.9). Among 46 cases with UPD testing, 13 (28.3%) were identified to be maternal UPD 15, and 8 of them had normal fetal karyotypes. 70.0% (49/70) of the patients experienced adverse pregnancy outcomes, including 32.9% fetal loss, 24.3% selective termination of pregnancy, and 12.9% with other pregnancy complications. The overall PPV of cfDNA positive trisomy 15 is 21.6%, and fetuses with cfDNA-positive trisomy 15 results showed a high risk (28.3%) of Prader-Willi syndrome resulting from maternal UPD 15. Our data suggests that amniocentesis should be recommended for cfDNA positive trisomy 15 case and exclusion of both UPD and true fetal mosaicism are highly warranted.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Session I

PB2063 Convergent coexpression of autism-associated genes implicates early immune dysfunction

Authors:

F. Segato Dezem[1], C. Cheng[2], J. Plummer[3], K. Aldinger[4]; [1]St Jude Children s Res. Hosp., Los Angeles, CA, [2]Boston Univ. Sch. of Med., Boston, MA, [3]St Jude Children s Res. Hosp., Memphis, TN, [4]UW/Seattle Children s Res. Inst., Seattle, WA

Autism spectrum disorder (ASD) is a highly variable neurodevelopmental disorder diagnosed based on a core set of social communication impairments and repetitive sensory-motor behaviors. Individuals with ASD often also exhibit a wide range of clinical features that vary in severity and parallel the heterogeneity among the common and rare genetic variants associated with ASD etiology. Despite this clinical and genetic complexity, pathway enrichment among ASD-implicated genes has suggested some functional commonalities, including impairments in synapse function and chromatin modification. However, most prior studies focused on the potential impact of genetic variants on brain development even though the genetic changes occur in all cells of the body and their impact is expected to contribute to both brain and non-brain clinical phenotypes. Here, we examined the expression of high-confidence ASD-associated genes (345 curated from SFARI high-confidence gene list) across multiple organs during human development to more comprehensively investigate the tissues and cell types that are vulnerable to ASD pathophysiology. Paired enrichment analysis revealed ASD genes enriched in the thymus, cerebellum, cerebrum, eye, and kidney ($P<0.05$). We employed an empirical binary classifier (ROCit) of the ASD gene set against single-cell gene expression profiles from the Descartes Atlas of human gene expression during development which profiled 4 million cells from 15 prenatal tissues. Partitioning cell type identities revealed a set of ASD genes (SCAI, MBD5, CASK) with enriched expression in the thymocytes. Immune dysfunction has been implicated in ASD etiology and the thymus is a critical site for T cell development. Our results suggest there may be an early primary role for immune-mediated dysfunction in ASD pathogenesis. Understanding ASD genetics in the context of body-wide single-cell gene expression during development provides a framework for understanding which genetic perturbations could lead to the complex clinical features recognized among individuals diagnosed with ASD.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Session II

PB2064 Correlation between maternal copy number variations and preterm birth in Indian women: A pilot study

Authors:

D. Charan[1], D. Chandel[2], T. Patel[3]; [1]Vellore Inst. of technology, Vellore, Vellore, India, [2]Sch. of Sci.s, Gujarat Univ., Ahmedabad, India, [3]VIT Univ., Vellore, India

Preterm birth (PTB) is a condition of spontaneous parturition before the completion of 37 weeks of gestation. PTB causes multiple birth complications, lifelong disabilities and can be frequently fatal to the neonates. India has the highest number of preterm births second only from African countries. Multiple factors like maternal malnutrition, inflammation/infection, endocrinological or genetic causes could be attributed to PTB. In this study we performed certain biochemical and molecular analysis for three patients [Institutional Ethical Committee of - GU (GUJIEC_03_2017)/GMERS-(GMERSMCS/IEC/37/2018).The preliminary survey showed they belong to the lower economy strata indicating maternal malnutrition as one of the reasons. Besides this, the ROS biochemical panel showed an increased oxidative stress, along with increased micronuclei. These outcomes can be a result of inadequate nourishment, unhygienic living environment and/or genotoxic exposures. Pan-genome copy number variation study (Illumina) in these 3 maternal samples showed that one of the samples had only three CNVs (Sample 2) while the other two sample have 46 and 83 CNVs (Sample 1/3) respectively. Few genes showed loss of heterozygosity while some showed increase in the copy number across all the samples studied. STRING analysis did not give us any common genes among the three samples establishing that more sample size is required to conclude the involvement of specific genes in PTB. We conducted in silico gene enrichment studies. Most of the genes for sample 1 were clustered under infection, inflammation, physiologic conditions like hypertension and endocrinological disturbances. Sample 2 with not many variations in gene copy number, showed genes clustered only around viral infection. Sample 3 had genes clustered under infections and a few other genes under non-heritable genetic conditions. More sample size is required to conclude the outcomes of preterm births. However, the primary data suggest that malnourishment and unhygienic living environment which leads to infections could be one of the early causes of PTB. This could be elevated by fragile genetic makeup and variations in copy number due to molecular events in genes involved in cellular, biologic, and molecular functions.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Session III

PB2065 Detection of fetal RhD status on SNP-based prenatal cell-free DNA screening

Authors:

T. Wang[1], Y. Altug[1], S-C. Kuo[1], X. Wang[1], N. Avent[2], S. Vege[3], H. Pawar[1], R. Swenerton[1], G. Heilek[1], B. Zimmermann[1], J. Babiarz[1]; [1]Natera, Inc., San Carlos, CA, [2]Consultant, Bristol, United Kingdom, [3]New York Blood Ctr., Long Island City, NY

Background: Rhesus D (RhD) alloimmunization can occur if a RhD negative (RhD(-)) individual is exposed to RhD positive (RhD(+)) blood. In the US, antenatal prophylaxis with anti-RhD immunoglobulin at 28 weeks gestation and after a potential sensitizing event is standard of care for RhD(-) pregnant people unless the father of the baby is also Rh(-) or the pregnant person is already RhD alloimmunized. In some countries, quantitative PCR (qPCR)-based NIPT for fetal RhD status in D(-) patients is used to avoid unnecessary treatment with anti-RhD immunoglobulin and to guide pregnancy management in RhD alloimmunized patients. Here, we establish the proof of concept of testing fetal RhD as a part of an SNP-based NIPT (Panorama^{TM}) that utilizes massively multiplex PCR (mmPCR) to screen for common chromosomal aneuploidies and microdeletions.
Methods: A set of primers for targeted amplicons covering regions differing between the *RHD* and *RHCE* genes were designed and incorporated into the NIPT assay. *RHD* and *RHCE* can be directly detected by targeting the differences between these two highly similar genes. To assess sensitivity of the approach, we developed models from cell lines with known RhD status to mimic RhD(-) pregnancies, carrying either an RhD(+) or RhD(-) fetus over a range of fetal fractions from 1% to 15%. Additionally, the method was tested on 996 de-identified pregnant patient plasma samples using qPCR for concordance analysis.
Results: Cell line mixture experiments that modeled RhD(-) pregnancies, including minor RhD(-) genotypes, showed the mmPCR approach was able to determine fetal RhD status. These experiments also demonstrated accurate detection of RhD status across a range of fetal fractions, even below the current 2.8% threshold for aneuploidy calling. Among 996 pregnant patient samples, all 110 RhD(-) were correctly identified, with 40 RhD(-) fetuses and 70 RhD(+) fetuses. These results were confirmed with qPCR. This yielded a Negative Percent Agreement (NPA) of 100% [95% CI: 91.2-100%] and a Positive Percent Agreement (PPA) of 100% [95% CI: 94.9-100%]. Importantly, the addition of RhD targets did not interfere with overall NIPT performance.
Conclusions: mmPCR and sequencing allows for the direct observation and quantification of fetal *RHD* alleles, enabling the accurate prediction of fetal RhD status, even at low fetal fraction. Our results established the proof of concept for adding RhD detection to the Panorama^{TM} NIPT assay to screen for fetal Rh status as early as 9 weeks gestation. Information about fetal RhD status has potential to be incorporated into the management of RhD(-) pregnancies.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Session I

PB2066 Diagnosing rare conditions in low-resource settings: the experience of the Differences of Sex Development clinic in the Democratic Republic of the Congo

Authors:

K. Mosema-Be-Amoti[1,2,3,4,5], D. Spencer[6,7], F. Kabala-Feruzi[8], J. Ebhele-Angundru[9], R. KABANGA[10], T. Mubenga-Lubenga[11], Y. Musawu[12], A. Mayani[13], L. Kiamfu-Musimbi[14], J. Magene[15], S. Kpabo-Nyomi[16], L. Muzama-Luko[17], A. Kilembe-Manzanza[18], K. Kamate[19], D. Diessa-Nyongombe[20], J-B. Essimbo-Djoma[21], D. Mutoba[22], D. Tshala-Katumbay[1], D. Mumba Ngoyi[1], E. Vilain[7], E. C. Delot[6]; [1]Inst. Natl. de Recherche BioMed.e, Kinshasa, Congo, Democratic Republic of the, [2]Biamba Marie Mutombo Hosp., Kinshasa, Congo, Democratic Republic of the, [3]Université de l'Uélé, Isiro, Congo, Democratic Republic of the, [4]Université de Kisangani, Kisangani, Congo, Democratic Republic of the, [5]Programme Natl. de Santé de l'Adolescent (PNSA)/Ctr. des Jeunes de Matete, Kinshasa, Congo, Democratic Republic of the, [6]Children's Natl. Res. and Innovation Campus, Washington, DC, [7]Univ. of California, Irvine, Irvine, CA, [8]Ctr. Provincial de Transfusion Sanguine, Kindu, Congo, Democratic Republic of the, [9]Mercy Surgical Polyclinic, Bunia, Congo, Democratic Republic of the, [10]Gen. Hosp. of Bunkonde, Bunkonde, Congo, Democratic Republic of the, [11]Ctr. Med. Operationnel Type A, Kinshasa, Congo, Democratic Republic of the, [12]Programme Natl. des Urgences et Actions Humanitaires, Kinshasa, Congo, Democratic Republic of the, [13]CSR Kimbulu, Kimbulu, Congo, Democratic Republic of the, [14]Gen. Hosp. of Makala, Kinshasa, Congo, Democratic Republic of the, [15]Hosp. de Makeke, Makeke, Congo, Democratic Republic of the, [16]Programme Natl. de Sante Mentale, Ituri, Congo, Democratic Republic of the, [17]Sendwe Hosp., Lubumbashi, Congo, Democratic Republic of the, [18]Clinique Ngaliema, Kinshasa, Congo, Democratic Republic of the, [19]Service de Chirugie Pédiatrique, Cliniques Univ.ires de Kinshsasa, Kinshasa, Congo, Democratic Republic of the, [20]Ctr. Médical ISTM Tshumbe, Tshumbe, Congo, Democratic Republic of the, [21]Ctr. de Santé de Foyer Matete, Kinshasa, Congo, Democratic Republic of the, [22]Policlinique Louis Second, Dungu, Congo, Democratic Republic of the

Background: Differences of Sex Development (DSD) are a spectrum of conditions with atypical sex chromosomes and/or anatomical or gonadal development. While accurate diagnosis is difficult even in high-income countries, little is known about DSD in low-resource settings, specifically the Democratic Republic of Congo (DRC). High cost of and limited access to care, and lack of trained specialists are obstacles to optimal care for rare diseases in the DRC. Standard karyotyping and other tests necessary to distinguish DSD conditions with different management are unavailable locally. To address these gaps we established a care network with local and foreign experts in genetics, pediatrics, gynecology, surgery, endocrinology, psychology. **Methods:** Health professionals were sensitized to the recognition of DSD in newborns, children, and adults through social media and clinical visits. Patients' clinical evaluations were standardized using an adaptation of the DSD-Translational Research Network (DSD-TRN) forms. To assess etiology, saliva was sent to Children's National in Washington, DC for exome (n=23 for 10 probands) or genome (n=43 for 20 probands) sequencing. **Results:** Thirty-one participants (2 months to 26 years) and their bio-relatives were recruited from December 2021 to May 2023: 17 had a male sex-of-rearing, 12 female, and 2 had experienced a change from female to male. All were brought to medical attention because of genital atypia. The 10 exomes identified only one possible etiology, a sex chromosome aneuploidy. Genome sequencing identified 6 causative variants. An intronic variant in *AR* established a diagnosis of partial androgen insensitivity in a 22 year-old XY male, raised as a female until age 5. Trisomy 21 explained the hypospadias in an XY boy. A recurrent homozygous *SRD5A2* variant described in the historical cases of 5α-reductase deficiency in the Dominican Republic was found in a family with two XY females and an unrelated individual. Others were a de novo stop gain in *NR5A1* and, *NR5A1* and *AR* missense variants. No definitive diagnosis could be made in the 8 individuals where short-read sequencing suggested XX sex complement. A single *CYP21A2* Gln319* variant was found in an XX male, but a diagnosis of Congenital Adrenal Hyperplasia could not be established as short-read sequencing is inadequate to identify the gene conversions responsible for most pathogenic alleles in this gene. **Conclusion:** This experience of DSD multidisciplinary clinic in DRC showed local challenges for the diagnosis and optimal management of these complex conditions. Collaborations between international and local experts may help advance diagnosis in low resources settings.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Session II

PB2067 † Early commitment to neuroectoderm is reduced in Down Syndrome iPSCs due to dysregulation of a Primary Cilium-Autophagy-Nrf2 (PAN) axis

Authors:

M. Chaklader, B. Rothermel; UT Southwestern Med. Ctr., Dallas, TX

Down syndrome (DS), caused by trisomy of human chromosome 21, is the most common genetic cause of intellectual disabilities and is associated with an increased risk for a wide range of developmental and degenerative processes. Analyses both in human and mouse models have shown delays in cortical neurogenesis during the development of DS embryos. The underlying cause of this defect has yet to be fully understood. We have found that iPSCs derived from individuals with DS (3S-iPSCs) are intrinsically less capable of committing to a $PAX6^+$ neuroectodermal lineage (NE) when compared to isogenic, diploid iPSCs controls (2S-iPSCs), both under conditions of spontaneous differentiation and when subjected to directed differentiation protocols. The primary cilium is a threadlike projection that extends from the cell membrane of many cell types, acts as a signaling hub, and is involved in many developmental and adaptive processes. Other investigators have proposed that the primary cilium helps direct stem cells toward a neuroectodermal fate during normal development by increasing autophagic activity to reduce NRF2-dependent transcription. Because the cilium grows during G_1/G_0 of the cell cycle, rapidly dividing cells are disfavored for signaling through this pro-neurogenic Primary Cilium-Autophagy-Nrf2 (PAN) Axis. Interestingly, DS cells have previously been noted to have abnormal primary cilia, be deficient in autophagy, and have elevated NRF2 activity. We provide evidence that 3S-iPSCs proliferate more rapidly than 2S-iPSCs leading to the suppression of PAN signaling. This suggests that the neurogenic deficiency of DS may be initiated during the earliest steps of lineage commitment. Studies are ongoing to test whether inhibitors that lengthen the G1 phase of the cell cycle can increase cilia formation and, thus, neuroectoderm differentiation in 3S-iPSCs. Because primary cilia impact various organs and organ systems, we postulate that defects in this signaling axis may contribute to other DS-associated pathologies.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Session III

PB2068 Early post-zygotic mutations are inherited by the germline and reveal its polyclonal origin in humans

Authors:

Y. Jang[1], L. Tomasini[2], A. Szekely[3], F. Vaccarino[2], A. Abyzov[1]; [1]Mayo Clinic, Rochester, MN, [2]Yale Univ., New Haven, CT, [3]Yale Univ. Sch. of Med., Hamden, CT

Little is known about the origin of germ cells in humans. We previously leveraged mutations to reconstruct zygote-rooted cell lineage ancestry trees in a phenotypically normal woman, termed NC0. Here, by sequencing the genome of her children and their father, we analyzed the transmission of early pre-gastrulation lineages and corresponding mutations across human generations. We found that the germline in NC0 is polyclonal and consists of at least two clones descending from each of the first two blastomeres. Our results, complemented by public data from multi-children families, also show that previously detected unbalances in early lineages allocation in the soma are not detected in the germline, suggesting a fundamental difference in mechanisms of lineage allocation between the soma and the germline.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Session I

PB2069 Epigenetic regulator *Tet2* modulates ovarian dysfunction in a model of the Fragile X premutation

Authors:

K. Shelly[1], S. Call[1], P. Jin[2], E. Allen[1]; [1]Emory Univ., Atlanta, GA, [2]Emory Univ Sch Med, Atlanta, GA

Fragile X-associated Primary Ovarian Insufficiency (FXPOI) is the clinical diagnosis given to roughly 20% of women who carry a premutation (PM) allele. This disease is driven by the expansion of a CGG-repeat tract in the 5' untranslated region (UTR) of the *FMR1* gene to 55-200 repeats. Limited availability of primary tissue and patient cohorts with elevated FSH levels and cessation of menses for 4-6 months prior to age 40 mean the molecular mechanism of FXPOI is understudied. We utilize mouse models of FXPOI with an ectopically expressed tract of CGG_{99} repeats capable of producing expanded repeat RNA (RNA-only) or rCGG and a non-canonically translated polyglycine product (FMRpolyG+RNA) from the *Rosa26* locus. Previous work demonstrated that follicular development, steroidogenesis, and ovulation were perturbed. Using untargeted metabolomics on FXPOI and control women we found that 2-oxoglutarate was altered in FXPOI women plasma. By crossing the PM mice with the mice haploinsufficient for the 2-oxoglutarate dioxygenase-dependent enzymes *Tet1* and *Tet2*, we observed that female mice expressing RNA-only alleles with haploinsufficiency of *Tet2* display reduced breeding compared to controls, which is not seen in mice only expressing CGG RNA. Decreased progeny results from a combination of reduced litters and pups per litter. Surprisingly, FMRpolyG+RNA;$Tet2^{+/-}$ females are not significantly different from controls or FMRpolyG+RNA females in producing progeny, suggesting a partial, but not complete, rescue of ovarian function. While the intraindividual variation seems to blunt a significant change from FMRpolyG+RNA itself, half of FMRpolyG+RNA;$Tet2^{+/-}$ females do not exhibit the consistent cessation of breeding observed in all FMRpolyG+RNA females. Ovaries collected reveal that corpora lutea are also present in the ovaries of FMRpolyG+RNA;$Tet2^{+/-}$ females in contrast to the FMRpolyG+RNA group, supporting that loss of ovulation was corrected. This work adds to the growing body of human omics data and previous mouse studies, that point to impaired ovulation as central to FXPOI. We are currently determining whether the role of Tet2 in demethylating ovulation-associated genes drives this phenotype using transcriptomic- and 5hmC-profiling data and whether these markers may be detected in human patients, enabling their use as biomarkers of FXPOI.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Session II

PB2070 Establishment of Nepal's first Non-Invasive Prenatal Testing (NIPT) Service Laboratory

Authors:

S. Thapa[1], N. Thakur[1], G. Joshi[1], S. Khanal[1], U. Thakur[2], A. Jang Kunwar[1]; [1]Kathmandu Ctr. for Genomics and Res. Lab. (KCGRL), Lalitpur, Nepal, [2]Nova Intl. Diagnostics, Lalitpur, Nepal

Design: Non-Invasive Prenatal Testing (NIPT) for fetal aneuploidies using cell-free DNA (cfDNA) has been widely adopted in clinical practice due to its improved accuracy. Several NIPT tests have been developed and validated. The core goal of cell-free based prenatal testing is to provide minimally invasive, clinically accurate screening for fetal chromosomal aneuploidies in the early stages of pregnancy. The purpose of this study was to establish a validated NIPT workflow for cell-free fetal DNA (cffDNA) sequencing from maternal plasma for the detection of trisomy 13, 18, and 21 on a semiconductor sequencing instrument. A total of ninety-one standard samples from healthy pregnant women were provided by Yourgene Health; their cfDNA library was prepared and loaded on the Ion 540 chip for sequencing. The sequencing output data was analyzed by using the bioinformatics pipeline of Yourgene Health. **Results:** Eighty-one samples were successfully validated out of a total of 91 samples. However, 3% of samples did not meet quality, and 3% of sample libraries had low reads and failed validation. Moreover, the validation of NGS workflow on the Ion Genestudio S5 for fetal aneuploidy detection would allow us to build up a reference database. Based on the validated database 224 clinical samples were analyzed out of which 43 samples failed QC and the 37 samples were analyzed for chromosomal aneuploidy risks, all had Z-score below 2.
8. **Conclusion:** This research aims to introduce and set up Nepal's first NGS laboratory for NIPT, utilizing the Ion Torrent technology to provide a complete in-house solution for pregnant women in Nepal. The establishment of this prenatal test signifies a significant advancement in prenatal care and reproductive health.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Session III

PB2071 Evaluating Gene Expression Profiles in Clinically Accessible Tissues for Prenatal Diagnostic Testing of Genetic Disorders

Authors:

M. Vladoiu[1], R. Zemet[1], C. Parobek[1], S. Zhao[1], J. Sinson[1], S. Tankersley[2], I. Van Den Veyver[1], P. Liu[1]; [1]Baylor Coll. of Med., Houston, TX, [2]Baylor Genetics, Houston, TX

Next generation sequencing (NGS) assays such as whole exome sequencing (WES) have considerably improved diagnostic yield in invasive prenatal testing particularly for monogenic disorders and after negative cytogenetic results. While approximately 40% of fetal abnormalities are diagnosed prenatally by conventional cytogenetic assays, WES has an incremental diagnostic yield of 31% across various prenatally detected malformations. With the advent of NGS technologies and the increasing number of variants detected in prenatal samples, there is a pressing need for confirmatory assays to promote accurate interpretation. RNAseq using whole blood and skin biopsies is expanding variant detection and provides functional evidence for variant interpretation. Our study aims to investigate the utility of RNAseq in prenatal diagnosis using chorionic villus sampling (CVS) and amniotic fluid (AF) samples. To determine clinically relevant expressed genes in prenatal samples, we performed RNAseq using polyA library preparation and standard RNAseq pipeline of 21 prenatal, 10 fibroblast and 10 blood samples. We reviewed the literature and queried existing databases (OMIM, HPO, PanelApp) to identify genes whose disruption leads to a prenatal phenotype. Gene expression profile comparisons were conducted between prenatal/postnatal samples and the prenatal phenotypes gene list, GTEx and single-cell RNAseq fetal datasets. Our results demonstrate an important overlap between the prenatal expression profiles and genes associated with prenatal phenotypes. A trend for a higher proportion of genes associated with urogenital, stillbirth, hydrops and central nervous system disorders was observed among AF samples, versus musculoskeletal and cardiovascular associated genes among CVS samples. Differential gene expression also revealed unique gene differences between CVS and AF samples. Some examples include LHX1, KCNJ16 and HNF1B showing higher expression levels in AF samples, and GATA4, PITX1 and GPC3 in CVS samples. Comparison between expression profiles of prenatal samples with fetal tissue-specific single-cell RNAseq data and GTEx database shows gene expression overlap with multiple organ types, particularly for AF samples suggesting that this sample type can provide a broader overview of genes associated with specific organ development. In conclusion, prenatal RNAseq in clinically accessible tissues reveals the detection of variety of unique genes with implications for prenatal conditions as well as fetal organ development. Our findings highlight the potential of RNAseq as a valuable assay supporting NGS variant interpretation in the prenatal setting.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Session I

PB2072 † Genetic Ancestry, Exome Sequencing and Classification of Variants Identified in Genes Associated with Inborn Errors of Metabolism

Authors:

J. Najera[1], Y. Mavura[1], A. Adhikari[2], M. Kvale[1], N. Risch[1], F. Chen[1], R. Gallagher[3]; [1]Inst. for Human Genetics, Univ. of California San Francisco, San Francisco, CA, [2]Illumina Corp., San Diego, CA, [3]Div. of Med. Genetics, Dept. of Pediatrics, Univ. of California San Francisco, San Francisco, CA

Background Newborn screening (NBS) plays a crucial role in identifying and treating inborn errors of metabolism (IEMs), a group of largely recessive genetic diseases. Traditional NBS methods, such as tandem mass spectrometry (MS/MS), have been successful in identifying IEMs, but the use of genetic screening pipelines could increase the positive predictive value. While the role of exome sequencing (ES) in NBS is promising, the pipeline approach is highly dependent on variant pathogenicity curation. Consequently, information disparity across populations related to pathogenicity of variants could lead to reduced accuracy of variant classification in patients of non-European descent. In this study we examine ES for IEMs and the relationship of genetic ancestry to diagnostic outcomes. **Methods** The California Department of Public Health provided dried blood spots from 1,728 neonates who screened positive by MS/MS for one of 48 IEMs between July 2005 and December 2013, including all individuals with a confirmed diagnosis. ES was performed and 833 confirmed true positive cases passed QC for further analysis. A variant annotation pipeline was created that analyzed 78 genes underlying the 48 IEMs, which included previously documented pathogenic variants, CNVs, and all protein altering variants with a global minor allele frequency <0.5%. The true positive individuals who were homozygous or with two annotated variants in a matching gene were considered exome positive (diagnosed), all others were considered exome negative. We characterized the genetic ancestry of exome positive (N=668) and exome negative (N=165) cases. Individual admixture was estimated from the exome data using HGDP WGS data as reference and the program ADMIXTURE. Differences in ancestries for positive and negative cases were assessed by linear regression. **Results** Overall genetic ancestries were: 7.5% African, 11.2% East Asian, 46.4% European, 8.5% Middle Eastern, 20.5% Native American, 5.6% South Asian. Native American ancestry was positively associated (P<0.02) with an exome positive diagnosis (22.7% in positive, 16.9% in negative); all other ancestries showed no association, either positive or negative. **Conclusion** Diagnostic yield of IEMs by ES was positively associated with Native American ancestry, and not negatively associated with any non-European genetic ancestry. While we found numerous ancestry-specific founder variants of varying frequency, there was no preferential bias in terms of diagnosis. These results suggest ES diagnosis of recessive IEMs is not negatively impacted by non-European genetic ancestry.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Session II

PB2073 † Genetic determinants of spontaneous preterm birth in Indian women: Transethnic and population specific markers.

Authors:

E. Bhattacharjee[1], R. Thiruvengadam[2], *. Ayushi[3], GARBH-Ini study group, P. Kshetrapal[3], S. Bhatnagar[3], P. P. Majumder[4], A. Maitra[1]; [1]Natl. Inst. of BioMed. Genomics, Kalyani, India, [2]Pondicherry Inst. of Med. Sci., Puducherry, India, [3]Translational Hlth.Sci. and Technology Inst., Faridabad, India, [4]John C. Martin Ctr. for Liver Res. and Innovation, Kolkata, India

Preterm birth (PTB), or live birth before 37 weeks of gestation, is the major cause of neonatal deaths and adverse health severities. Prevalence of PTB varies widely across countries and is considerably high in India and other South-Asian countries. To investigate whether such variation is due to genetic differences, we derived weighted genetic risk score (GRS) using the genomewide significant SNPs from the 23andMe spontaneous PTB (sPTB) genome-wide association study (GWAS) summary statistics and tested it on 4682 Indian women from GARBH-Ini cohort. We tested GRS prediction using Nagelkerke's pseudoR^2 adjusting for age, BMI, parity, occupational status and top 10 PCs. The European-derived GRS performed poorly in Indian women, explaining 0.9% of sPTB risk (p=0.44). Hence, to identify SNPs that can robustly predict risk of sPTB in Indian women, we conducted GWAS using genotype data of these 4682 women adjusting for above mentioned covariates. We found 38 SNPs with p<1e-5, and one genomewide significant SNP, rs57480735 (OR=1.57, p= 2.78e-8), which was also associated with gestational length (B= -0.16, p=2.05e-3). Of these 39 SNPs, 15 were also associated with early sPTB (<33 weeks). Additionally, from the cross-ancestry meta-analysis with 23andMe sPTB GWAS summary statistics, we identified 199 transethnic SNPs, of which rs35760881 was the most significant (p=8.45e-7). Functional investigation revealed that rs57480735 is a cis-eQTL of *NR2F2* which participates in maternal placenta development. Using placental single cell RNA sequencing data we found its high expression in the cells of maternal origin. Further, integrating genotype data with clinical data from the cohort, we found that preterm delivering women who carried minor allele at rs57480735 had higher uterine arterial pulsatility index, suggesting poor placental development. Interestingly, the transethnic SNP, rs35760881 is a cis-eQTL of *AKIP1,* which via inflammation may cause early myometrial contraction. Several immune cells in the placenta showed high expression of *AKIP1*. Moreover, we found that women who delivered preterm and carried minor allele at rs35760881 had high systemic mean arterial pressure, suggesting early myometrial contraction. Finally, from the AUROC analysis, we found that the 39 significant SNPs along with 199 transethnic SNPs have high predictive ability for sPTB (AUC 0.8). Our findings thus suggest that inflammation may result in sPTB in a population independent manner, while, poor placental development might be a population specific event. The significant SNPs and transethnic SNPs may help to stratify at-risk group of women for interventions leading to reduction of sPTB.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Session III

PB2074 Genetics services are an integral part of interdisciplinary care of individuals with disorders of sex development.

Authors:

C. Keegan, M. Hannibal, **L. Mohnach**; Univ Michigan, Ann Arbor, MI

Disorders of sex development (DSD) include conditions involving anomalies of the sex chromosomes, gonads, reproductive ducts and genitalia. Patients and families with DSD are faced with complex medical and psychosocial needs and benefit from an interdisciplinary care team composed of specialists from urology, endocrinology, gynecology, psychology, and genetics. A clinical diagnosis is often made using karyotype analysis, hormone testing, imaging, and physical examination. However, a specific genetic diagnosis helps to tailor management by providing more insight into fertility potential, gonadal tumor risk, and response to surgery. Many DSD conditions have phenotypic overlap, which can confound management and inheritance counseling. A genetic diagnosis is important for families to understand recurrence risks, reproductive options, and identify at-risk relatives. Further, some conditions have an increased risk for tumor development, but their diagnosis may be missed or not clinically recognized until later in life. Lastly, gene-specific identification allows a better understanding of the phenotypic variability and natural history of these distinct conditions. We evaluated the diagnostic rate of genetic testing in our DSD patient population to determine the utility of genetic testing in this clinic. Patient records from 2011-2023 were reviewed. Since the inception of the program, 256 patients have been evaluated in our DSD clinic. Common referral indications include: 46,XY phenotypic female, 46,XY undervirilized phenotypic male, atypical genitalia, congenital adrenal hyperplasia, sex chromosome abnormalities, and complex cloacal/urogenital malformations. Genetic evaluation differed by phenotypic presentation but most often included a DSD gene panel or whole exome sequencing. A genetic diagnosis was made in 100% of patients with a 46,XY female phenotype with absent uterus and in 54% of patients with 46,XY karyotype and female phenotype, with a uterus present. Our diagnostic rate was the lowest (29%) in those presenting with 46, XY male phenotype, most likely due to the diversity of etiologies associated with hypospadias and cryptorchidism. Genetic testing in the DSD patient population has a high yield, especially in those with 46,XY female phenotype without a uterus. Genetic services are valuable for all patients seen in the DSD clinic as a genetic diagnosis provides the family and interdisciplinary team with more details to formulate multisystem medical management decisions. Genetic providers are integral to the DSD team to provide genetic testing and serve a key role in patient education and psychosocial support.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Session I

PB2075 Genome-wide investigation reveals significant interaction of *FURIN* with maternal vitamin intake on orofacial cleft risk.

Authors:

Z. Erdogan-Yildirim[1], N. Mukhopadhyay[1], J. C. Carlson[2], J. R. Shaffer[2,1], S. M. Weinberg[1], M. L. Marazita[1,2,3]; [1]Ctr. for Craniofacial and Dental Genetics, Dept. of Oral and Craniofacial Sci., Sch. of Dental Med., Univ. of Pittsburgh, Pittsburgh, United States, Pittsburgh, PA, [2]Dept. of Human Genetics, Sch. of Publ. Hlth., Univ. of Pittsburgh, Pittsburgh, United States, Pittsburgh, PA, [3]Clinical and Translational Sci., Sch. of Med., Univ. of Pittsburgh, Pittsburgh, United States, Pittsburgh, PA

Nonsyndromic orofacial clefts (NSOFC) are congenital defects affecting the lip and/or the palate. NSOFC have a multifactorial etiology involving genetic factors and intrauterine environmental exposures. Hence, exploring gene-environment (GxE) interactions may elucidate the complex biological mechanisms of NSOFC. Previous research, including our own, has investigated GxE interactions in multiethnic or European cohorts using both trios and case-control designs.In this study, we explore the genome-wide GxE interactions between genetic variation and intrauterine exposures to maternal smoking, alcohol or vitamin intake within ± 3 months of conception.Our study cohort consists of 547 Filipinos with cleft lip or cleft lip and palate (CL/P) and 264 unrelated controls with no family history of clefting. Genotyping was done via the Illumina GDA-8v1-0 array and additional genotypes were imputed using TOPMed reference panel via minimac4. Variants with imputation rsq &lt 0.8 were excluded.We conducted a two-stage design GxE analysis on a total of 5,245,942 variants. We used logistic regression models accounting for the first five principal components of ancestry (PCA), and assuming an additive genetic model in PLINK v1.9. In the first stage, we screened for genetic and/or GxE associations with a p &lt 0.05 using a joint 2 degree of freedom test (*2df*). In the second stage, we tested the GxE effect alone while applying a significance threshold accounting for the dependence between the joint test and the GxE test (as determined via simulation) and further adjusted by the number of independent loci (p_{GE} &lt 0.00275/716).Preliminary findings revealed a significant interaction effect influencing CL/P between variants on 15q26.1 and maternal vitamin intake. The strongest signal was obtained at rs881431 ($p_{2df} = 8.23 \times 10^{-6}$ and $p_{GE} = 3.55 \times 10^{-6}$) in the intergenic region between *VPS33B* and *SV2B*. It locates about 170kb distal from *FURIN* encoding a serine protease involved in the cleavage of protein precursors including growth factors, their receptors, and extracellular matrix proteins. Furin is vital in very early embryonic development. Germline knockout mouse embryos lacking *FURIN* cannot survive beyond 11 days. Folic acid impacts the proteolytic activity of furin. Additionally, mutations in *GDF11* that prevent cleavage by furin are known to cause CL/P. In summary, our study identified a significant interaction near *FURIN* with maternal vitamin intake associated with risk of CL/P in Filipinos. Notably, several loci showed suggestive interaction with alcohol use and smoking. Further follow-up analyses and validation are critical to strengthen and corroborate the findings.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Session II

PB2076 Genotype - phenotype identification in a case of genetically described (5p inversion deletion and duplication syndrome) developmental epileptic encephalopathy.

Authors:

E. Yaman[1], Y. Ataş[2], G. Şen[2], D. E. Toprak[2], S. Kanmaz[2], E. Şimşek[3], T. Atik[3], S. Yılmaz[2], H. Tekgül[2]; [1]Ege Univ. Faculty of Med., Izmir, Turkey, [2]Ege Univ. Faculty of Med., Dept. of Pediatric Neurology, Izmir, Turkey, [3]Ege Univ. Faculty of Med., Dept. of Pediatric Genetics, Izmir, Turkey

Background:
The inverted duplication with deletion of 5p, shortly known as *inv dup del 5p*, is a rare chromosomal rearrangement disorder that has been described in detail by only a handful of cases in the literature. The resulting phenotype, which was first described by Sreekantaiah et al in 1999, is a complex one, differing depending on both the size and the position of the deletion and duplication in each case; generally presents with development delay accompanied with cardiac, neurocranial and musculoskeletal anomalies and varying dysmorphic facial features.

Introduction:
Here we report a case of a 15 month old male presenting ***minor facial dysmorphologies***, ***developmental delay***, and ***epileptic encephalopathy***. The patient was receiving dual anti-seizure treatment for behavioral pause and automatism seizures and presented to us with focal motor status epilepticus. Microarray analysis for the genetic etiology-specific identification of developmental and epileptic encephalopathy showed a **large partial deletion** (**2.7 Mb**) located at 5p15.33 and **a much larger duplication** (**22.9 Mb**) located at 5p15.33p14.1. The family segregation study had normal karyotypes indicating that the rearrangement was **de novo**.

Methods:
Cranial magnetic resonance imaging, metabolic tests, electroencephalography, genetic tests (microarray and karyotyping) and karyotyping of the family were performed to investigate the etiology of developmental epileptic encephalopathy and the de novo status of genetic rearrangement.

Case Discussion:
The patient was noted *not* to have the characteristic findings of 5p deletion syndrome (Cri-du-chat) and cat-like-cry was *not* observed. Consistent with the variable clinical findings associated with 5p duplication, no anomalies or system findings were detected, except for developmental delay, minor dysmorphic findings, and epilepsy. The patient was characterized by ***focal motor*** and ***non-motor seizures*** and ***developmental epileptic encephalopathy*** with EEG characteristics. *Focal epileptiform activities* were present on initial presentation and control EEG monitoring.

Conclusion:
To the best of our knowledge, **our patient is the eighth reported case in the literature and the fifth reported case of postnatal pure 5p inversion duplication deletion syndrome.** The genotype-phenotype and electro-clinical description of this rare case of developmental epileptic encephalopathy caused by this dual rearrangement is presented. *Describing the clinical and genetic features of such cases, collecting and detailing information on genotype-phenotype correlations will have great prognostic value and will be an important contribution to the literature.*

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Session III

PB2077 GWAS meta-analysis on heavy menstrual bleeding in three large biobanks reveals 28 loci suggesting multiple contributing pathways

Authors:

A. Johnson[1,2], J. Cunha[1,2], J. Huffman[3], P. Wilson[4,5], K. Cho[6], J. Johnsen[7], M-H. Chen[1,2], F. Thibord[1,2]; [1]Natl. Heart Lung and Blood Inst., Bethesda, MD, [2]The Framingham Heart Study, Framingham, MA, [3]Palo Alto Veterans Inst. for Res., Boston, MA, [4]Emory Univ., Atlanta, GA, [5]Atlanta VA Med. Ctr., Decatur, GA, [6]VA Boston Hlth.care System, Boston, MA, [7]Univ. of Washington, Seattle, WA

Background: Heavy menstrual bleeding (HMB) is the most prevalent global clinical bleeding issue and contributes to various medical complications and negative quality-of-life outcomes. Large biobanks such as UK Biobank (UKBB), the VA Million Veteran Program (MVP), and FINNGEN could help discover HMB genetic risk loci using a well-powered genome-wide association study (GWAS). To our knowledge, this is the first study to report a GWAS focused specifically on HMB in the general population.**Methods**: ***Study Population***: Data from individuals with HMB ICD-9/10 codes were sourced from UKBB (13,500 cases/247,045 total), FINNGEN (19,024 cases/119,687 total), and MVP (7,550 cases/56,898 total). ***GWAS and Meta-Analysis***: UKBB, MVP, and FINNGEN used generalized mixed effects models implemented in SAIGE and REGENIE that account for relatedness and population structure with adjustments for age, genetic principal components, and cohort-specific covariates. For the meta-analysis, we applied a fixed-effects approach using METAL. Transcriptome-wide association studies (TWAS) were conducted using FUSION and selected relevant GTeX tissues, followed by *coloc* analysis.
Results: ***Multi-population meta-analysis***: GWAS signals (p<5E-8) were identified in regions previously associated with uterine fibroids (UFs) (e.g. *CDC42/WNT4*, *GATA2*, *TP53*, *SYNE1*), coagulation (*FV*), and several relevant cancers (e.g., ovarian, uterine, breast). Some novel candidate genes were highly expressed in uterine tissues according to GTeX (e.g., *GMNC*, *STK26*). The strongest associations were a protective effect of the Factor V Leiden variant (OR=0.69, P=1.8E-33), and a protective regulatory region variant near *FSHB* (follicle-stimulating hormone B) (OR=0.88, P=1.35E-28), a gene pivotal in regulating menstrual cycles. Conditional analyses revealed multi-SNP independent signals only at 2 loci (chr1, chr10). TWAS revealed several additional loci including several with high co-localization probabilities (CALCRL, CDK2AP1, CCDC170) in hematological and genitourinary tissues (spleen, liver, vagina).
Conclusions: Variants were identified in loci with genes highly expressed in the uterus, or identified in prior UF or malignant female growth loci, suggesting major enrichment of these in HMB. Aside from *FV* Leiden and *GATA2*, there were relatively few overlaps with genetic signals for platelet phenotypes, anemia, hemostasis, or coagulation loci as had been hypothesized. This suggests that at least in older women the additional genetic drivers of HMB may be more related to regulation of hormones and menstrual cycles, and abnormal growth in female genitourinary tissues.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Session I

PB2078 Identification of novel genes for Coloboma using single cell RNA-Sequencing

Authors:

E. Boobalan[1], D. M. McGaughey[1], U. Neelathi[1], D. Sanchez-Mendoza[1], T. D. Fufa[2], R. B. Hufnagel[3], B. P. Brooks[1]; [1]Pediatric, Dev.al & Genetic Ophthalmology Section, Ophthalmic Genetics & Visual Function Branch, NEI, NIH, Bethesda, MD, [2]Div. of Genome Sci., NHGRI, NIH, Bethesda, MD, [3]Med. Genetics and Ophthalmic Genomics Unit, Ophthalmic Genetics & Visual Function Branch, NEI, NIH,, Bethesda, MD

Uveal coloboma is a congenital disorder of ocular development caused by failure of optic fissure closure (OFC) beginning around the 5th week of human gestation. Although rare (0.5-2.6 cases/10,000 live births), it is estimated to cause up to 10% childhood blindness. Although significant progress has been made in identifying genes for coloboma in humans, many cases remain unsolved molecularly. To identify genes important for optic fissure closure we used single cell RNA-sequencing of wild-type C57BL/6J eyes at E11.5—the time point at which the optic fissure is closing.

Single cells were made by digesting eyes with Trypsin for 10-20 minutes at 37^{o}C. Libraries were made with 10x Chromium controller using Chromium Next GEM Single Cell 3' Reagent Kits v3.1 (Dual Index). After quality control filtering, we recovered 4034 cells across 11 cell types, including 418 cells from the optic fissure/stalk, 1598 presumptive neural retina (pNR), 538 presumptive RPE (pRPE), and 1017 fibroblasts. We identify a cluster of cells on enriched genes from previously known to occur in the optic fissure (e.g., Adamts16, Ntn1, Pax2, Vax1) as well as potentially novel genes (e.g., Afapl2, Sox1, Lrrn1, Shtn1). Immunohistochemistry (IHC) and In Situ techniques were used to evaluate the temporal and spatial expression of genes of interest in mouse embryonic eye sections. Morpholino knock down assays were done on to assess the role of genes in zebrafish ocular development.

Sox1 expression was confirmed in the ventral pNR and pRPE at the optic fissure and lens by IHC at E11.5 and found to be significantly down regulated in the optic fissure (but not the lens) at E12.5. A similar pattern of expression was noted in zebrafish by ISH and morpholino knockdown causes coloboma like phenotype in zebrafish. Similarly, Lrrn1 and Shtn1 mRNA expression were noted by RNAScope assay in the distal naso-ventral retinal neuroblasts, proximal ventral pRPE and weaker expression in dorsal pNR as well as at the extra ocular regions for Lrrn1; and ventral pNR and pRPE for Shtn1 at E11.5 and significantly downregulated at E12.5. These data provide further insight into the molecular mechanisms of OFC and may identify candidate genes for human disease.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Session II

PB2079 Illuminating Prenatal Diagnosis: Enhancing Clinical Relevant Variant Interpretation via Transcriptome Analysis in Amniotic Fluid Cells.

Authors:

Y. Cao[1], S. Ketkar[2], H. Dai[2], M. Shi[1], Y. Zheng[1], Z. Liu[1], Y. Kwok[1], Z. Dong[1], P. Liu[2], T. Leung[1], B. Lee[2], R. Choy[1]; [1]The Chinese Univ. of Hong Kong, Hong Kong, China, [2]Baylor Coll. of Med., Houston, TX

Transcriptome analysis has garnered significant interest as it can offer additional information about aberrant expression and splicing that complements DNA-based genetic testing. It aids in clinically relevant variant interpretation as well as helps detect variants that may otherwise go unnoticed with genetic testing alone. RNA sequencing (RNA-seq) has been shown to improve the diagnostic rate by 8-36% for postnatal patients with suspected Mendelian disorders but inconclusive genetic investigations. However, its use in prenatal diagnosis is currently limited. Our study aimed to assess the feasibility and utility of RNA-seq supplementing prenatal genetic testing analysis. We retrospectively recruited 50 prenatal subjects with amniotic fluid cells samples, which includes 4 cases with normal live birth, 6 cases with known pathogenic variants and two deletions, 40 cases with various fetal structural abnormalities but inclusive genetic testing results. RNA libraries of AF samples are formed via polyA capture and rRNA removal and reverse transcription of cDNA. PE150 technology is employed to sequence the RNA extracted from AF samples. Principle complement analysis showed clustered and unique gene expression of expression in these amniotic fluid cells compared with published dataset of fibroblasts and whole blood. Outlier-Based differential gene expression analysis of the reads counts/genes flagged a Lissencephaly case with an outlier *PAFAH1B1* gene which matched its 1.9Mb deletion on 17p13.3, including *PAFAH1B1* gene. Alternative splicing analysis detected another positive case with a single-exon deletion of the *PIEZO2* gene by showing in-frame exon skipping in transcripts without obvious decrease in expression. Allele-specific expression analysis reveals no dominant transcripts with four heterozygous missense variants or abnormal splicing events. Additionally, we examined the alternative transcript/s generated by genome-wide rare variants with a spliceAI delta score>0.2 in 30 cases using 30x genome sequencing data. The higher the score, the greater the likelihood that the variant could result in transcript changes, while commonly occurring in a leaky pattern, suggesting that PVS1 Strength evidence be used with caution. Our study highlights the great potential of transcriptome analysis through RNA sequencing as a promising prenatal diagnostic tool.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Session III

PB2080 Inferring the number of founding primordial germ cells in humans from *de novo* mutation data

Authors:

Z. Gao; Univ. of Pennsylvania, Philadelphia, PA

Primordial germ cells (PGCs), the precursors of mature germ cells, are established around the gastrulation stage during embryogenesis in mammals. Extensive research has shed light on the developmental process of germ cells in mice, but studying the corresponding process in humans presents practical challenges. Moreover, the distinct morphological structure of the post-implantation epiblast in humans and mice raises questions regarding the generalizability of knowledge acquired from mice to human development. To address these questions, I propose a genetic approach to learn about germ cell development in humans, leveraging the recently available genomic sequencing data of pedigrees. Deep sequencing of human families with multiple children reveals a considerable fraction of *de novo* mutations to be shared among siblings. The presence or absence of these sibling-shared mutations in each child is informative about the genotype of the corresponding diploid germ cell in the parent. Based on this rationale, I developed an algorithm to determine the *minimum* number of founding cells at the time of PGC specification. Analysis of existing human pedigree datasets indicates that least 2-7 cells commit to the germ cell fate in each individual. This finding aligns with previous observations in mice, suggesting that germ cells are not monophyletic relative to somatic cells. The estimates do not significantly differ between sexes, consistent with the timing of PGC specification before sex determination in mammals. Additionally, the estimate is not influenced by the number of offspring in a family but correlates with the number of informative mutations detected, indicating that detecting more sibling-shared mutations, such as those at short tandem repeats, can further refine the lower bound estimate. Next, I developed a formula to estimate the *expected* number of founding PGCs, under a model of synchronized, symmetric division during germ cell proliferation. Analysis of human pedigree data suggests a small number (<50) of founding cells at PGC specification. Although this estimate seems surprisingly small, it is consistent with observations from live imaging of developing mouse embryos. Alternatively, if germ cell proliferation involves asymmetric division, asynchrony, or cell death, the actual number might exceed the estimate. Overall, our study provides refined estimates of the number of founding PGCs that contribute to the gamete pool in humans. More generally, the new genetic approach developed in this study, combined with classical anatomical, molecular, and developmental approaches, promises to uncover a more detailed model of germ cell development.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Session I

PB2081 Integrative analysis of transcriptome dynamics during human craniofacial development identifies novel candidate disease genes

Authors:

J. Cotney[1], T. Yankee[1], S. Oh[1], E. Winchester[1], A. Wilderman[1], K. Robinson[2], T. Gordon[3], J. Rosenfeld[3], J. VanOudenhove[1], D. Scott[4], E. Leslie[5]; [1]UConn Hlth., Farmington, CT, [2]Emory Univ., Decatur, GA, [3]Baylor Coll. of Med., Houston, TX, [4]Baylor Coll. Med, Houston, TX, [5]Emory Univ., Atlanta, GA

Craniofacial disorders are among the most common of all congenital defects. A majority of craniofacial development occurs early in pregnancy. To fully understand how craniofacial defects arise, it is essential to observe gene expression during this critical time period. To address this, we performed bulk and single-cell RNA-seq on human craniofacial tissue obtained from 4 to 8 weeks post conception, establishing the most comprehensive profile of the transcriptome in the early developing human face to date. We identified 239 genes that were expressed more strongly in craniofacial tissues relative to dozens of other human tissues and stages. We found that craniofacial specific enhancers are enriched within 400 kb of these genes, establishing putative regulatory interactions. To further understand how genes are organized in this program, we constructed coexpression networks and show that disease candidates are likely genes that are coexpressed with many other genes, serving as regulatory hubs within these networks. We leveraged large functional genomics databases including GTEx and gnomAD to reveal hub genes that are more strongly expressed in craniofacial tissue and genes which are resistant to mutation in the normal healthy population. We validated our network findings using sequencing data from orofacial clefting trios and single-nucleus gene expression from the developing human face. Our unbiased methods revealed dozens of novel disease candidate genes that warrant further study.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Session II

PB2082 † Making a human specific model: Maternal Immune Activation and Neuroinflammation.

Authors:

B. Spiegelberg; Johns Hopkins Univ., Baltimore, MD

The human placenta is essential for the development and survival of a developing fetus. It is the main communicator between fetus and mother and is responsible for producing hormones and growth factors for correct fetal development during pregnancy. Maternal Immune Activation (MIA) plays a role in antagonizing the placenta and increasing risk of neurodevelopmental disorders, by causing a proinflammatory state. This information is known from human epidemiological studies and animal models, but the human placenta is a severely understudy. There is a gap in the scientific field for a human specific placenta model to better understand the interplay of MIA and neurodevelopmental disorders.
Primary human trophoblast stem cells (hTSC) and human pluripotent stem cells (hPSC) differentiated to hTSC can potentially model placental processes in vitro. Yet, it remains controversial how the differentiation of human pluripotent stem cells to trophectoderm relates to in vivo development and the factors required for this differentiation. Here, we demonstrate that the primed pluripotent state retains potency to generate trophoblast stem cells by activating EGF and WNT and inhibiting TGFb, HDAC and ROCK signaling without exogenous BMP4 (named TS). We map this specification by temporal single cell RNAseq compared to activating BMP4 or activating BMP4 and inhibiting WNT. TS conditions generate a stable proliferating cell type that is highly similar to six-week placental cytotrophoblasts with activation of endogenous retroviral genes and without amnion expression. Multiple primed iPSC and ES lines differentiate to iPS-derived-TSCs that can be passaged for at least 30 passages and differentiate to pure populations of multinucleated syncytiotrophoblasts (STB) and extravillous trophoblast cells. Having establish that primed IPSC and ES lines can be used to make an in vivo model of TSC, we were able to use this to create a model to study MIA and neuroinflammation with, specifically, STB as the mediator. We show that by exposing STBs to a RNA viral analog, they produce proinflammatory cytokines, mimicking MIA.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Session III

PB2083 Mutational and gene expression profiles of early-onset preeclampsia.

Authors:

N. Guo, H. Chavez, M. Huang, C. Mason, K. K. Nelson, M. Meads, O. Aisagbonhi, M. Parast, M. Horii, K. M. Fisch; UC San Diego, La Jolla, CA

Preeclampsia affects 3-8% of pregnancies and is a leading cause of severe maternal morbidity and mortality. Early-onset (EO) preeclampsia, defined as preeclampsia with an onset of <34 weeks gestation, is characterized by placental dysfunction and fetal growth restriction. The role rare genetic mutations play in placental dysfunction observed in preeclampsia is unexplored although widespread placental mosaicism has recently been described. To address this, we investigated rare genetic variants and gene expression from healthy pregnancies (>34 weeks gestation and <34 weeks gestation) and those affected by preeclampsia (late-onset (LO) >34 weeks gestation and EO <34 weeks gestation). We analyzed RNA sequencing data of 112 human placentas with detailed clinical adjudication and histopathological review. We identified variants (SNPs and INDELs) using the GATK Best Practices Workflow for RNAseq short variant discovery implemented in bcbio-nextgen and annotated variants with VEP. We implemented stringent variant quality filters including DP >50 and then prioritized rare variants by filtering for allele frequency in gnomAD v3.1.2 < 0.05, removed non-exonic variants, filtered out INDELs that were not present in gnomAD v3.1.2 or dbSNP, and removed frequent variants (>70% of individuals) and genes (>50% of individuals), which resulted in a total of 8,127 variants in 2,736 genes. We performed clinical enrichment analysis with a Fisher's Exact Test in maftools to identify genes significantly enriched for variants in each clinical subset (Preeclampsia vs Healthy; Preeclampsia LO vs EO; Healthy >34 weeks vs <34 weeks; Preeclampsia LO vs Healthy >34 weeks; Preeclampsia EO vs Healthy <34 weeks), which identified 6 genes differentially mutated in Preeclampsia EO vs Healthy <34 weeks (UCHL5, CFLAR, GOGLB1, SH3D21, DVL1 and CSHL1) and 5 genes differentially mutated between Preeclampsia LO and EO (UCHL5, CSHL1, COL6A2, CFLAR and CD36; FDR < 0.05). We also uncovered significantly higher mutational burden in Preeclampsia EO relative to all other groups (p.adj < 0.05). Differentially expressed genes were identified after TMM-normalization using the limma-voom method, resulting in 538 differentially expressed genes between EO Preeclampsia compared to Healthy <34 weeks (adj-p < 0.05 and |log2FoldChange| > 1), which were significantly enriched for PI3K-Akt signaling, Calcium signaling, and cytokine-cytokine receptor interactions using g:Profiler (p-value < 0.05). This study has characterized rare variants, differentially expressed genes and increased mutational burden in placentas associated with early-onset preeclampsia warranting further investigation.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Session I

PB2084 NGS based Genetic Analysis of Embryos for Chromosomal Abnormalities in Nepal

Authors:

S. Khanal, S. Thapa, G. Joshi; Kathmandu Ctr. for Genomics and Res. Lab. (KCGRL), Lalitpur, Nepal

Introduction: Chromosomal abnormalities occurring in embryos can result in various reproductive and developmental disorders. Preimplantation Genetic Screening (PGS) is a method employed to detect chromosomal abnormalities in embryos before their implantation. This research entails an examination of the genetic data obtained from embryos of multiple patients in order to ascertain the prevalence of chromosomal abnormalities and their impact on embryo viability.**Methods:** Genetic analysis was performed on embryos from a cohort of patients using NGS. The embryos were assigned unique Patient IDs and Embryo IDs. Chromosome abnormalities were identified and categorized based on the impacted chromosome. The results were interpreted to determine whether the embryos were euploid (normal), aneuploid (abnormal), or Mosaic. **Results:** A total of 312 embryos from 78 patients were analyzed. Among them, 80 embryos were identified as euploid and 92 embryos were identified as aneuploid. The most common chromosomes impacted by abnormalities were 1 and 16. Mosaicism was observed in 128 embryos.**Interpretation:** The high occurrence of aneuploidy in this patient cohort suggests that chromosomal abnormalities are prevalent in embryos. These abnormalities can lead to a heightened risk of implantation failure, miscarriage, and developmental disorders. The presence of mosaicism indicates the possibility of both normal and abnormal cell populations within an embryo, potentially influencing overall embryo viability. These findings underscore the significance of genetic analysis through PGS, which enables the selection of euploid embryos for successful implantation and healthy pregnancies.**Conclusion:** This study provides valuable insights into the genetic characteristics of embryos within a patient cohort, revealing a substantial prevalence of chromosomal abnormalities and mosaicism. The identification of such abnormalities through PGS facilitates the selection of euploid embryos, thereby improving the success rates of assisted reproductive technologies and reducing the risk of genetic disorders in offspring. Further research and larger-scale studies are necessary to validate these findings and optimize the clinical application of PGS.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Session II

PB2085 Non-immune hydrops fetalis is associated with autosomal recessive variants in the MYB Binding Protein 1a (*MYBBP1A*) gene

Authors:

J. Tenorio[1], E. Mansilla[2], G. Fe Amalia[1], R. Reojo[3], P. Arias[1], A. Parra[4], P. Pascual[4], N. Gallego[4], M. Cazalla[4], R. Rodriguez-González[5], J. Nevado[6], P. Lapunzina[6]; [1]Inst. of Med. and Molecular Genetics (INGEMM) - La Paz Univ. Hosp. - CIBERER, Madrid, Spain, [2]Inst. of Med. and Molecular Genetics (INGEMM) - La Paz Univ. Hosp.-CIBERER, Madrid, Spain, [3]Dept. of Pathology, La Paz Univ. Hosp., Madrid, Spain, Madrid, Spain, [4]Inst. of Med. and Molecular Genetics (INGEMM) - La Paz Univ. Hosp., Madrid, Spain, [5]Div. of Maternal and Fetal Med., Dept. of Obstetrics and Gynecology, Hosp. Univ.rio La Paz. IdiPAZ, 28046 Madrid, Spain., Madrid, Spain, [6]INGEMM-HULP-CIBERER, Madrid, Spain

Non-immune hydrops fetalis (NIHF) is an extremely infrequent entity usually characterized by an excessive accumulation of fetal fluid within the fetal extravascular compartments and body cavities. The etiology of NIFH is highly variable with a proportion of idiopathic unknown cases, or associated with syndromic disorders. Here we present a fetus died at 27+3 weeks with a NIHF presenting with oligohydramnios, cystic hygroma, pleural effusion, and generalized hydrops with predominance of subcutaneous edema. The fetus also presented with ascites, severe and precocious IUGR and some skeletal anomalies. Trio whole exome sequencing was applied in order to screen for a possible genetic pathogenic variant. Variant prioritization according to a custom in-house algorithm allowed to identify two variants in *MYBBP1A*, one nonsense (NM_001105538.2:c.238G>T:NP_001099008.1:p.Gly80Ter) and one canonical splice-site variant (NM_001105538.2:c.3196-2A>G:NP_848696.1:p.Leu177Argfs*20), each inherited from a healthy parent. A previous report (PMID: 28425981) described another case with similar phenotype with a compound heterozygous variant in MYBBP1A (one identical to our patient). The two variants are predicted to be damaging by the *in silico* bioinformatic tools applied. The protein encoded by *MYBBP1A* play a role in many cellular processes including response to nucleolar stress, tumor suppression and synthesis of ribosomal DNA. Therefore, we suspect that *MYBBP1A* can be a strong candidate gene associated with the development of hydrops fetalis. It is necessary to collect more cases and further studies to understand the role of this gene and the mechanism associated with the development of the prenatal malformation.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Session III

PB2086 Noninvasive Preimplantation Genetic Testing for Aneuploidy from Spent culture Media in infertile couples.

Authors:

A. Patel[1], S. S. Chettiar[1], D. Jhala[2]; [1]Genexplore Diagnostics and Res. Ctr. Pvt. Ltd., Ahmedabad, Gujarat, India, [2]Sch. of Sci., Gujarat Univ., Ahmedabad, India

Preimplantation genetic testing for aneuploidy (PGT-A) is a primary method utilized to select euploid embryos for transfer and have supported numerous couples with infertility worldwide. PGT includes methods that allow embryos to be tested for Chromosomal aneuplodies and inherited conditions, relevant to embryo viability. PGT-A helps patients who have advanced maternal age, recurrent pregnancy loss, male infertility, or recurrent implantation failure to enhance ART outcomes. However, in order to obtain embryonic genetic material to perform PGT-A, a biopsy is required, involving the removal of one or more cells from Day 3 or Day 5 of cultured embryos. This invasive procedure requires highly sophisticated instruments, biopsy skills and also there have been concerns about mechanical damage to the developing embryos as embryo viability could be compromised in some cases. Recent development of technology using cell free DNA present in Spent culture media (SCM) overcomes requirement of embryo biopsy and has potential to detect chromosomal abnormalities termed as Noninvasive Preimplantation Genetic Testing for Aneuploidy (niPGT-A). niPGT-A is a novel approach to the genetic analysis of embryos allow clinicians to select embryos with normal genomic content without damaging embryos. Here, we are presenting niPGT-A testing of 3400 embryos from 624 females. Subjects included in the study were both women and men who were then sub-grouped based on four major reasons viz. 1) advanced maternal age, 2) more than three IVF failures, 3) low sperm count and 4) genetic indications. Culture medium in which an embryo had grown from day 3 until the day of last day of culture (Day5) was collected. After a blastocyst was removed from its culture dish,15-20 µL of its SCM was collected in PCR tubes. Total 3400 SCM were collected from 624 females undergoing IVF treatment. These collected SCMs were subjected to whole genome amplification using a picoplex WGA kit. Amplified product was then taken up for customized NGS based chromosomal aneuploidy analysis where a lower mark of 10MB of deletion duplication was taken for reporting and a mosacisim of minimum 25% was reported. NGS data revealed that 816 (24%) embryos had segmental or whole chromosomal aneuploidy. Whereas, 484(14.2%) embryos showed low mosaicism. 74 (2.17%) samples exhibits low DNA and were not be able to analyzed. Euploid reported embryos were evaluated for morphology score and AI score and are consistent with the invasive PGT-A testing. However, embryos with low mosaic is of concern but progress of niPGT-A will overcome challenges and become a choice of method to detect chromosomal abnormalities in developing embryos.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Session I

PB2087 Novel application of NIPT to indirectly diagnose paternal sex chromosome aneuploidy via parent-of-origin analysis

Authors:

A. Zalesny[1], A. Boyd[2], B. Polendo[1], S. Whitehead[1], R. Higgins[2], M. D. Maxwell[1]; [1]Univ. Hlth.System, San Antonio, TX, [2]Univ. of Texas Hlth.Sci. Ctr. at San Antonio, San Antonio, TX

In 2011, non-invasive prenatal testing (NIPT) was introduced as the first maternal screen for fetal sex chromosome abnormality (SCA). It has since been shown to incidentally detect maternal SCA, microduplications, and malignancy. Turner syndrome (TS) is the most common chromosomal abnormality in first-trimester miscarriage and often results from sporadic paternal nondisjunction, making recurrence unlikely. Possible explanations for recurrence include undiagnosed parental SCA or gonadal mosaicism. Here we present two consecutive 45,X intrauterine fetal demises (IUFDs) in which NIPT's maternal cell-free DNA background was applied to determine X chromosome inheritance and thereby the monosomy's parent of origin.

A G4P1 28-year-old female and her 28-year-old partner had one first-trimester miscarriage at age 24 followed by a term pregnancy, delivering a live-born male. At ages 26 and 27, they had two second-trimester IUFDs, both with cystic hygromas, monosomy X risk on single nucleotide polymorphism (SNP)-based NIPT, and 45,X karyotypes on amniocentesis. Given her two consecutive pregnancies associated with SCA, a maternal karyotype was performed on peripheral leukocytes which revealed 46,XX(20). To further investigate potential 45,X mosaicism, the patient consented to comparative analyses of maternal and fetoplacental SNP profiles from each IUFD to determine X chromosome inheritance, which confirmed both to be maternally inherited and suggested mosaic paternal SCA. A paternal karyotype was then performed on peripheral leukocytes, which revealed 46,XY(20), and supplemental X-/Y-centromere FISH of 200 cells was also normal, suggesting paternal 45,X gonadal mosaicism. Paternal fibroblast and semen analyses have not been performed due to logistic limitations.

While somatic mosaicism is unconfirmed, paternal 45,X gonadal mosaicism is suspected within the context of consecutive fetal TS, SNP-based NIPT comparisons, and Y-chromosome constitution of the couple's only viable pregnancy. The data support paternal 45,X gonadal mosaicism as another, possibly more likely differential diagnosis for their first fetal loss and consecutive fetal TS than recurrent sporadic meiotic error. To our knowledge, this is the first case of NIPT applied to indirectly diagnose paternal SCA via parent-of-origin analysis.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Session II

PB2088 Optical genome mapping for detection of chromosomal aberrations in prenatal diagnosis

Authors:

Q. Zhang, Y. Wang, Z. Xu, P. Hu; Women's Hosp. of Nanjing Med. Univ., Nanjing Maternity and Child Hlth.Care Hosp., Nanjing, China

Introduction: Chromosomal aberrations are the most important etiological factors for birth defects. Optical genome mapping is a novel cytogenetic tool for detecting a broad range of chromosomal aberrations in a single assay, but relevant clinical feasibility studies of OGM in prenatal diagnosis are limited. **Material and methods:** We retrospectively performed optical genome mapping analysis of amniotic fluid samples from 34 fetuses with various clinical indications and chromosomal aberrations detected through standard-of-care technologies, including karyotyping, fluorescence in situ hybridization, and/or chromosomal microarray analysis. **Results:** In total, we analyzed 46 chromosomal aberrations from 34 amniotic fluid samples, including five aneuploidies, 10 large copy number variations, 27 microdeletions/microduplications, two translocations, one isochromosome, and one region of homozygosity. Overall, 45 chromosomal aberrations could be confirmed by our customized analysis strategy. Optical genome mapping reached 97.8% concordant clinical diagnosis with standard-of-care methods for all chromosomal aberrations in a blinded fashion. Compared with the widely used chromosomal microarray analysis, optical genome mapping additionally determined the relative orientation and position of repetitive segments for seven cases with duplications or triplications. The additional information provided by optical genome mapping will be conducive to characterizing complex chromosomal rearrangements and allowing us to propose mechanisms to explain rearrangements and predict the genetic recurrence risk. **Conclusions:** Our study highlights that optical genome mapping can provide comprehensive and accurate information on chromosomal aberrations in a single test, suggesting that optical genome mapping has the potential to become a promising cytogenetic tool for prenatal diagnosis.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Session III

PB2089 Optical genome mapping versus chromosomal microarray analysis and karyotyping in prenatal diagnosis

Authors:

P. Hu, Y. Xu, Q. Zhang, Y. Wang, Z. Xu; Women's Hosp. of Nanjing Med. Univ., Nanjing Maternity and Child Hlth.Care Hosp., Nanjing, China

Background: Emerging studies demonstrate that optical genome mapping (OGM) is able to detect nearly all classes of structural variations in a single assay with high concordance with standard of-care methods. However, the application and evaluation of OGM in prenatal diagnosis remains limited.**Objectives:**This study aimed to evaluate the feasibility, efficacy, and incremental yield of OGM compared with chromosomal microarray analysis (CMA) and karyotyping for routine prenatal diagnosis.**Study design:**In this prospective study, a total of 200 singleton fetuses with ultrasound abnormalities were recruited. OGM and CMA plus karyotyping were performed in each sample in parallel. Triploidies, aneuploidies, copy number variations and microscopic balanced translocations and inversions were detected and analyzed in blinded fashion. Small (< 50 kb) insertions, deletions, duplications and submicroscopic translocations and inversions that beyond the detection range of CMA and karyotyping were validated by long-rang PCR plus Sanger sequencing, multiplex ligation-dependent probe amplification or fluorescence in situ hybridization.**Results:**Overall, 41 (20.5%, 41/200) cases with chromosome aberrations were identified by OGM, including one with triploidy, 22 with aneuploidies, 15 with pathogenic or likely pathogenic copy number variations and three with balanced translocations. Compared with CMA plus karyotyping, OGM provided an additional detection of one case with pathogenic intragenic duplication and two cases with cryptic balanced translocations (one translocation resulted in the disruption of OMIM gene associated with dominant disorders), but failed to identify one case with pericentric inversion. Additionally, OGM added information of orientation and location for eight duplication segments and identified eight (4.0%, 8/200) cases with repeat contraction disorders.**Conclusions:**Compared with CMA and karyotyping, OGM provided a more comprehensive detection of structural variations at a higher resolution in prenatal diagnosis. In addition, OGM revealed the location and orientation of duplication segments, refined breakpoints of structural variations and identified specific repeat contraction disorders. Our results suggest that OGM has the potential to be an alternative technology to CMA and karyotyping for prenatal evaluation of fetuses with ultrasound abnormalities.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Session I

PB2090 Optimization of preimplantation genetic testing workflow for a higher percentage of euploid embryos

Authors:

Y. Goikhman, S. Avdeichik, I. Mirzalimova, A. Kullyev; Israeli Med. Ctr., Tashkent, Uzbekistan

Upon establishing preimplantation genetic testing for aneuploidy (PGT-A) laboratory in a new location in Tashkent, Uzbekistan, AB-PGT (AB Vector) kit was chosen as the PGT-A sample preparation method.
With this setup, 62.8% of embryos were diagnosed as euploid, consistent with published data for euploidy rate among younger patients (the average patient's age is 31.9 years).
As the number of samples grew we switched to the EmbryoMap PGT-A kit to streamline the sample processing. EmbryoMap sample processing reduces hands-on time in half. It also reduces the chance of human errors in sample processing and results interpretation due to the following measures: The order of samples with unique Sample IDs is locked at the beginning of processing in the eMap Laboratory Planner. Primers for the library come pre-mixed for every sample. Data analysis and reporting are performed using eMap software with automated calling for whole chromosome aneuploidy, segmental changes, and mosaic changes.
The euploidy rate with the EmbryoMap is 48.2%, 23% lower relative to AB Vector. The percentage of embryos with mosaic and segmental aneuploidy is 60% higher with EmbryoMap (22.3% with EmbryoMap compared to 13.9% AB Vector).
IVF clinic with clinicians performing ovarian stimulation, embryologists performing a biopsy, IVF laboratory setting, genetic laboratory personnel, and clinical geneticist reading and interpreting results remained the same for all the samples processed. All samples were sequenced on the Illumina MiSeq Sequencing System. The lower euploidy rate could be due to false positive results. To improve calling accuracy we are cataloging artifacts specific to EmbryoMap. To reveal false positive results we established a practice of routinely re-testing the inner cell mass (ICM) of embryos diagnosed as abnormal. Any discrepancies found during re-testing are used to improve the accuracy of PGT-A.
Conclusions: The sample processing kit in PGT-A impacts both probability of human errors and the percentage of embryos recommended for transfer. The percentage of euploid embryos among younger patients is a useful reference point to compare against published data. Results interpretation should be tailored to each PGT-A workflow. Diagnosis for embryos not recommended for transfer should be routinely reconfirmed by analyzing ICM, as part of PGT-A quality control.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Session II

PB2091 High-Resolution and Non-Invasive Prenatal Exome Screening

Authors:

M. Duyzend[1,2], H. Brand[1,2], C. Whelan[2], J. Lemanski[1], M. Salani[1], S. Hao[1], I. Wong[1], E. Valkanas[1,2], C. Cusick[2], C. Genetti[3], L. Dobson[4], C. Studwell[4], K. Gianforcaro[4], L. Wilkins-Haug[4], S. Guseh[4], B. Currall[1], K. Gray[4], M. Talkowski[1,2]; [1]Massachusetts Gen. Hosp., Boston, MA, [2]Broad Inst., Cambridge, MA, [3]Boston Children's Hosp., Boston, MA, [4]Brigham and Women's Hosp., Boston, MA

Non-invasive prenatal screening (NIPS) involves the assessment of circulating cell-free fetal DNA extracted from maternal plasma. While screening for chromosomal aneuploidies is now routine in clinical practice and advances have enabled detection of some targeted single gene conditions, current NIPS approaches capture only a small fraction of genetic variants relevant to prenatal diagnosis. Building on earlier efforts that relied on deep sequencing of cell-free and parental DNA, our studies demonstrate the feasibility of a high-resolution non-invasive prenatal screen that can capture most pathogenic coding variation (single nucleotide variant (SNV), indel, copy number variant (CNV)), including *de novo* variation, from a maternal blood draw alone. To demonstrate the feasibility of our method, we performed a study on 51 pregnancies with gestational ages ranging from 9 to 40 weeks and benchmarked our method against 11 of these samples with gold-standard germline exome sequencing from cord blood or amniocentesis. We prepared cell-free DNA (cfDNA) libraries from maternal plasma, enriched for exonic regions, and generated an average 211x exome-wide short-read sequencing coverage. We detected small variants using a customized pipeline built to filter out candidate variant sites generated by sequencing or mapping artifacts and genotyped sites using a Bayesian Gaussian mixture model which simultaneously estimates fetal fraction and considers features such as the DNA fragment sizes supporting each allele. We identified CNVs using read-depth techniques. Our methods demonstrate a high median sensitivity/precision for SNV (92.9% / 93.3%) and indel (77% / 71.6%) genotyping from cfDNA. Prior to filtering, median sensitivity is 99.6% for SNVs and 94% for indels, indicating that most coding variants are captured, but opportunities remain to optimize filtering and genotyping. We assessed utility across 14/51 pregnancies referred for diagnostic genetic testing and discovered all clinically identified variants (n=4), including a likely pathogenic splice variant in *COL2A1* (associated with Stickler syndrome) and a 4MB terminal deletion on chr7 (in a fetus with multiple congenital anomalies). Because maternal cfDNA is also sequenced, our approach has added value as a maternal carrier screen with high median sensitivity / precision for maternal SNVs (98.4% / 95.9%) and indels (86.2% / 81.9%), and we found at least one reportable variant in over half of pregnant individuals assessed. This method represents a substantial increase in the resolution of NIPS techniques and suggests that the complete non-invasive prenatal exome is accessible at scale.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Session III

PB2092 Performance of a genome-wide association study of neural tube defects allows an unbiased search for genetic contribution.

Authors:

F. Pangilinan[1], J. L. Mills[2], A. M. Molloy[3], L. C. Brody[1]; [1]Natl. Human Genome Res. Inst., NIH, Bethesda, MD, [2]Eunice Kennedy Shriver Natl. Inst. of Child Hlth.and Human Dev., NIH, Bethesda, MD, [3]Sch. of Med., Trinity Coll., Dublin, Ireland

Neural tube defects (NTDs) occur during if the neural tube, which gives rise to the brain and spinal cord, fails to close completely in the first month post fertilization. NTDs are among the most common major birth defects, causing severe morbidity and mortality. The etiology involves genetic and environmental components. Periconceptual folic acid supplementation has been shown to prevent 50-70% of NTDs, suggesting that susceptible genetic factors negatively affect folate related processes at the time of neural tube closure.

To date, researchers have largely used a candidate gene approach to uncover the genetic component of NTDs, where known developmental genes or known folate pathway genes have been tested for association with increased risk. This approach has been less successful than anticipated, with many reported associations not found to be replicable. Although valuable, candidate gene studies are limited because the only genes selected are those known a priori by researchers.

A genome-wide association study (GWAS) has the advantage of being an unbiased by assumptions. For this study, nearly two million genetic markers will be genotyped by the Center for Inherited Disease Research using an array that is the equivalent of the Global Diversity Array (Illumina). Designed to capture common and low frequency variants in global populations, as well as incorporate functional variants, this array will be used in a genome-wide screen of association with NTDs. An exceptional number of participants is required to reach a sufficient sample size to correct for multiple comparisons in the primary GWAS and to replicate positive signals.

The International Consortium on the Genetics of Neural Tube Defects is a collaborative effort to perform an NTD GWAS. Study participants have been drawn from the National Birth Defects Prevention Study (USA) and the Baby Biobank (UK), and will also involve NTD-specific cohorts, including a large, established cohort from the Republic of Ireland. Additional participants originate from cohorts collected in New York, Texas, Chile, the Netherlands, and Malaysia. After pooling, more than 1400 NTD cases and their parents from the US and Europe will be available for an initial GWAS of NTDs. Case-control and family-based tests of association will be used to evaluate genetic risk of offspring developing an NTD or mothers developing an NTD affected pregnancy. Over 1200 NTD cases and their parents are available for replication testing of positive signals. This study is expected to reveal new sources of genetic contribution to NTDs, thereby illuminating its etiology and potentially its treatment and prevention.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Session I

PB2093 Perinatal detection of HPV infection using a wearable device: Prevention of genomic instability using DNA editing.

Authors:

E. McGhee[1], A. Bahmani[2], J. LIllard[3]; [1]Martin Luther king Jr. Community Hlth.Hosp., Los Angeles, CA, [2]Stanford Univ., Palo Alto, CA, [3]Morehouse Sch. of Med., Atlanta, GA

The observation of infants showing signs from receiving perinatal transfection of human papillomavirus (HPV)-induced genomic instability has been demonstrated in several small studies. The direct contact of HPV infection carried by a pregnant woman can be perinatally transferred to the unborn baby. Wearable devices may be important for detecting HPV Infection in pregnant women and may be helpful in halting the perinatal transmission of the virus by this detection. However, HPV DNA has been detected in umbilical cord blood, and HPV can spread to amniotic cells that are then ingested by the fetus. Cells involving p53 and Rb, gate keeper proteins protect against DNA damage, if not degraded by E6/E7 HPV complex. This research will investigate the possibility for Mendelian mutations, associated with increased chromosomal instability through Robertsonian aberrations. Robertsonian translocations are common constitutional abnormalities in humans but are considered rare occurrences in cancer cells. Research is sparse surrounding Robertsonian translocations in viral infected cells; therefore, we examined *in vitro*, whether HPV-E6/E7 transfection will facilitate the formation of Robertsonian aberrations. We used single nucleotide polymorphism (SNP) spectral karyotyping (SKY) and array-comparative genomic hybridization (array-CGH) to score Robertsonian aberrations in cells, we analyzed metaphases to determine the types of Robertsonian translocations present in cells. The results showed non-homologous Robertsonian translocations in 100% of metaphases scored and homologous acrocentric rearrangements-isochromosome types in 75% of metaphases scored. We studied two types of Robertsonian translocations: der(13;14), the non-homologous type, and i(13), the homologous isochromosomes type. The results showed that HPV E6/E7 oncoproteins induce genomic instability, causing Robertsonian translocations on chromosomes 13 and 14. This research may provide a reliable detection method for future research involving HPV infection using wearable devises.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Session II

PB2094 Pleiotropic effect of de novo variants in structural birth defects and neurodevelopment disorder

Authors:

W. Gao[1,2], H. Jake[3,1], Q. Lu[1], Z. Guojie[2], C. Shu[1], Y. Shen[4], W. Chung[5]; [1]Dept. of Pediatrics, Columbia Univ. Irving Med. Ctr., New York, NY, [2]Dept. of System Biology, Columbia Univ. Irving Med. Ctr., New York, NY, [3]Dept. of Systems Biology, Columbia Univ. Irving Med. Ctr., New York, NY, [4]Columbia Univ., New York, NY, [5]Dept. of Pediatrics, Columbia Univ., New York, NY

Structural birth defects like congenital diaphragmatic hernia (CDH), orofacial cleft and sex development disorders have heterogenous genetic causes. Previous studies suggest that hundreds of genes can contribute to risk of these conditions through de novo variants. Many genes involved in core developmental pathways may have pleiotropic effect in contributing to risk of multiple birth defects and neurodevelopmental disorders. The statistical power of identifying new risk genes is typical low in studies of birth defects without large sample size. In this work, we aim to estimate the degree of pleiotropic effect of de novo variants across these conditions and to identify pleiotropic risk genes that have a new role in birth defects. We obtained genome data of a comprehensive set of birth defects from the NIH Gabriella Miller Kids First (GMKF) Pediatric Research Program, including 1078 CDH trios, 204 orofacial cleft birth defect trios, 218 Kidney and Urinary Tract trios, 107 nonsyndromic craniosynostosis, 44 sex development disorders trios, 185 Esophageal Atresia and Tracheoesophageal Fistulas (EA/TEF) trios. We called de novo coding variants from the GMKF affected individuals. We also obtained published de novo coding variants from studies on congenital heart disease (The Pediatric Cardiac Genetics Consortium, N~3000) and NDD (N~30,000). Consistent with previous studies, there are significant enrichment of de novo protein truncating variants in all GMKF conditions, especially among constrained genes (gnomAD pLI≥0.5). The enrichment of de novo missense variants is marginal; its significance depends on sample size, as expected. To estimate pleiotropic effect, we applied a simple Poisson test of damaging variants (PTV and missense variants with CADD phred-scale score≥25), in the combined data of all the conditions, and identified 219 genes with exome-wide significance (p-value<2.5e-6). Among these genes, 90 have at least one damaging variants in at least two conditions, including 85 estimated NDD genes and five putative novel risk genes. A notable gene is *ZFHX3* (*p*=7.8e-9), with damaging variants in individuals with NDD, CHD, or CDH, and a previously reported individual with orofacial cleft (not included in our data). Finally, we focused on CDH. Overall, the population attributable risk (PAR) of de novo coding variants to CDH is estimated to be 27%, including 7.5% from established CDH genes by human or mouse studies and an additional 6.5% from known NDD and CHD risk genes, most of which have not been implicated with CDH. These results provide new insight of finding new risk genes for rare birth defects through methods that leverage pleiotropic effect.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Session III

PB2095 Preimplantation genetic testing data from 129,479 IVF embryos reveals the landscape of haplo- versus triplo-sensitivity prior to blastocyst formation

Authors:

S. Carioscia[1], A. Biddanda[1], I. Vogel[2], E. Hoffman[2], R. McCoy[1]; [1]Johns Hopkins Univ., Baltimore, MD, [2]Univ. of Copenhagen, Copenhagen, Denmark

Less than half of all human conceptions survive to live birth. Extra or missing chromosomes (aneuploidies) are the primary cause of pregnancy losses, many of which occur during the preimplantation stages. Data from preimplantation genetic testing for aneuploidy (PGT-A) from trophectoderm biopsies of blastocyst-stage in vitro fertilized (IVF) embryos offers a unique view into the origins of aneuploidy and the landscape of dosage sensitivity during the first five days after fertilization.

Here we report detailed patterns of chromosome abnormalities detected in biopsies of 129,479 day-5 blastocyst-stage IVF embryos from 20,900 sets of patients. Using genome-wide array-based genotyping of embryo biopsies, as well as DNA from parents, we infer transmission of individual parental homologs and assign parental origins of aneuploidies. The scale of the data offers unique insight into the origins of rare abnormalities in genome-wide ploidy. This includes the observation that triploidy nearly exclusively involves an extra copy of the maternal genome (1365 cases of maternal triploidy vs. 66 cases of paternal triploidy) and that haploidy/genome-wide uniparental isodisomy (GW-isoUPD) nearly exclusively involves the sole presence of a maternal genome (224 cases of maternal haploidy/GW-isoUPD vs. 35 cases of paternal haploidy/GW-isoUPD).

Gains and losses of all individual chromosomes were observed among the blastocyst-stage embryos. Maternal meiotic-origin trisomies were strongly enriched on chromosomes 15, 16, 21, and 22, consistent with previous literature. Because the predominant mechanisms of maternal meiotic error produce equal ratios of monosomies and trisomies (via reciprocal gain and loss in the egg and polar body, or vice versa), imbalance in this ratio offers insight into lethality of gains versus losses during cleavage-stage embryonic development. While all chromosomes were enriched for trisomies compared to monosomies, the magnitude of enrichment was greater for larger chromosomes with more protein-coding genes (Pearson $r = 0.683$, $p = 4 \times 10\text{-}4$). The enrichment was strongest on chromosomes 1 (89.3% trisomies; binomial test, $p < 1 \times 10\text{-}10$) and 9 (87.6% trisomies; binomial test, $p < 1 \times 10\text{-}10$) and weakest on chromosomes 21 (56.4% trisomies; binomial test, $p = 1.3 \times 10\text{-}26$) and 22 (56.2% trisomies; binomial test, $p = 6.8 \times 10\text{-}40$). These observations support a model whereby the cumulative burden of dosage effects of numerous genes, as opposed to few genes of large effect, drives embryonic mortality in response to aneuploidy. Our results offer insight into the strong and recurrent forces of purifying selection that shape the first days of human development.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Session I

PB2096 Prenatal CMA using exon-level coverage for 'disease gene' CNV interrogations in the era of non-invasive prenatal testing

Authors:

W. Bi[1,2], S. Anderson[2], R. Song[2], S. Cheung[1], A. Breman[1], J. Lupski[1], A. Beaudet[1], P. Ward[1,2], I. Van Den Veyver[1,3], C. Bacino[1,2], N. Owen[1,2], J. Smith[1], C. Settler[2], R. Rigobello[2], M. Chau[1,2]; [1]Dept. of Molecular and Human Genetics, Baylor Coll. of Med., Houston, TX, [2]Baylor Genetics, Houston, TX, [3]Dept. of Obstetrics and Gynecology, Baylor Coll. of Med., Houston, TX

Background: Non-invasive prenatal testing (NIPT) with cell-free DNA is widely accepted as standard of care to screen for common aneuploidies, and some platforms also include screening for microdeletions/duplications. This shift in clinical practice affects the use of chromosomal microarray analysis (CMA) in the prenatal setting. In addition, microarrays with exon-by-exon coverage for disease genes enable detection of intragenic copy number variants (CNVs) leading to increased sensitivity to postnatal diagnosis but such arrays are rarely used in prenatal CMA. This study investigated the role of prenatal CMA in the era of NIPT based on a single laboratory's experience.

Methods: We retrospectively reviewed the results of all the amniotic fluid or chorionic villus samples that were analyzed by CMA using custom-designed Agilent arrays during last 11 years at Baylor Genetics. For 80% of samples, CMA was performed using arrays that include exon-by-exon coverage for >1,700 genes. Parental samples were received concurrently to evaluate maternal cell contamination and to facilitate data interpretation.

Results: The overall CMA detection rate of clinically significant findings was 10.4% and the diagnostic rate for pregnancies with abnormal ultrasound findings was 13.7%. The diagnostic rate was highest (32.1%) for cases with abnormal or atypical NIPT results as primary indications. Among those, the most frequent NIPT findings were increased risk or positive result for autosomal aneuploidy with a diagnostic rate of 48.3%, followed by increased risk or positive result for sex chromosome aneuploidy, microdeletion/duplication, and inconclusive or nonreportable findings. Of the cases with increased risk for microdeletion/duplication, CMA confirmed the CNVs in 13% of cases including cases with increased risk of 22q11.2 deletion and cases with supernumerary/ring chromosomes of >14 Mb in size. Due to the increased probe coverage in genes for established rare disease traits, CMA detected clinically significant exonic CNVs involving nine different genes. The smallest one is a *de novo* 5.6 Kb deletion of the *ZIC2* gene in a fetus with multiple holoprosencephaly.

Conclusions: Prenatal CMA remains essential for the detection of microdeletions/duplications because NIPT is limited in screening for submicroscopic copy number changes. We also show that increased probe coverage of disease genes on prenatal microarrays enables detection of single gene copy number changes. Finally, this confirmation data demonstrates the importance of a diagnostic test such as CMA in follow up to NIPT results that are positive or increased risk for microdeletions/duplications.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Session II

PB2097 Products of Conception analysis by chromosomal microarray: a single center experience.

Authors:

J. Smith, N. Owen, W. Bi; Baylor Coll. of Med., Baylor Genetics, Houston, TX

Background: Missed abortions and fetal demises are often caused by chromosome aberrations detectable by conventional cytogenetics. However, chromosome analysis of products of conception (POC) may be unsuccessful or uninformative due to maternal cell contamination (MCC) of the sample and reduced viability of cells in culture. Chromosomal microarray (CMA) may circumvent such issues, as it does not require culturing, and MCC studies may be performed on extracted DNA prior to running CMA. **Methods:** Data from this clinical laboratory's offering of POC CMA over a 5-year period were reviewed specifically for aneuploidy/triploidy and compared to conventional cytogenetic results for the same period. CMA was performed on a custom designed Agilent 400k array with SNP probes. MCC was evaluated on CMA cases by comparative analysis of maternal and fetal DNA using multiple unlinked polymorphic markers when a maternal specimen was received. **Results:** Reportable data were obtained on 71% and 82% of POCs submitted for CMA and conventional cytogenetics, respectively. Unreportable CMAs were due to insufficient DNA quantity/quality or overwhelming MCC. Notably, 45% of cytogenetic studies were reported as 46,XX with no way to distinguish which were maternal vs. fetal in origin. Data review showing significantly more 46,XY results reported by CMA than by cytogenetics (31% vs. 12.6%) is evidence that a significant number of 46,XX results are maternal in origin. Discrepancies between CMA results (various aneuploidies and triploidy) and cytogenetics (46,XX) were noted in several cases. The use of MCC studies and avoidance of maternal cell overgrowth in culture by analysis of direct preparations gives CMA an advantage over conventional chromosome analysis. Among more than 100 CMA cases with abnormalities/incidental findings were 9 triploidies and 73 single chromosome aneuploidies including Monosomy X (18), trisomy 21 (15), trisomy 22 (9), trisomy 13 (8), trisomy 18 (7), trisomy 9 (4), trisomy 15 (4), trisomy 16 (2) and one case each of XXX, XXY and trisomies 4, 8, 14, and 20. Additional anomalies detected were 1 case of monosomy 3, 1 case of mosaic +X,+19, 8 other mosaic cases and 2 double aneuploidies (1 trisomy 15 and 16 and 1 trisomy 5 and 20). **Conclusions:** These data support the use of MCC studies and avoidance of maternal cell overgrowth in culture by analysis of direct preparations giving CMA an advantage over conventional chromosome analysis. Our CMA data are consistent with the literature in that autosomal trisomies as a group are the most common class of anomalies in pregnancy losses followed by monosomy X and triploidy; it does not support the published frequency of trisomy 16.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Session III

PB2098 Quantitative trait loci mapping: Identifying inbred strain-specific loci that modify cleft secondary palate expression in*Prdm16*^{csp1}mouse mutants.

Authors:

D. Melwani, M. Elibiary, B. Kimberly, E. Leslie, B. Bjork; Midwestern Univ., Lombard, IL

Cleft secondary palate (CP) is among the most common congenital craniofacial anomalies with complex inheritance. *Prdm16* is a zinc finger transcription factor whose loss in mouse embryogenesis results in CP resembling Pierre Robin complex. *Cleft secondary palate 1* mutant embryos *(Prdm16*^{csp1}*)* on a congenic FVB/NJ strain background (*csp1-FVB*) exhibit fully penetrant CP, whereas *csp1* mutants on a congenic C57BL/6J strain background (*csp1-B6*) do not exhibit CP. We hypothesize that quantitative trait loci (QTLs) that contribute to this phenotypic variation are present in these strains' genomes. To this end, we performed a F2 genetic modifier screen with genome-wide SNP analysis using the Mini Mouse Universal Genotyping Array (MiniMUGA) and Giga Mouse Universal Genotyping Array (GigaMUGA). *csp1-B6* carrier males were outcrossed to inbred FVB females to produce F1 B6.FVB-*csp1* carriers. Male and female F1 carriers were intercrossed to generate F2 homozygous *csp1* mutant pups on a mixed B6xFVB background. Resulting F2 mutants were placed in 3 distinct phenotypic classes (wide CP, narrow CP, and no CP), and whole-genomeSNP genotyping was carried out. QTL.gCIMapping.GUI software was utilized to identify putative QTLsthrough genome-wide composite interval mapping (GCIM) and likelihood of the odds (LOD) assessment. We also used r/QTL2 as an alternative method. We identified 6 significant QTL regions on 5 chromosomesassociated with variation in CP expression in *csp1* mutants with differing contributions of FVB and B6 alleles. Next, we performed secondary QTL analyses by adding high-resolution GigaMUGA SNPs on these 5 associated chromosomes to refine the QTL intervals. We are currently evaluating positional candidate genes for potential damaging sequence variants that may contribute to differences in CP phenotype between the *csp1* mutants on the FVB and B6 strains. In future studies we hope to employ existing mouse mutant alleles and/or CRISPR/Cas9 genome editing to introduce variant alleles to the FVB or B6 genomes to substantiate potential modifier loci identified in our screen. Our findings may contribute to a greater understanding of genetic factors that influence normal craniofacial development that may contribute to the complex etiology of cleft palate in humans.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Session I

PB2099 Rare Heterozygous DMRT1 Variants In Spermatogenic Failure And Primary Ovarian Insufficiency (POI) cases

Authors:

H. Madeen[1,2], A-G. Juchnewitsch[3], T. Marić[4], K. Lillepea[3], A. Dutta[3], A. Stendahl[1], M. Punab[5], K. Pomm[5], D. Moreno Mendoza[6], A. Lopes[7], J. Gonçalves[8], K. Almstrup[9], K. Aston[10], D. Ježek[4], A. Katušić Bojanac[4], M. Barbalić[4], M. Laan[3], D. Conrad[1,2]; [1]Oregon Hlth.and Sci. Univ., Beaverton, OR, [2]Oregon Natl. Primate Res. Ctr., Beaverton, OR, [3]Inst. of Biomedicine and Translational Med., Univ. of Tartu, Tartu, Estonia, [4]Univ. of Zagreb Sch. of Med., Zagreb, Croatia, [5]Andrology Clinic, Tartu Univ. Hosp., Tartu, Estonia, [6]Andrology Dept., Univ.t Autònoma de Barcelona, Barcelona, Spain, [7]Univ. of Porto, Porto, Portugal, [8]Departamento de Genética Humana, INSA, Lisbon, Portugal, [9]Copenhagen Univ. Hosp., Copenhagen, Denmark, [10]Univ of Utah, Salt Lake City, UT

Non-obstructive azoospermia (NOA) is one of the more severe forms of male infertility, however many cases remain unexplained. DMRT1 is a transcription factor that is important in regulating development and sex determination. Disease-causing variants with autosomal dominant inheritance in *DMRT1* have been linked to many human infertility phenotypes, including disorders of sexual development and NOA. The connection between *DMRT1* variation and primary ovarian insufficiency (POI) in humans has not been determined. This study aimed to understand the effect of rare heterozygous *DMRT1* mutations in patients with spermatogenic failure (SPGF) and POI, and the genetic burden of carrying these mutations. We utilized whole exome sequencing (WES) from the GEnetics of Male INfertility Initiative (GEMINI) and Estonian Andrology (ESTAND) cohorts; n=1,940 SPGF, n=105 POI, and n=644 normozoospermic controls. To analyze causative *DMRT1* variants, a computational pipeline was developed to filter variants and ACMG guidelines were followed to determine variant pathogenicity. Gene-based burden testing was performed using the GEMINI+ESTAND cases and population-based controls from publicly available data in gnomADv2.1. We identified an excess of rare potentially pathogenic *DMRT1* variants in the combined male infertility cases (n=14), compared to normozoospermic controls (n=0, Fisher's Exact test $p<0.05$). Burden testing with the larger set of population controls from gnomAD with damaging *DMRT1* variants (n=519/141,456) compared to GEMINI+ESTAND was significant ($p<0.05$). Nine variants, including two in POI cases, clustered in the DNA binding (DM domain) of the DMRT1 protein. Six variants clustered together in two hotspot amino acid positions, both of which were highly conserved across species and previously reported as functionally important in the DM domain. We identified three rare *DMRT1* variants in POI cases, a significant enrichment compared to gnomAD (p=7.26E-03). One POI case with a substitution in the DM domain was also detected in two NOA cases. Phenotypes observed in patients with DM domain variants included smaller testes, spermatogenic arrest, and germ cell neoplasia in situ (GCNIS). These findings increase the understanding of the role of rare heterozygous *DMRT1* variants in male infertility, revealing how missense substitutions in the DM domain could be a dominant cause of male infertility. The connection between *DMRT1* mutations and POI in humans has not yet been determined and the shared positions between NOA and POI cases is a further avenue for investigation on *DMRT1* and POI.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Session II

PB2100 Recurrent spontaneous abortion related to balanced translocation of chromosomes - A case report

Authors:

N. DUKUZE; Univ. of Rwanda, Kigali, Rwanda

ABSTRACT INTRODUCTION: Recurrent spontaneous abortion (RSA) is defined as three or more consecutive pregnancy losses before the 20th week of gestation. RSA is often idiopathic, but structural chromosomal abnormality is an important cause. An unbalanced karyotype in the conceptus of a couple when one partner has a structural chromosomal abnormality may result in failure to implant, miscarriage, or pregnancy of a fetus with an unbalanced karyotype. CASE PRESENTATION: We report a rare case of RSA associated with balanced translocation of chromosomes. a woman who had four spontaneous abortions, all pregnancy loss happened before 12 weeks of gestation, no other known chronic diseases reported to the family nor medications taken during pregnancy. The karyotype was 46.XX, t(13p,21p) The abnormal karyotype was not found in any other chromosomes. Further spectral karyotyping was performed to rule out the involvement of any other chromosomal aberrations present in the genome.The cytogenetic analysis of the husband revealed a normal karyotype 46.XY. CONCLUSION: Couples with more than three miscarriages should be referred to the genetist for chromosomal analysis for possible hereditary etiology and chromosomal abnormalities responsible for miscarriages to plan prenatal diagnostics and genetic counseling for subsequent pregnancies.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Session III

PB2101 Residual risk of clinically significant copy number variations in fetuses with nasal bone absence or hypoplasia after excluding non-invasive prenatal screening-detectable findings

Authors:

Y. Wang[1], P. Hu[2], Z. Xu[1]; [1]Women's Hosp. of Nanjing Med. Univ., Nanjing Maternity and Child Hlth.Care Hosp., Nanjing, China, [2]Women's Hosp. of Nanjing Med. Univ., Nanjing Maternity and Child Hlth.Care Hosp., Nanjing, China, Nanjing, China

Objective: To evaluate the residual risk of clinically significant copy number variations (CNVs) in fetuses with nasal bone (NB) absence or hypoplasia after excluding theoretically non-invasive prenatal screening (NIPS)-detectable abnormalities, and to assess their clinical outcomes. **Methods:** This prospective study encompassed 400 fetuses with NB absence or hypoplasia undergoing chromosomal microarray analysis (CMA) testing between 2015 and 2022. Clinically significant CMA findings were categorized into three subgroups, including three-NIPS-detectable (trisomies 21, 18 and 13), five-NIPS-detectable (trisomies 21, 18 and 13 and sex chromosome aneuploidies) and genome-wide NIPS-detectable. We calculated the theoretical residual risk following the exclusion of NIPS-detectable findings and compared it with the results of a published control cohort of low-risk pregnancies. We further evaluated their clinical outcomes. **Results:** The overall diagnostic yield in our cohort was 7.8% (31/400). The modeled residual risk of clinically significant CNVs ranged from 3.1% (three-chromosome NIPS) to 1.7% (genome-wide NIPS) in fetuses with isolated NB absence or hypoplasia and 9.7% (three-chromosome NIPS) to 6.7% (genome-wide NIPS) in fetuses with non-isolated NB absence or hypoplasia. The theoretical residual risks were significantly higher in all NIPS models when compared with the control cohort. The normal infant rate in fetuses with normal CMA results was 97.9% (323/330). **Conclusions:** The residual risk of clinically significant CNVs in fetuses with NB absence or hypoplasia following the exclusion of theoretically NIPS-detectable findings was higher than that in low-risk pregnancies.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Session I

PB2102 Screening over 600 candidate genes for disease-causing variants in male factor infertility.

Authors:

K. Lillepea[1], A-G. Juchnewitsch[1], L. Kasak[1], A. Valkna[1], A. Dutta[1], K. Pomm[2], O. Poolamets[2], M. Punab[1,2,3], M. Laan[1]; [1]Inst. of Biomedicine and Translational Med., Univ. of Tartu, Tartu, Estonia, [2]Andrology Ctr., Tartu Univ. Hosp., Tartu, Estonia, [3]Inst. of Clinical Med., Univ. of Tartu, Tartu, Estonia

Failure to conceive a child affects 5-10% of men, and in most cases, it is due to spermatogenic failure (SPGF, $<39\times10^6$ sperm/ejaculate). Currently, >50% of SPGF patients remain idiopathic[1]. Although genetic factors are considered as a key etiology behind SPGF, today's clinical guidelines include limited genetic tests. Exome sequencing (ES) has led to the discovery of hundreds of candidate genes and variants for male factor infertility[2]. However, their added value in increasing the diagnostic yield and improving patient management in routine clinical practice is to be determined.

We performed ES to screen 519 idiopathic SPGF patients recruited to the ESTonian ANDrology (ESTAND) cohort for disease-causing variants in over 600 previously established and recently proposed candidate genes. Subgroups of patients with different andrological profiles were compared: non-obstructive azoospermia (NOA, 186 men), oligozoospermia (182), and patients with idiopathic cryptorchidism and SPGF (151). Rare disease-causing variants (MAF<0.01) were filtered, interpreted, and prioritized using available data for clinical and functional consequences, matchmaking with literature cases, and in silico predictions of variant pathogenicity. The final pathogenicity assessment was based on the ACMG guidelines[3] and gathered extended phenotype and pedigree data.

Molecular diagnosis was identified for >10% of ESTAND cases, and there was no significant difference between the three andrological subgroups for the proportion of patients with genetic findings. This is an important outcome with broad clinical consequences. Most of previously published ES studies in male infertility have focused only on NOA[2], whereas the fraction of idiopathic patients presenting oligozoospermia or cryptorchidism with SPGF is manifold higher compared to men with no sperm[1]. Although the screened gene panel comprised over 600 genes, the identified (likely) pathogenic variants were distributed only to 40 genes, including recurrently affected loci and recurrently detected disease-causing variants in more than one patient in ESTAND or overlapping with findings in other cohorts.

Our data suggest that gene panel-based ES could be included in the genetic testing of SPGF patients as it will contribute to the personalized management of patients' reproductive and overall health.

References:

[1] Punab et al (2017) Hum Reprod 32:18-31

[2] Laan et al (2021) Br Med Bull 16:5-22

[3] Richards et al (2015) Genet Med 17:405-424

Funding: Estonian Research Council, grant PRG1021.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Session II

PB2103 Sex differences in placental gene expression in hypertensive versus normotensive pregnancies

Authors:

S. Plaisier[1], A. Abdulahad[2], T. Powell[2], L. Nordstrom[2], D. Coonrod[2], M. Wilson[1]; [1]Arizona State Univ., Tempe, AZ, [2]Valleywise Hlth.Med. Ctr., Phoenix, AZ

Preeclampsia is a pregnancy disorder characterized by gestational hypertension and proteinuria and is thought to be rooted in placental dysfunction. Many studies have reported fetal sex biases in the incidence of pregnancy complications including hypertension and preeclampsia, but the mechanisms underlying the relationship between fetal sex and pregnancy complications are poorly understood. Having previously demonstrated sex differences in placental gene expression in uncomplicated pregnancies, we sought to determine if gestational hypertension affects placental gene expression differently based on fetal sex. We performed RNA sequencing and sex-informed transcriptomics of placental tissue from full-term deliveries of ten hypertensive and ten normotensive pregnancies, each with half carrying a fetus assigned female at birth (XX) and half with assigned male at birth (XY). We found that XY male placentas had distinct gene expression profiles from XX female placentas and only XY males separated by hypertension status. XY male placentas showed differential gene expression of 540 genes between hypertensive and normotensive conditions, which were enriched for genes involved in metabolism, cell adhesion, angiogenesis, signaling and proliferation. In contrast, XX female placentas had only 303 differentially expressed genes between hypertensive and normotensive conditions, with little overlap to XY male placentas, and were enriched for spliceosome and immune functions. We found that the XY male placental hypertension gene signature had much higher correlation to published signatures of preeclampsia than XX female placentas. We used single cell deconvolution to show that specific immune cell types had different imputed cell fraction in XY male versus XX female hypertensive placentas, suggesting potential mechanisms by which sex differences in placental gene expression could lead to different health outcomes. In conclusion, our study demonstrates the importance of sex as a biological variable as there are large differences in the placental gene expression pattern of gestational hypertension based on the sex assigned at birth and that this may be due in part to different proportions of specific immune cell types in the placental tissue.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Session III

PB2104 Single cell RNA-seq processing pipeline and data integration method decisions impact significance of immune cell clusters in the uterus

Authors:

J. Garifallou[1], M. Simoni[2], S. Mani[2], M. Mainigi[2], S. M. Gordon[1]; [1]Children's Hosp. of Philadelphia, Philadelphia, PA, [2]Dept. of Obstetrics and Gynecology, Univ. of Pennsylvania, Philadelphia, PA

Introduction: The diverse landscape of single cell RNA-seq technologies that have been developed have led to an ever-increasing number of pipelines and tools used to analyze these complex datasets. Initial data processing pipelines to call cells, map reads to a reference genome, and collapse UMIs, have been previously shown to produce variations in datasets while typically maintaining similar cell yields and mapping quality. Resultant cell by gene matrices are further analyzed in other software packages such as Seurat (Hao, 2021) and include mechanisms to integrate diverse libraries. Efficiency of the integration method across the libraries and proper accounting of unwanted sources of data variance are essential to elucidate the true, experimental biological heterogeneity of the dataset.
Methods: Sorted CD45+ immune cells from 1) endometrial biopsies obtained during the window of implantation (pre-pregnancy) and 2) the maternal decidua from first trimester elective terminations of pregnancy were collected. Single cell RNA-seq libraries were prepared using the 10x single cell library preparation and sequenced on the Illumina NovaSeq6000. Data was demultiplexed using the cellranger v6.0 pipeline (10x genomics). Cells were called and reads were then mapped to the same GRCh38 genome reference using the following pipelines: cellranger v6.0, Kallisto-Bustools, STARsolo, and Alevin. Separately, downstream mapped read matrices were analyzed using the seurat anchor integration and the harmony methods while regressing out differences between tissue source (Endo vs. FT) and UMI count. Post processing data quality was assessed for each of the pipelines regarding the number of molecules, number of genes, number of mapped reads per cell. To assess the effect of the initial processing pipeline choice on cell clustering and cell annotation, cells were automatically annotated using singleR package with a published literature dataset of an appropriate tissue reference. Finally, significance of cell type proportions was assessed using the propellor R package.
Results: Differing numbers of total cells per library and per dataset (all libraries integrated) were observed. Differences in the proportion of cell types were observed depending on the analysis pipeline. Examination of these differences hopes to yield insights into the optimal analytical methodology for this cell-tissue type.
Conclusion: Each data set has unique biological complexities that should be considered during analysis. Optimization and testing of multiple methods allows for maximizing biological data yield and subsequent analysis on a per sample and per data set basis.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Session I

PB2105 Single-cell analysis of the human neonatal infant airway epithelium reveals age-related changes in cellular composition and transcriptional programming

Authors:

S. Bhattacharya[1], B. Bera[2], C. Elizabeth[1], K. Salka[3], A. L. Welham[1], H. Morizono[4], M. J. Gutierrez[5], J. Jaiswal[1], N. Gustavo[1]; [1]Children s Natl. Hosp., Washington DC, DC, [2]Childrens Natl. Hosp., Washington DC, DC, [3]The State Univ. of New York at Buffalo, Buffalo, NY, [4]Children's Natl. Hosp., Washington DC, DC, [5]Johns Hopkins Univ. Sch. of Med., Baltimore, MD

The heterogeneous composition of the human airway epithelium may allow plasticity during early development and explain age-related differences in structure and function. How airway epithelial cell (AEC) populations change from the naïve neonatal state to the mature adult state is not well understood. Defining the early developmental differences in the airway epithelium is critical to understand the pathogenesis of respiratory disorders that begin to originate during this critical period of life. Here, we used single cell RNA seq analyses to define age-related changes in cellular composition and transcriptional programming comparing differentiated human neonatal infant AECs and adult AECs. We performed single-cell RNA sequencing of approximately 9990 nasal AECs from three neonatal infant donors and three adults. Primary AECs were differentiated at air-liquid interface (ALI), dissociated into a single-cell suspension (10X) and, followed by sequencing (Illumina) . For each sample, CellRanger, was used to estimate the gene expression per cell from the raw sequence files, Seurat and DoubletFinder to remove low quality reads, scTransform for normalizing the reads and Seurat for clustering and visualization. To estimate the differential expression (DE) between infants and adults, we integrate both groups using Seurat. Gene set enrichment analysis (GSEA) was used for functional annotation. Using single-cell RNA-sequencing and clustering based on known epithelial marker genes, we identified significant cell heterogeneity in primary human AEC populations, with 7 major clusters (basal, mitotic basal, basal-KRT15, rare cells, deuterosome and ciliated). Age-dependent differences observed included an overabundance of basal epithelial cells with mitotic markers(e.g. MKI67) and basal cells expressing KRT15 in Newborn infants, whereas Goblet/Secretory cells had overabundance in adults. Pseudobulk differential expression (DE) between infants and adults identified cell proliferation pathways in infants and immune-related pathways in adults. At cellular level there is higher expression of cell proliferation in infants whereas adults have higher expression of immune genes. Single-cell analysis of the human neonatal infant and adult airway epithelium revealed age-related changes in cellular composition and transcriptional programming. Additional studies are needed to further elucidate the early developmental program of the human airway epithelium, its environmental modifiers during infancy, and the functional consequences for the pathogenesis of respiratory disorders that begin during early life.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Session II

PB2106 SPEN deficiency contributes to the development of cleft palate in humans and mice

Authors:

B-J. Kim, A. Hernandez-Garcia, D. Scott; Baylor Coll. of Med., Houston, TX

Normal 0 false false false EN-US KO X-NONE /* Style Definitions */ table.MsoNormalTable {mso-style-name:"Table Normal"; mso-tstyle-rowband-size:0; mso-tstyle-colband-size:0; mso-style-noshow:yes; mso-style-priority:99; mso-style-parent:""; mso-padding-alt:0in 5.4pt 0in 5.4pt; mso-para-margin-top:0in; mso-para-margin-right:0in; mso-para-margin-bottom:8.0pt; mso-para-margin-left:0in; line-height:107%; mso-pagination:widow-orphan; font-size:11.0pt; font-family:"Calibri",sans-serif; mso-ascii-font-family:Calibri; mso-ascii-theme-font:minor-latin; mso-hansi-font-family:Calibri; mso-hansi-theme-font:minor-latin;} Deletions of chromosome 1p36 are the most common telomeric deletions in humans. Deletion/phenotype mapping suggests the existence of two critical regions that are sufficient to cause most of phenotypes seen in 1p36 deletion syndrome. *SPEN* is located in the proximal critical region and encodes a transcriptional repressor. Pathogenic *SPEN* variants have been shown to cause a new neurodevelopmental disorder, Radio-Tartaglia syndrome, characterized by anomalies of the brain, palate, and/or heart. SPEN deficiency also causes defects in the brain and heart in mouse models. However, malformations of the palate have not been described in SPEN-deficient animal models. Here we report that *Spen* null mouse embryos have a cleft palate. The palatal shelves are normally formed in wild-type and *Spen* null embryos at E13.5. At E14.5, the palatal shelves are elevated and ready to fuse in the middle line in wild-type embryos. In contrast, the palatal shelves are not elevated in *Spen* null embryos, which leads to cleft palate at E15.5. We conclude that loss of SPEN function contributes to the development of anomalies in the palate, including cleft palate, and that SPEN is required for normal palatal development in humans and mice.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Session III

PB2107 *SYCP2* variants present with variable age-related progressive SPGF and male infertility

Authors:

R. Hvasta[1], C. Pombar[1], M. Malizio[1], M. Olszewska[2], A. Malcher[2], M. Kurpisz[2], S. Qureshi[3], M. Khan[3], K. Hwang[1], A. N. Yatsenko[1]; [1]Dept. of Obstetrics, Gynecology, and Reproductive Sci., Magee-Womens Res. Inst., Sch. of Med., Univ. of Pittsburgh, Pittsburgh, PA, [2]Inst. of Human Genetics, Polish Academy of Sci., Poznan, Poland, [3]Dept. of Molecular Biology and Genetics, Inst. of Basic Med. Sci., Khyber Med. Univ., Peshawar, Pakistan

Infertility is a rising public health crisis with poorly understood foundations, affecting ~17.5% of the global population. Male infertility accounts for nearly half of the global burden and often presents with varying degrees of severity. Spermatogenic failure (SPGF) encompasses the variable presentation of defects in spermatogenesis from oligozoospermia to non-obstructive azoospermia (NOA). Considering many cases are due to genetic variation in meiotic genes, it is essential to investigate the breadth of these genes' impact on spermatogenesis and their pathogenicity. One such gene, *SYCP2*, a member of the synaptonemal complex, essential for meiotic recombination, has been reported to cause autosomal dominant SPGF. Our analysis has identified a potential age-related progression of dominant SPGF associated with loss of function (LoF) *SYCP2* variants in 3 siblings and 6 unrelated infertile patients.

Within our cohort, familial DNA (father, mother, proband with NOA, sibling with NOA, and sibling with reduced spermatogenesis [20 mil/mL]) and DNA from one sporadic patient with severe oligozoospermia (SO) (<5 mil/mL) were sequenced by Whole Exome Solution V1 (SOPHiA Genetics) and variants were analyzed via Fabric Genomics and SOPHiA DDM software.

In the family, a novel, single heterozygous, LoF *SYCP2* variant co-segregated with SPGF in the 3 brothers. At evaluation, the two brothers with NOA were 32 and 35, and the brother with reduced spermatogenesis was 25. In the sporadic case, a novel CNV deletion was identified. The patient was found to be normozoospermic at 28; however, at 41, the patient was revaluated and diagnosed with SO.

In conjunction with findings from published reports, it appears that *SYCP2* positive patients under age 30 typically present with less severe forms of SPGF or even clinically normal, whereas older patients appear to progress to more detrimental forms of SPGF. These variants could be insidiously inherited by offspring from females and notably, younger male carriers who were fertile at the time of conception, discretely passing on deleterious mutations through generations. We suggest that age-related accumulation of meiotic defects causes progressive SPGF, and potentially causes other cellular defect comorbidities such as cancer. To test our hypothesis, we will generate transgenic mouse models and evaluate their genomic instability, homologous recombination capacity, and perform fertility assessments at different time points. If confirmed, these findings could provide invaluable insight into the delicate balance required for meiotic preservation and propose a mechanism for dominant progressive onset of SPGF.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Session I

PB2108 Targeted exome sequencing in fetuses with ultrasound findings: A powerful tool in prenatal diagnosis.

Authors:

C. Pangalos, V. Chini, V. Velissariou; INTERGENETICS, Athens, Greece

One of the biggest challenges in prenatal diagnosis are the limitations of existing ultrasonographic tools regarding the true fetal phenotype. Even at an advanced gestation age, ultrasound findings may not be clear in many cases. Thus, prenatal genetic testing of syndromes *in utero* is a one-way approach. Following the initial publication (Pangalos *et al*, 2016), which presented a first approach on prenatal targeted exome sequencing at a cohort of fetuses with ultrasound anomalies and normal microarray results, this paper presents a retrospective analysis of a greater number of cases, in order to assess the potential diagnostic yield of this approach. In the period 2016-2022, 113 pregnancies with fetal ultrasound anomalies with normal prenatal microarray results were referred for additional molecular testing following genetic counseling. Only cases that underwent invasive prenatal sampling (chorionic villi sampling or amniocentesis) were included in this cohort. Fetal anomalies from this cohort were suspected by routine ultrasound scanning, mostly in the setting of second-trimester ultrasound screening. Twin pregnancies, where both fetuses had different anomalies and were both sampled, were counted as two separate cases. Targeted analysis through next-generation sequencing (NGS) of 874 genes associated to syndromes with fetal ultrasound findings was performed. Reporting and evaluating of findings followed the guidelines of ACMG (Monaghan *et al*, 2020). The advantage of this approach is that only pathogenic or likely pathogenic variants were reported avoiding complications within prenatal genetic counseling. In 34 out of 113 cases (30%), clinically relevant to the referred ultrasound findings, were detected. This agrees with the pooled incremental yield of exome sequencing (31%) reported (Mellis *et al*, 2022), while it seems to be 4-fold higher than the diagnostic yield of microarray analysis alone in fetuses with abnormal ultrasound scans (7%) (Callaway *et al*, 2013). It should be noted that a number of such cases result from parental inheritance, with parents being at risk for future pregnancies. By this approach, both diagnosis and prevention can be provided. Considering the aforementioned, prenatal targeted exome sequencing, following a normal chromosomal microarray, would notably increase the diagnostic yield in fetuses with ultrasound anomalies by ~30% and would allow early diagnosis of a genetic disorder, irrespective of the limited fetal phenotype.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Session II

PB2109 † Temporal changes in gene expression and chromatin accessibility in the mouse gut identify candidate postnatal modifiers of Hirschsprung disease (HSCR)

Authors:

H. Chubaryov[1], W. Hou[2], A. Chakravarti[3]; [1]NYU, New York, NY, [2]Columbia Univ., New York, NY, [3]NYU Sch. of Med., New York, NY

Introduction: The enteric nervous system (ENS) is established during embryonic development via the migration, differentiation, and proliferation of enteric neuronal crest cells (ENCCs) into the developing gut. Hirschsprung disease (HSCR) arises from developmental genetic defects in ENCC-derived neuronal and glial cells as well as cells in the supporting mesenchyme. Given that 30-50% of HSCR patients have persistent gastrointestinal symptoms after surgical repair, we hypothesized that postnatal genetic factors can modify HSCR. To unveil such genes, we studied gene expression changes, and the corresponding temporal *cis* regulatory elements that drive them, in wildtype mice across pre- and postnatal gut development. Methods: Bulk RNA-seq and ATAC-seq of the colon in the mouse were collected for E12.5, E14.5, P0, P07, P21, and 6-month-old mice. snRNA-seq was also conducted at P0 to identify the cell types contributing to these temporal changes. Results: Hierarchical clustering of bulk RNA-seq data revealed 10 distinct temporal patterns with many genes showing increased expression in postnatal development. The majority (13/20) of*temporally* varying known HSCR genes, including the genes most commonly mutated in HSCR, *Ret* and *Ednrb*, clustered together and decreased in expression in the postnatal gut. This cluster comprised genes enriched for RNA processing and chromatin modification (GO analysis). The remaining HSCR genes were distributed in clusters 2,7 and 9. Cluster 2, with *Eno3* and *Mmaa*, showed increased gene expression in the postnatal gut and was enriched for immune system and ribosomal biogenesis functions (GO). The ATAC-seq data similarly showed strong time dependent clustering by developmental stage (60% of total PCA variance). We observed that the *Ret* enhancer MCS+9.7, containing a common disease variant, is active during early development and corresponded to peak *Ret* gene expression but is absent later when *Ret* gene expression declined significantly. Similarly, for *Mmaa* we identified an enhancer that is first accessible at P0 and remained so throughout postnatal development, corresponding to its increased expression. Finally, snRNA-seq data were instructive of cell types with HSCR gene expression. *Eno3* and *Mmaa* are not restricted to the ENS: *Eno3* is expressed in epithelial cells and *Mmaa* across a wide variety of cell types. Conclusion: Gene expression profiles, including at single nucleus resolution, and chromatin accessibility identified enhancers of their cognate genes that vary across pre- and postnatal development provide specific coding and noncoding targets to search for variants affecting post-surgical persistence of HSCR.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Session III

PB2110 Temporal profiling of transcriptomics and chromatin accessibility during development in fetal 22q11.2 deletion syndrome reveled by single-cell sequencing

Authors:

Z. Zhou[1,2], Y. Zhao[1,2], Y. Zheng[1,2], S. Xue[1,2], Y. Cao[1,2,3], X. Chen[1,2,3,4], Y. Kwok[1], J. Liao[5], R. Choy[1,2,3,4,6], Z. Dong[1,2,3]; [1]Dept. of Obstetrics and Gynaecology, The Chinese Univ. of Hong Kong, Hong Kong, China, [2]Key Lab. for Regenerative Med., Ministry of Ed. (Shenzhen Base), Shenzhen Res. Inst., The Chinese Univ. of Hong Kong, Shenzhen, China, [3]Fertility Preservation Res. Ctr., Dept. of Obstetrics and Gynaecology, The Chinese Univ. of Hong Kong, Hong Kong, China, [4]Maternal-Fetal Med. Inst., Shenzhen Baoan Women's and Children's Hosp., Shenzhen Univ., Shenzhen, China, [5]6. Dept. of Chemical Pathology, Faculty of Med., The Chinese Univ. of Hong Kong, Hong Kong, China, [6]The Chinese Univ. of Hong Kong-Baylor Coll. of Med. Joint Ctr. for Med. Genetics, Hong Kong, China

22q11.2 deletion syndrome (22q11DS) is commonly presented with immunodeficiency, hypoparathyroidism and congenital heart disease (CHD). However, the underlying mechanism is largely unknown. Currently, RNA-seq in cultured amniotic fluids (AFs) provides additional information of RNA aberration in assistance to disease investigation. Here, we applied single-cell RNA-seq and ATAC-seq to cultured AFs from 12 22q11DS fetuses [six with CHD diagnosed at Gestational Weeks (GW) of 20 by routine morphology scan (as later GW) and six at 16GW as (early GW)], and 10 GW-matched normal controls (5 from 20GW and 5 from 16GW). The single-cell analysis revealed consistent composition (adherent cells) of cell populations and expression profilings from different cultured passages of AFs from the same samples. As cardiac anomalies were identified in all six cases collected at 20GW, we first investigated the unique expression pattern from these six cases as reference. ScATAC-seq showed alteration of chromatin accessibility in cases not only in the deleted region, but also in 1Mb upstream region. In addition, inflammation fibroblast populations (*CCL2*$^{+}$) showed reduced expression of *CXC* gene-family in 22q11DS, suggesting a blunt hemostatic and inflammatory response. Further, a novel fibroblast subpopulation (*FTX*$^{+}$) presented with strong expression correlation with published fetal myocytes and cardiac fibroblast, but with significantly differential expressions detected in CHD related genes compared with controls. Lastly, we identified a subgroup mesenchymal stem cells in controls, the expression pattern of which showed activation of VEGF/EDN signaling pathways related to hemostasis and vasculature development, but it was absent in cases. In comparison, 22q11DS at 16GW showed a transient expression/chromatin-accessibility profiling different from that at 20GW; however, the subpopulations in fibroblast and mesenchymal stem cells showed mostly consistent expression patterns, indicating single-cell analysis in AFs in early 2nd trimester would be a potential source for phenotype prognosis for 22q11DS fetuses. Co-immunofluorescent staining confirmed the findings. Overall, our study demonstrates a phenotype-related single-cell profiling during fetal development by using 22q11DS as a model.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Session I

PB2111 The experience of one newborn sequencing study in evaluating potential clinical sequencing laboratory partners and genomic sequencing platforms

Authors:

K. Kucera[1], L. Milko[2], H. Cope[3], J. Berg[4], D. Bailey, Jr[1], C. Powell[5], H. Peay[3]; [1]RTI Intl., Research Triangle Park, NC, [2]Univ North Carolina at Chapel Hill, Chapel Hill, NC, [3]RTI Intl., Durham, NC, [4]UNC Chapel Hill, Chapel Hill, NC, [5]UNC - Chapel Hill, Chapel Hill, NC

Background: Early Check is a voluntary newborn screening research program available to birthing parents in North Carolina. Since 2018 Early Check has been offering screening for a small panel of conditions with targeted assays. In 2021, we began planning for the addition of genomic sequencing to screen for hundreds of genetic conditions that are actionable in early childhood. During the planning phase we sought to answer two vital questions: 1) What sequencing platform should be used - custom capture sequencing of only the genes of interest, or bioinformatic selection of genes of interest from exome sequencing (ES) or genome sequencing (GS) backbone? 2) Are commercial clinical sequencing laboratories interested in partnering to advance the field of newborn sequencing? This work provides insight to other groups that seek to conduct population-based sequencing studies.
Methods: Study leadership identified priority metrics to evaluate sequencing options including accepted specimen type, platform flexibility, turnaround time, throughput, and cost. Meetings were scheduled with eight clinical sequencing laboratories selected from an initial list of fifteen candidates to investigate interest and preferred approach.
Results: All eight interviewed laboratories were interested in newborn sequencing research and indicated ability to return results from standard clinical short read sequencing within ≤6 weeks for 150-200 specimens/week. Two laboratories had a validated pipeline that included DNA extraction from dried blood spots (DBS) and five indicated the capacity to validate within 6-9 months. All laboratories preferred informatic selection of gene panels from ES, GS, or large medical capture panel backbone quoting panel design flexibility, and time and costs associated with production and validation of custom capture panels. Five laboratories preferred ES and two preferred a large medical capture panel backbone. Three offered an option to design a custom capture panel. While GS was discussed as the ideal backbone for informatic panel selection, it was generally not considered feasible due to reagent costs. However, decreasing cost of GS, along with public-private partnerships make GS a feasible option. Motivations for laboratories to partner in newborn sequencing research included philanthropy, data use for technology development, and investment into research in clinical NGS for NBS.
Conclusions: Commercial clinical laboratories are interested in partnering on newborn sequencing research and prioritize informatic selection of gene panels. Cost, throughput, and TAT were identified as current limitations to large scale NBS.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Session II

PB2112 Transcriptome profiling during early postnatal development in humans

Authors:

V. Stranecky[1], L. Steiner Mrazova[1], A. Vrbacka[1], T. Kobets[2], P. Janovska[2], J. Kopecky[2]; [1]Dept. of Paediatrics and Adolescent Med., 1st Faculty of Med., Charles Univ. and Gen. Univ. Hosp., Prague, Czech Republic, [2]Inst. of Physiology, Czech Academy of Sci., Prague, Czech Republic

Premature birth and its complications are the number one cause of death among children < 5 years of age and could impose long-term developmental and health problems. To improve the prognosis of premature newborns, it is necessary to understand the causes of their insufficient postnatal adaptation that depends on the changes in energy metabolism and other mechanisms. However, this research is limited by the lack of clinical materials. Here we present results of RNA sequencing that we used to characterize perinatal and especially early postnatal development using a unique biobank of mostly extremely premature newborns (n=37) who died during first 2 months after birth from a number of different causes. We used two continuous variables (gestational age and length of survival after birth) to find differentialy expressed genes and other tools such as gene ontology, pathway enrichment analysis and protein-protein interaction networks to reveal the biological significance of these genes.*This project is supported by Ministry of Health of the Czech Republic (AZV* NU20-07-00026*).*

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Session III

PB2113 Trio-based exome sequencing incorporating deep phenotyping improves the diagnostic yield of prenatal skeletal dysplasia

Authors:

G. Li[1], Q. Qi[1], N. Wu[2], Y. Jiang[1], X. Zhou[1], Y. Lv[1], N. Hao[1], Z. Zhao[2], K. Xu[2], J. Cai[2], Q. Li[2]; [1]Peking Union Med. Coll. Hosp., beijing, China, [2]Peking Union Med. Coll. Hosp., Beijing, China

Prenatal skeletal dysplasia refers to a group of rare genetic disorders that disrupt fetal skeletal development, resulting in abnormal bone growth, morphology, and even prenatal death. In this study, we conducted trio-based exome sequencing on 198 Chinese pedigrees with prenatal skeletal dysplasia. Indications for testing included abnormal ultrasound findings, previous pregnancy abnormalities, and relevant family history. Through deep phenotyping by experienced obstetricians, we identified a positive gene testing rate of 73%. Employing two research strategies, we classified subgroups based on the phenotype of prenatal skeletal dysplasia and categorized skeletal dysplasia according to molecular diagnosis. Notably, we curated hot genes associated with prenatal skeletal dysplasia, including *COL1A1* (21/198), *FGFR3* (19/198), *COL1A2* (12/198), and *COL2A1* (10/198).Furthermore, we conducted a comprehensive review of postnatal patients with skeletal dysplasia who carried these hot disease-causing genetic mutations identified in our study, as well as those from the Deciphering disorders Involving Scoliosis and COmorbidities (DISCO) study (http://www.discostudy.org/). This analysis aimed to elucidate the phenotypic and genotypic distinctions between prenatal and postnatal patients with skeletal dysplasia.This study represents the largest cohort of prenatal skeletal dysplasia in China, with a high positive rate of gene testing. We recommend incorporating hot genes such as *COL1A1*, *FGFR3*, *COL1A2*, and *COL2A1* into prenatal skeletal dysplasia panels. Ultimately, the integration of trio-based exome sequencing and deep phenotyping improves diagnostic yield for prenatal skeletal dysplasia. The phenotypic and genotypic distinctions between prenatal and postnatal patients with skeletal dysplasia expand our understanding of skeletal dysplasia.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Session I

PB2114 Triploidy complicated with Hyperinsulinism hypoglycemia with an unusually long lifespan.

Authors:

R. Mendelson[1], K. Mezeivtch[1], A. Crichton[1], R. Saadeh-Haddad[2], L. Rubin[1]; [1]Medstar Georgetown Univ. Hosp., Washington, DC, [2]Georgetown Univ. Hosp., Mclean, VA

Triploidy has been estimated to cause up to 20% of spontaneous abortions and is an uncommon cause of premature birth and perinatal death. Reports of survival beyond several days is extremely rare. In the few known cases of liveborn triploid infants, the clinical phenotype may include severe intrauterine growth restriction, macrocephaly, total syndactyly of third and fourth fingers, and brain, heart, kidney anomalies.

We present a case of an unusually long-surviving infant with triploidy. The diagnosis was unsuspected during pregnancy. The newborn was delivered at 36 weeks gestational age to a 33 year-old G2P2 mother via cesarian section due to breech presentation. Birthweight was 2.14 kg. He was admitted to the birth hospital NICU for respiratory distress, requiring non-invasive support. His course was complicated by apneas and development of necrosis of a portion of the small bowel that required a 35 cm small bowel resection and subsequent presumed parenteral nutrition-associated cholestatic liver disease. At one month of age, he was transferred to our center for management of his short gut syndrome and evaluation for liver and potentially intestinal transplantation.

Upon arrival, he was intubated and hypotensive requiring dopamine and dobutamine infusions and hydrocortisone. During air transport, he had experienced new onset seizures treated with multiple anticonvulsants. He was growth restricted and had multiple dysmorphic features including downward slanting palpebral fissures, hypertelorism, a thick/short neck, low set and posteriorly rotated ears with underdeveloped helices, bilateral syndactyly of fingers, single palmar creases, talipes equinovarus of both feet, syndactyly of left 2nd-3rd toes, and generalized hypotonia. A chromosomal microarray revealed triploidy, 69, XXY. Karyotyping of two tissues did not show mosaicism. His subsequent hospital course included an inability to wean from respiratory support, hyperinsulinemic hypoglycemia, progressive liver failure, a bacterial bloodstream infection, and progressive edema. After increasing cardiac rythmn disturbances and discussions with his parents, he expired at two months of age. Autopsy confirmed cholestatic liver disease, visceromegaly, and a small brain with focal polymicrogyria.

This case adds to the range of phenotypes of neonates with triploidy. Our patient displayed several previously undescribed triploidy features including hyperinsulinemic hypoglycemia, enterocolitis, and liver failure. We believe he is the longest reported survivor having complete triploidy.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Session II

PB2115 Uncovering human specific gene regulatory elements in heart organogenesis by integrating primary tissues and organoid model

Authors:

s. oh, K. Child, J. Cotney; UConn Hlth., Farmington, CT

Organogenesis is highly conserved but ultimately results in distinct differences in tissue phenotypes across species. Changes in gene regulation have been suggested to be major contributors to phenotypic differences, and many studies have suggested that particular non coding regions known as enhancers have contributed to substantial human specific phenotypes. However, there is still a lack of complete understanding for human organogenesis itself at the comprehensive level of gene expression dynamics and their regulations. To address this gap in knowledge, we focused on uncovering human specific gene expression dynamics in organogenesis by integrating primary tissue data and *in vitro* organoid model. To figure out human specific gene expression dynamics in embryonic stages, we identified human biased gene regulatory landscapes for 15 developing organs by jointly comparing previously generated datasets from human and mouse. We initially focused on heart organogenesis, which is reported to have less conserved enhancers than other tissues at the sequence level. With gene expression, we figured out that heart development in human CS13(Carnegie Stage 13) to CS23 is most comparable to mouse E10.5(Embryonic 10.5 day) to E14.5. Interestingly, some heart genes such as MYL7, ACTC1 and TBX20 show different gene expression dynamics between human and mouse despite the overall expression levels in heart being high, suggesting that human specific enhancers contribute to the different trends of gene expression. From ChromHMM based genome segmentation analysis, we found that these genes have some species specific strong enhancer segments as well as conserved enhancer segments, which have the potential to affect the species specific expression pattern of these genes during organogenesis. To fully understand the effects of these enhancers on early organogenesis, we generated heart organoid models from H9 human Embryonic Stem Cells(hESCs), which successfully showed morphological and structural features of developing heart. Our organoid model also shows inducible expression of KRAB dCas9 early in differentiation, which allows us to suppress the activity of such regions. We also performed joint profiling of gene expression and chromatin accessibility in single nuclei (snMultiome) on both primary heart tissue and cardiac organoids, which allowed us to explore these findings at the cell type level. Taken together, it is expected to identify such regions that are important for understanding human specific features of heart development by giving human specific gene expression features in organogenesis.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Session I

PB2116 Understanding caudal developmental abnormalities using 5' end single-nuclei RNA sequencing data from wild type and Danforth's short tail mouse E9.5 tailbuds

Authors:

C. Zajac[1], R. D. Albanus[2], N. Manickam[1], E. Curka[1], C. Keegan[1], S. Parker[1]; [1]Univ. of Michigan, Ann Arbor, MI, [2]Washington Univ., St. Louis, MO

Background: Caudal embryo development is important for multiple organ systems arising from all three germ cell layers, including the gastrointestinal system, genitourinary system, and the lower spinal cord. Embryonic malformations associated with caudal birth defects affect 1 in 10,000 human live births, but their mechanisms are largely unknown. The Danforth's short tail (*Sd*) mouse exhibits cessation of the vertebral column at the lumbar level and malformations of urogenital and gastrointestinal organs and therefore provides a tractable model to study the etiology of human caudal birth defects. The *Sd* mutation is caused by an endogenous retroviral (ERV) insertion 12.5 kb upstream of the *Ptf1a* gene in the promoter region of the lncRNA gene *Gm13344*. This region is orthologous to a human pancreatic developmental enhancer, which provides motivation for the use of the *Sd* mouse as a translational model to study caudal dysgenesis. **Methods:** To better understand the cell-specific regulatory elements upstream of the promoter regions of the *Sd* mouse genome, we generated 5'end single nuclei RNA sequencing (5'end snRNA-seq) data from control (WT) and *Sd* E9.5 tailbuds. We used a custom pipeline for droplet QC before clustering. Then, we performed SCAFE (Single-Cell Analysis of Five-prime Ends) on this dataset to visualize gene expression and identify transcription start sites and transcribed cis-regulatory elements. **Results:** We obtained 5,804 (WT) and 9,768 (*Sd*) high-quality nuclei for joint clustering analyses. We found 18 clusters representing different cell types, some of which have differential chromatin accessibility and are linked to caudal embryonic developmental gene markers including *Ptf1a*, *Shh*, and *Gm13344*. In addition, we validated these findings using the SCAFE results from our 5'end snRNA-seq dataset, where we observed co-expression of Gm13344 and *Ptf1a* solely in a restricted subset of clusters in the *Sd* sample. **Future direction:** Preliminary results provide insights into cluster characterization. We identified a subset of cells that may represent the early cellular events that are causal to the *Sd* caudal phenotype. The next steps in this project involve further integration of the 5' RNA and chromatin accessibility modalities to identify transcription factors perturbed as a result of the *Sd* mutation. We expect these integrative analyses will improve our understanding of the causal cell types and regulatory networks that influence the *Sd* phenotype and, consequently, caudal developmental disorders in humans.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Session I

PB2117 † Understanding splicing regulation during human heart organogenesis

Authors:

K. Child, J. Cotney; UConn Hlth., Farmington, CT

Congenital Heart Defects (CHDs) are the most common congenital abnormality worldwide, affecting approximately 1% of all births. However, ~60% of these cases remain unexplained. Properly controlled gene expression during cardiac development, particularly at the initiation of organogenesis, is crucial for normal development. This can occur at multiple levels including chromatin state, transcription initiation, and RNA splicing, as well as many post-transcriptional mechanisms. Recent analysis of RNA splicing patterns across multiple species and organs revealed that dynamics and complexity of splicing of the developing heart are significantly greater than most tissues and on par with brain. These findings were based primarily on heart samples from fetal and adult stages with little data from the critical organogenesis phase. To address this gap in knowledge, our lab has directly examined both chromatin state and composite level gene expression ranging from Carnegie stages (CS) 13 to 23 (4 to 8 post-conception weeks of development). We identified a module of genes with dynamic expression through this period that significantly enriched for genes involved in RNA splicing and processing. Given the highly dynamic nature of splicing at later timepoints and a coherent module of genes directly related to regulation of splicing, we sought to examine splicing and isoform utilization during the organogenesis phase. When we compared alternative splicing (AS) between our 8 individual CS timepoints, we find 3,749 genes with AS events across the various stages of heart organogenesis. These are distinct from our previously reported differentially expressed genes but also enriched for genes related to cardiac development. In addition to these trends, we found evidence for splicing events which result in previously unannotated transcripts. To confirm these results, we performed Nanopore long-read sequencing of cDNA libraries from our samples. Using a combination of long- and short-read data we generated a comprehensive, novel cardiac-centric gene annotation. From this analysis we identified 13,137 novel transcripts including variants of well-known cardiac genes. One example is *TBX5*, a well-known cardiac gene implicated in Holt-Oram Syndrome. We discovered a novel TSS for *TBX5* which is highly conserved across species and constrained in human populations. Understanding transcript utilization and AS will lead to better screening for CHDs clarifying the unexplained CHD cases and provide a better understanding of gene expression dynamics during this critical period of development.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Session II

PB2118 Validation of low-pass genome sequencing as a replacement for chromosomal microarray in prenatal diagnostics

Authors:

E. Greenfeld[1], C. Mighton[2], N. Watkins[1], V. Di Gioacchino[1], J. Lerner-Ellis[3], E. Mukharryamova[1], N. Anggala[1], A. Noor[4]; [1]Mount Sinai Hosp., Toronto, ON, Canada, [2]St. Michaels Hosp. & Univ. of Toronto, Toronto, ON, Canada, [3]Mount Sinai Hosp., Sinai Hlth., Toronto, ON, Canada, [4]Mount Sinai Hosp, Toronto, ON, Canada

Background: Low-pass genome sequencing (LP-GS) is emerging as an alternative to chromosomal microarray (CMA) for the detection of copy number variants (CNVs). This study aimed to assess whether LP-GS can replace SNP-based CMA in prenatal diagnostics. **Methods:** A validation study was conducted to evaluate the performance of LP-GS at 10x compared to CytoSNP 859 (Illumina). 96 samples were included, of those, 40 positive samples with a range of abnormalities including CNVs of varying sizes, low-level mosaic aneuploidies, complex structural abnormalities (mosaic and non-mosaic), and AOH. 40 negative control samples were included. 14 samples were run by LP-GS at 5x and 10x for comparison. Concordance, sensitivity, specificity, reproducibility, repeatability, limit of detection, turnaround time, and cost-effectiveness were assessed. Tertiary analysis of LP-GS data employed the Genoox Franklin pipeline. Data from 34 samples were run through both the Illumina DRAGEN CNV pipeline and Genoox Franklin pipeline for comparison of two CNV analyses. **Results:** The results from all samples run by LP-GS at 10x and CMA were concordant with the exception of one false negative (69,XXX triploidy) which was missed by LP-GS. Sensitivity of LP-GS for detecting clinically relevant variants previously characterized by CMA was 97.5%, and specificity was 100%. Clinically relevant findings were concordant between the samples run at 10x and 5x, however, there were more failed and low confidence variant calls from the samples run at 5x, reflecting artefacts. There was 100% concordance between LP-GS results from 150 ng input and higher input. Reproducibility and repeatability experiments resulted in 100% concordance. Results were concordant between the Genoox Franklin and Illumina DRAGEN pipelines, however the calls from DRAGEN were fragmented, in particular for larger CNVs and whole chromosome aneuploidies. Workflow was shorter for LP-GS (1.5 days) than for CMA (2.5 days). Overall, the LP-GS turnaround time was comparable to CMA. LP-GS at 5x was cost-effective compared to the Illumina CytoSNP-850K array. LP-GS at 10x was almost cost neutral using S2 flow cell and 33 samples, and cost-effective with S4 flow cell. **Discussion:** LP-GS is a promising replacement for CMA in prenatal diagnostics. The one false negative case indicates that better integration of genotyping data to enhance CNV pipelines is needed. Strengths of LP-GS include a simplified protocol and faster turnaround time, gene content agnostic detection of variants, and integrated workflow for detecting single nucleotide variants and CNVs.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Session III

PB2119 Variants in DOK7 gene results in fetal akinesia deformation sequence

Authors:

A. TIWARI, D. Saxena, S. Phadke; Sanjay Gandhi Postgraduate Inst. of Med. Sci., Lucknow, India

INTRODUCTION: Fetal akinesis deformation sequence (FADS) is characterized by reduced fetal movements, joint contractures, facial anomalies, pulmonary hypoplasia and IUGR. DOK7 (Downstream of Tyrosine Kinase 7) gene was identified as a cause of FADS in 2009. Here, we report a case of FADS due to biallelic variants in DOK7 gene.

PATIENT DETAILS: At 24 weeks, antenatal scan showed 22 weeks fetus with subcutaneous edema, absent stomach, IUGR, arthrogryposis and polyhydramnios. Postnatally, fetus had facial dysmorphism (flat facies, depressed nasal bridge, upturned nose, microretrognathia), joint contractures (overlapping digits, flexion at elbow and hyperextension at knee joint, prominent calcaneum), short neck and subcutaneous edema on dorsum of hands and feet. On internal examination, lungs were hypoplastic. Whole exome sequencing identified compound heterozygous variants, c.68G>A (p.W23X) in exon 2 and c.1373dupC (p.Q460Pfs*59) in exon 7 of DOK7 gene (NM_173660). These are predicted to be likely pathogenic by ACMG/ AMP guidelines.

DISCUSSION AND CONCLUSION: First report of DOK7 related FADS concluded that complete loss of function due to variants in pleckstrin homology (PH) or phosphotyrosine binding domains result in a lethal fetal akinesia phenotype, whereas the variants associated with CMS are present in all domains. However, this was refuted by Radhakrishnan P, et al. who reported the 2nd case of DOK7 related FADS due to biallelic variant c.1263dupC, reported earlier in CMS. This is the third report of FADS due to biallelic DOK7 variants and further strengthens the association of DOK7 gene with this lethal phenotype and lack of genotype-phenotype correlation and further management of future pregnancy.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Session I

PB2120 Variants of uncertain significance identified with prenatal exome sequencing: to report or not to report?

Authors:

G. Santen, M. de Koning, M. Haak, E. Nibbeling, C. Ruivenkamp, M. Hoffer, M. Suerink; Leiden Univ. Med. Ctr., Leiden, Netherlands

Objectives: Congenital anomalies affect a substantial number of pregnancies and prognoses are greatly influenced by the presence of an underlying genetic disorder. Prenatal exome sequencing (pES) is increasingly being implemented in routine genetic diagnostic work-up. Diagnostic rates of pES are around 10%. In addition, variants of uncertain significance (VUS) are detected. Their interpretation is challenging due to lack of a complete clinical phenotype in a prenatal setting. Furthermore, the choice to report or not is a balancing act between reproductive autonomy and paternalism. VUS therefore provide a challenge for molecular and clinical geneticists, MFM specialists and future parents. We describe our experiences with reporting VUS in a prenatal setting.
Methods: In the Netherlands, pES is offered in pregnancies with a congenital anomaly detected by ultrasound screening. We retrospectively analyzed all consecutive pregnancies in which VUS identified by pES were reported to parents during pregnancy (n=20).
Results: VUS were mainly reported in genes associated with the detected fetal phenotype (n = 13). Most VUS were in genes associated with an autosomal dominant inheritance pattern and were de novo (n = 6) as well as inherited (n =6). Of the parents considering TOP (n = 14), four times the added uncertainty of the detected VUS contributed to pregnancy termination. In two cases, parents chose to continue the pregnancy either because the gene was not associated with intellectual disability or because the variant was inherited from a mildly affected parent. When TOP was not considered (n = 6), the VUS changed perinatal management in two cases: neonatologists were standby during labor in a case with increased risk of tracheomalacia due to a variant in *DCHS1,* and testing for metabolic acidosis was requested in a case with a variant in *UQCRC2*. Of the 13 live births, two probands were lost to follow-up and two demised at age of 1 and 15 months respectively. Of the other nine live born children, four have some degree of developmental delay.
Conclusions: Uncertain findings can be viewed as undesirable, especially in the prenatal setting with added stress of time pressure and incomplete phenotyping. Nevertheless, VUS might be helpful in guiding prenatal decision making or adjusting perinatal management. We advocate for future research to focus on patients' and clinicians' perspectives regarding VUS, in order to formulate much needed guidelines balancing reproductive autonomy and paternalism.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Session II

PB2121 Whole exome sequencing in fetuses with abnormal ultrasound finding.

Authors:

D. Saxena, A. K. Tiwari, A. Moirangthem, K. Mandal, S. Srivastava, H. Sait, S. Phadke; Sanjay Gandhi Postgraduate Inst. of Med. Sci., Lucknow, India

Background: The incidence of major structural malformations is around 2-3%, which may be due to chromosomal abnormalities, single gene disorders, teratogens etc. Without molecular diagnosis, prenatal diagnosis in next pregnancy cannot be provided. Diagnostic yield of fetal karyotyping is around 8-10%. Use of cytogenetic microarray has further increased the diagnostic rate by around 6%. Still, around 80% of the cases remain undiagnosed. Use of next generation sequencing in prenatal cases had increased the diagnostic rate by around 20-30%. Objectives:To study the role of whole exome sequencing (WES) in fetuses with abnormal ultrasound findings.Methods: WES (trio/ solo) was performed on cases with fetal malformations detected on antenatal ultrasound, selected over a period of two years from September, 2020 to September, 2022. Results: A total of 50 families were included, comprising fetuses with skeletal defects (32%, 16/50), central nervous system malformations (22%, 11/50), hydrops (10%, 5/50), renal (10%, 5/50) and cardiac malformations (6%, 3/50), recurrent neural tube defects (8%, 4/50) and miscellaneous conditions including caudal regression syndrome (2 cases), severe oligohydramnios (2 cases), multiple pterygium and congenital diaphragmatic hernia (1 case each). Trio WES was done in 21 (42%) cases whereas solo fetal exome sequencing was done in 29 (58%) cases. Diagnostic yield was found to be 48% (24/50) and was highest in fetuses with skeletal dysplasia (81%, 13/16) and hydrops (80%, 4/5) followed by renal (40%, 2/5), CNS (36%, 4/11) and cardiac malformations (33%, 1/3). No causative variations were identified in fetuses with recurrent NTDs and other miscellaneous defects.We found pathogenic variants in 10 cases (CTSA, COASY, FGFR3, GNPTAB, PEX1, P3H1 in 2 cases, THSD1, TMEM67 and OBSL1), likely pathogenic variants in 9 cases (COL1A1, CREBBP, DOK7, KIF5B, L1CAM, LZTR1, PIEZO1, P3H1 and ZNF699) and variant of uncertain significance (VOUS) in 5 cases (MTOR, CHST3, COL2A1, COL6A1 and TRAPPC12). Most common disorder observed in our cohort was osteogenesis imperfecta, seen in 16.7% (4/24) cases. Most of the disorders detected (62.5%, 15/24) were autosomal recessive, 33.3% (8/24) had autosomal dominant disorder (de novo-4 and inherited - 2) and 4.2% (1/24) had an X-linked recessive disorder.Conclusion: Our study further emphasizes the role of whole exome sequencing in fetal abnormalities. Confirmation of molecular diagnosis improves pregnancy management, prenatal counselling, and assessment of recurrence risk in further pregnancies. Identification of novel variants further expanded the mutational spectrum of fetal disorders.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Session III

PB2122 Whole genome sequencing identifies very rare de novo and inherited pathogenic variants in European cleft case-parent trios.

Authors:

E. Aladenika[1], A. Alade[2], M. Olujitan[1], W. Awotoye[1], A. Oladayo[3], S. Miller[4], T. Busch[1], A. Butali[3]; [1]Univ. of Iowa, Iowa, IA, [2]Univ. of Iowa, Iowa City, IA, [3]Univ. of Iowa, Coralville, IA, [4]Univ Iowa, Iowa, IA

Introduction One in 700 infants worldwide is born with orofacial cleft (OFC). OFC etiology is multifactorial involving genetics and environmental factors. In the current study, we investigated the contributions of de novo and inherited (novel and rare) variants in craniofacial genes using whole genome sequencing data of European cleft case-parent trios. **Methodology** Whole genome sequencing was conducted on 15 case-parent trios (syndromic and non-syndromic). Called variants were annotated using GATK. Annotated variants were filtered based on genotype quality ≥20 and read depth ≥10. Novel (MAF=0) and very rare (MAF < 0.001) de novo and inherited protein-altering variants were identified. Deleteriousness of these protein-altering variants were predicted using polyphen, SIFT and a CADD score cutoff of ≥ 15. Variants predicted to be deleterious across polyphen, SIFT and CADD were prioritized. ACMG criteria was also used to assess pathogenicity using VARSOME. Amino acid changes effect on protein structure and function were identified using HOPE. The genes harboring these variants were assessed for contributions to craniofacial development using Mouse Genome Informatics (MGI), Online Mendelian Inheritance in Man (OMIM) and CleftGeneDB. Genes involved in craniofacial development in any of the 3 databases were prioritized. DECIPHER and thorough literature search were used to assess the implication of these prioritized genes in OFC. **Results and discussion** We identified 186 de novo variants in 141 genes. Eighteen of these genes harbor 19 deleterious variants. *VCAN* which is expressed in relevant craniofacial tissues contains a very rare pathogenic de novo variant rs781121274 (MAF= 0.0001). We also identified 684 inherited (novel and very rare) variants, among these, 84 were predicted to be deleterious. Eight of these deleterious variants were identified in genes that are expressed in craniofacial tissues, namely, *CFAP45*, *NISCH*, *CDKN1A*, *NOTCH1*, *FUZ*, *COL6A2*, *IL11RA*, and *AKAP6*. The resulting amino acid changes in *NOTCH1* and *COL6A2* are predicted to be likely damaging by HOPE. Patients with deletion of loci that contain *CFAP45, NISCH, NOTCH1, COL6A2* or *AKAP6* have OFCs as reported phenotypes. Similarly, there are reports that implicate these genes in OFC development. **Conclusions** Our study provides additional evidence for the role of *VCAN, CFAP45, NISCH, NOTCH1, FUZ, COL6A2* and *AKAP6* in OFC development. This is the first time *CDKN1A* and *IL11RA* will be implicated in OFC development. Identification of very rare variants (<0.001) in these genes could explain the missing heritability for OFC. Future functional experiments to validate these observations are needed.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Session I

PB2123 Whole genome sequencing of fibroid samples identified potential novel loci associated with fibroid risk

Authors:

J. Jaworski[1,2], J. N. Hellwege[1,3], E. A. Jasper[1,4], T. L. Edwards[1,2], D. R. Velez Edwards[1,5,4]; [1]Vanderbilt Genetics Inst., Vanderbilt Univ., Nashville, TN, [2]Div. of Epidemiology, Dept. of Med., Vanderbilt Univ. Med. Ctr., Nashville, TN, [3]Div. of Genetic Med., Dept. of Med., Vanderbilt Univ. Med. Ctr., Nashville, TN, [4]Div. of Quantitative Sci., Dept. of Obstetrics and Gynecology, Vanderbilt Univ. Med. Ctr., Nashville, TN, [5]Dept. of BioMed. Informatics, Vanderbilt Univ. Med. Ctr., Nashville, TN

Uterine fibroids are benign smooth muscle growths affecting up to 75% of women by menopause. African American (AA) women have higher prevalence, earlier onset, and larger and more numerous fibroids than European American (EA) women. Genome-wide association studies have found numerous genetic loci associated with fibroids risk, however most of those studies were of European ancestry. Sequencing studies of fibroids to date have focused on somatic sequence changes. We performed whole genomic sequencing on AA and EA samples to determine the impact of rare variants on fibroid risk across populations. Samples from a de-identified electronic health record system linked to a DNA biorepository (BioVU) were selected using a previously described algorithm. A total of 548 AA (271 cases/ 277 controls) and 659 EA (371 cases/ 288 controls) samples were whole genome sequenced using the Illumina Dragen Germline application (v.3.10.4) on the Illumina Basespace sequence hub. EA and AA groups were joint called separately using the Dragen PopGen application (v.3.10.4). Sample and variant quality control was performed in PLINK2. The Ensembl Variant Effect Predictor was used for all annotation. Single variant and gene-based testing were performed using RVTESTS. Single variant analyses, adjusted for age and 10 principal components, were limited to variants with a minimum minor allele count (MAC) of 20. A cross-ancestry meta-analysis was completed using METAL. The gene-based tests were performed with the RVTESTS option SKAT-O, using rare, coding, and loss of function variants. For AA samples, the single variant association analysis showed nominally significant loci in 4p12 (p-value 1.5×10^{-8}) and 20q11.22 (p-value 1.4×10^{-7}). For the EA, there were variants close to significance on 12p13.33 (p-value 9.8×10^{-8}) in the gene *LINC02443*. Cross-ancestry meta-analysis detected a locus at 1p31.3 (p-value 4.9×10^{-8}) in the gene *ROR1*. The 4p12 locus was detected in the meta-analysis (p-value 5.6×10^{-7}) as well as the 12p13.33 (p-value 8.1×10^{-7}). For the gene-based tests, *SH2B3* (p-value 2.1×10^{-5}) and *EHD1* (p-value 2.4×10^{-5}) were nominally significant in the AA dataset. The top EA result was *SLC26A2* (p-value 5.2×10^{-6}). Whole genome sequencing of fibroid samples identified potential novel loci associated with fibroid risk within and across the AA and EA populations. Increasing the sample size with more sequencing and collaboration with existing datasets will further clarify these results.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Session II

PB2124 Whole genome sequencing to better understand the clinical heterogeneity of SHANK3-related PMS patient.

Authors:

A-C. Tabet[1], A. Vitrac[2], C. Leblond[3], A. Maruani[1], M. Rachid[4], F. Cliquet[5], A. Mathieu[5], F. Amsellem[6], J. Levy[7], Y. Capri[8], L. Perrin[1], S. Drunat[4], D. germanaud[1], N. Couque[9], C. Dupont[1], L. Ruaud[10], e. Pipiras[1], A. Verloes[11], R. Delorme[1], T. Bourgeron[5]; [1]Robert Debre Hosp., Paris, Paris, France, [2]Pasteur intitute, Paris, Paris, France, [3]Pasteur Inst., Paris, Paris, France, [4]Robert Debre Hosp., Paris, France, [5]Inst. Pasteur, Paris, France, [6]Inst. Pasteur, Paris, Paris, France, [7]AP-HP, PARIS, France, [8]Univ. Hosp. APHP-Nord Robert Debré, PAris, France, [9]Robert Debré Hosp., PAris, France, [10]Robert Debre Hosp., PAris, France, [11]Robert DEBRE Univ. Hosp., Paris, France

Phelan-McDermid syndrome (PMS) is a neurodevelopmental disorder (NDD) caused by a loss of the distal part of chromosome 22 or by a pathological *SHANK3* variant. There is a wide variability of the mechanism and the size of the deletion among PMS patients ranging from less than 100 kb to over 9 Mb. Indeed, the distribution of the breakpoint position is mostly uniform along the 22q13 region, with the exception of a recurrent breakpoint within the *SHANK3* gene. PMS patients present with global developmental delay, intellectual deficiency, language impairment ranging from severe delay to absence of speech, neonatal hypotonia, minor morphological features and other nonspecific clinical features which differ from a patient to another leading to a high phenotypic heterogeneity. A number of studies have explored the clinical diversity observed in PMS patients which has led to the identification of correlation between the size of the deletion and the presence or severity of the PMS symptoms. In addition to deletion size, additional genetic variants could also contribute to the phenotypic difference existing between each patient with PMS. In order to identify such variants, we selected 90 *SHANK3*-related PMS patients carrying a 22q13 deletions or a pathogenic variant affecting *SHANK3* and collected their DNA (and the DNA from their parents when available) to perform whole genome sequencing. The analysis of *de novo* variants shows the presence of multiple hits in genes associated to NDD and predicted as deleterious. In particular, for some patients with epilepsy history, we could detect *de novo* deleterious variants in an HPO risk gene for epilepsy. We also found LoF and predicted deleterious missense variant in the 22q13 region on the intact chromosome 22 for patients with 22q13 deletion. The genetic and clinical profile of the PMS *SHANK3* patients will be described and compared with three other cohorts. In a next step, the common variants of *SHANK3*-related PMS will be also explored using PRS analysis.

Session Title: Prenatal, Perinatal, and Developmental Genetics Poster Session III

PB2125 Zellweger spectrum disorder: The case of a missing *PEX1* allele in a fetus with multiple congenital anomalies.

Authors:

T. Sneddon, K. Gilmore, E. Jalazo, B. Powell, N. Vora; Univ. of North Carolina, Chapel Hill, NC

PEX1 and *PEX6* encode ATPases that are essential for peroxisome biogenesis, maintenance, and prevention of peroxisome degradation. Biallelic variants in either gene are associated with the peroxisome biogenesis disorder known as Zellweger spectrum disorder (ZSD). This is an autosomal recessive disorder with a broad phenotypic spectrum where severity is inversely correlated with residual protein function.

Trio genome analysis was performed on a fetal sample from a patient enrolled in the Prenatal Genetic Diagnosis by Genomic Sequencing (PrenatalSEQ) multicenter study due to ultrasound findings that included cerebral ventriculomegaly, unilateral left clubbed foot, and a cardiac ventral septal defect. Analysis of filtered sequence variants identified a maternally-inherited pathogenic nonsense variant in *PEX1* [c.2614C>T; p.(Arg872Ter)] and a paternally-inherited pathogenic frameshift variant in *PEX6* [c.1314_1321del; p.(Glu439fs)]. Digenic inheritance has not been reported for ZSD, and the fetal phenotype was non-specific, so this result was non-diagnostic.

Postnatally the baby had severe hypotonia, respiratory distress, and dysmorphic features. At 2 weeks of life, the state newborn screening reported an elevated C26:0 ratio of 1.11 µM (cut off <0.15 µM) that resulted in diagnostic biochemical testing, confirming a diagnosis of a peroxisomal biogenesis disorder. Subsequent manual review of the trio genome sequence read alignments revealed a paternally-inherited 740 bp *PEX1* deletion variant (c.2719-251_2783+425del) that removes exon 17 of 24 and was confirmed by Sanger sequencing. These compound heterozygous *PEX1* variants were reported as consistent with a diagnosis of ZSD.

Concurrently, a peripheral blood sample sent to a commercial lab for clinical sequence analysis and deletion/duplication testing of an 18-gene ZSD panel, that included *PEX1* and *PEX6*, failed to identify the 740 bp *PEX1* deletion. Communication with the lab resulted in an update to their bioinformatic pipeline such that the 740 bp *PEX1* deletion would be identified on subsequent samples, providing reassurance to the parents that clinical prenatal testing by that lab would identify both *PEX1* variants in future pregnancies.

This case highlights the benefits of research studies to inform and improve clinical care and clinical genetic testing, and also serves as a caution that single-exon deletions are not detected by all genetic tests or genomic methodologies and, as exemplified in this case, may account for some of the missing diagnostic yield from clinical genetic testing.

Session Title: Pharmacogenomics Poster Session I

PB2126 A blood-based multi'omic analysis of response to vedolizumab treatment in IBD

Authors:

J. Rioux[1,2], G. Boucher[3], J. Thompson Legault[3], A. Forest[1], C. Des Rosiers[1,2], A. Ananthakrishnan[4], R. Xavier[4,5]; [1]Montreal Heart Inst. | Res. Ctr., Montreal, QC, Canada, [2]Université de Montréal, Montreal, QC, Canada, [3]Montreal Heart Inst., Montreal, QC, Canada, [4]Massachusetts Gen. Hosp., Boston, MA, [5]Broad Inst., Boston, MA

INTRODUCTION The inflammatory bowel diseases (IBD) are chronic relapsing inflammatory diseases of the gastrointestinal tract that affect over one million Americans. A key challenge in IBD is matching patients with the most effective therapy as this would lead to improved clinical and endoscopic outcomes, less tissue damage and a decrease in societal costs. In the current study, we chose to focus on the response to treatment with vedolizumab. Rather than blocking cytokine signaling as most biologics, vedolizumab is believed to interfere with immune cell trafficking to the intestines. Given our ultimate objective to develop a biomarker panel to facilitate therapeutic decisions in the clinic, we have chosen to prioritize blood-based biomarker discovery.
METHODS The study cohort consisted of a prospective cohort of 92 IBD patients where vedolizumab was administered intravenously at weeks 0, 2, 6 and 14, and serum samples collected at weeks 0 and 14. These serum samples were evaluated for the level of a targeted set of 49 cytokines, chemokines, and growth factors, many of which are encoded by genes within IBD loci (PMID: 34106269). In addition, we performed targeted metabolomics (amino acids, organic acids, and acyl carnitines); all metabolite classes implicated in immune cell activity and/or signaling. Moreover, we used a non-targeted lipidomics approach measuring >1500 metabolites.
RESULTS We observed that in patients that responded well to vedolizumab treatment there was an important increase in serum levels of interferon-regulated cytokines between week 0 and week 14. These cytokines are hallmarks of pro-inflammatory "M1" macrophages, and are elevated in the serum of IBD patients as compared to matched controls. In terms of the serum levels of amino acids, organic acids and acyl carnitines, we found metabolite associations in baseline samples and/or changes that occurred during the 14 weeks of treatment in responders as compared to non-responders that were consistent with a metabolic reprogramming of macrophages. Finally, we identified changes in lipid metabolites that we previously associated with host and microbial metabolic functions associated with IBD (PMID: 36662167).
CONCLUSION The results support the hypothesis that there is biological heterogeneity in patients with IBD that can have an impact on important clinical outcomes such as response to therapy. More specifically, these results suggest that in the subset of patients that respond well to vedolizumab there is a selective interference of M1 macrophage trafficking to the intestines and/or metabolic reprogramming of circulating macrophages.

Session Title: Pharmacogenomics Poster Session II

PB2127 † A Flexible Network Based Framework for Identifying Drug-Disease Relationships including Repurposing Opportunities and Adverse Effects

Authors:

R. Chen[1], Q. Wang[1], Q. Wei[2], Y. Yan[1], Y. Tan[3], A. Tiwari[1], X. Zhong[4], B. Li[1]; [1]Vanderbilt Univ., Nashville, TN, [2]Vanderbilt Univ, Nashville, TN, [3]Vanderbilt Univ. - Nashville, TN, Nashville, TN, [4]Vanderbilt Univ. Med. Ctr., Nashville, TN

Given the high cost of drug development, there has been a growing interest in drug repurposing. In parallel, predicting adverse drug reactions (ADR) for a drug, another type of drug-disease relationships, has significant value on minimizing the cost of drug development as well as mitigating the occurrence of ADRs in medical care. For effective drug repurposing and ADR prediction, it is key to integrate heterogenous data at different levels to accurately link drugs and diseases. In this study, we develop a computational framework that builds multi-layer networks of three types of nodes: genes, diseases and drugs, based on disease-disease similarity networks, gene-gene networks, disease-gene association networks, drug-gene networks, and drug-drug similarity networks. It uses the Random Walk with Restart algorithm to traverse the networks to build and score relationships of genes, diseases and drugs. Our framework is flexible and allows for easy addition of networks of any of the three types of nodes. We applied our framework on 220 diseases, with integrative networks of 1) Gene Ontology, KEGG, PPI, tissue and cell type specific co-expression networks; 2) phenotypic and semantic similarity networks for diseases; 3) structure similarity networks for drugs; 4) bipartite networks of drug-gene, disease-gene, and drug-disease connections. Overall, we observed that the top scoring drug-disease pairs exhibit a remarkably enrichment in clinical trials (permutation test, $p < 10^{-4}$) and off-label uses ($p < 10^{-4}$), suggesting novel repurposing opportunities in our prediction. Specifically, we found that 55.8% of approved drug-indication pairs (DIPs), 43.1% of off-label DIPs and 53.9% of the clinical trial DIPs are among the top 100 (of 8120) of the predictions (all $P<10^{-4}$). In parallel to drug repurposing, we found that 51% of reported Drug-Adverse Effect Pairs (DAEPs) are in the top 100 predictions ($P<10^{-4}$). To further assess whether our framework is able to predict ADR for new drugs, we collected reported DAEPs for 77 drugs that were approved after 2015, and observed that 40.4% of reported DAEPs are within the top 100 predictions, demonstrating our framework's ability to predict DAEPs for new drugs. We created a comprehensive table of all drug-disease pairs ranked by their predicted scores, encompassing 1520 drugs and 8120 diseases, which we hope are useful for the drug repurposing and ADR field.

Session Title: Pharmacogenomics Poster Session III

PB2128 A structural biology approach for PheWAS analysis for antidepressant repurposing

Authors:

L. Peng, C-Y. Chen; Biogen, Cambridge, MA

Drug repurposing is becoming an increasingly attractive and economically viable option to identify new indications for an approved or investigational drug, given the risks, costs, and slow pace of drug discovery and development. Here we present phenome-wide association studies (PheWAS) as an approach to identify protein targets from four classes of commonly-prescribed antidepressants (ADs) for neurological disorders beyond mood and behavioral disorders (depression, anxiety), guided by knowledge of the AD inhibitor-bound and *apo* (AD inhibitor-unbounded) conformational states of the target. Namely, selective serotonin reuptake inhibitors (SSRIs) target the serotonin transporter (SERT, *SLC6A4*); serotonin-norepinephrine reuptake inhibitors (SNRIs) target the norepinephrine transporter (NET, *SLC6A2*); tricyclic antidepressants (TCAs) target SERT and NET, and monoamine oxidase inhibitors (MOAIs) target monoamine oxidase A (MAOA, *MAOA*) and monoamine oxidase B (MAOB, *MAOB*) enzymes. For genes encoding these proteins we define burden sets for loss-of-function (LoF) and missense variants based on their translational consequences in the protein (residue mutations) and their location in domains known for functional regulation of neurotransmitters. PheWAS analysis on 470K UK Biobank (UKB) samples with exome sequencing data for single variants and burden sets on *SLC6A4*, *SLC6A2*, *MAOA*, and *MAOB* was conducted for 500 phenotypes across ICD codes from in-patient diagnostic records to identify rare variants and gene burdens with strong disease associations for which antidepressants can be repurposed for treatment.We identified, for example, a single variant association (P = 8.515E-6) in *SLC6A2* with intracerebral hemorrhage, suggesting a repurposing opportunity for SNRIs. Our approach demonstrates how knowledge of a protein's conformational state can be used for genetics analysis to identify genetic variants that cause disease.

Session Title: Pharmacogenomics Poster Session I

PB2129 A synergistic effects of chemical chaperone and valproic acid on GBA mutations in Gaucher disease.

Authors:

J-J. Lee[1], M. Kang[2], I. Choi[3], G-H. Kim[1,4], H. Shin[2], H. Bae[5], S. Hwang[5], B. Lee[4,5]; [1]Asan Med. Ctr., Seoul, Korea, Republic of, [2]Asan Inst. for Life Sci., Seoul, Korea, Republic of, [3]Dept. of Genetic Counseling, Univ. of Ulsan Coll. of Med., Seoul, Korea, Republic of, [4]Univ. of Ulsan Coll. of Med., Seoul, Korea, Republic of, [5]Asan Med. Ctr. Children's Hosp., Seoul, Korea, Republic of

Gaucher disease (GD) is one of the most common lysosomal disorders caused by glucocerebrosidase (GCase) deficiency. The impairment of GCase results in the accumulation of glucosylceramide (Gb1), which is deacylated form, glucosylsphingosine (Lyso-Gb1). Enzyme replacement therapy (ERT) and substrate reduction therapy (SRT) have been approved for the treatment of GD, effectively improving visceral, hematologic, and skeletal abnormalities. However, neurological manifestations remain limited in patients with GD2/3. Chemical chaperones, such as Ambroxol (ABX), enhance the GCase activity and can penetrate the blood-brain barrier, thereby improving neurological symptoms in GD. Here, we evaluated the efficacy of high-dose ABX in various GBA mutations using the COS7 cell line. we confirmed an increase in GCase activity and modulation of EMT markers in response to high-dose ABX treatment in various GBA mutations. In addition, we observed that the co-administration of ABX and valproic acid (VPA) in HepG2 cells resulted in an increase in GCase activity and modulation of epithelial-mesenchymal transition (EMT) markers. In conclusion, our finding suggests that synergistic effect of the combined treatment with high-dose ABX and VPA in promoting GCase upregulation and regulating EMT markers in GD.

Session Title: Pharmacogenomics Poster Session II

PB2130 Allele frequency analysis of genetic variants related to the pharmacogenomics of Parkinson's disease in Latinos: A second-look between population groups, towards a feasible clinical reproducibility of results.

Authors:

H. Chaparro-Solano[1,2], T. Peixoto Leal[2], I. Fernandez Mata[1,2], Latin American Research Consortium on the Geneticsof Parkinson's Disease (LARGE-PD); [1]Dept. of Molecular Med., Cleveland Clinic Lerner Coll. of Med., Case Western Reserve Univ., Cleveland, OH, [2]Genomic Med. Inst., Lerner Res. Inst., Cleveland Clinic, Cleveland, OH

Background
Despite Parkinson's disease (PD) being the second most prevalent neurodegenerative pathology worldwide, available treatments result in wide variability in patient response in terms of dosage and side effects. Although several single nucleotide polymorphisms (SNPs) have been suggested as clinically useful pharmacogenetic markers to address this challenge, none of them have been reproducible. The aim of this study was to gain insights into the genetic variants related to PD treatment response by comparing allele frequencies (AF) of SNPs associated with treatment variability in PD among Latinos and other population groups.
Methods
We used data from the Latin American Research Consortium on the Genetics of PD (LARGE-PD). Genotypes of 452 healthy participants from seven different countries (Argentina, Brazil, Colombia, Chile, Mexico, Peru, and Uruguay) were included. SNPs of interest were selected through a systematic review. Control population-level allele frequencies of these SNPs were obtained from gnomAD v2.1.1 using the "total of genotypes" (TG) and the "Latino/Admixed American" (LAA) population. AFs for our sample were calculated. Fisher's exact test was employed to assess whether there were statistically significant differences in the AFs of the selected SNPs when comparing the three groups with each other. Statistical significance was set at $P<0.05$.
Results
A total of 52 SNPs were obtained from 76 studies. Most of these studies were conducted in European populations (41%), with only 17% including Latino individuals. LARGE-PD AFs compared to TG showed statistically significant differences. A similar result was observed when comparing TG with the LAA sample, except for four of them. When comparing the LARGE-PD sample with the LAA sample, a third of the analyzed SNPs also exhibited statistical differences, including the majority of those located in the *COMT* gene, with the exception of rs4633.
Discussion
Our results suggest that differences in AF between population groups could explain the lack of reproducibility of previously published results. While the AF differences between the gnomAD TG and our sample were expected, the differences observed between the two Latino groups were not. This may be attributed to the limited inclusion of participants from all Latin-American countries, a factor that was considered and managed in our recruitment process. The consistency in rs4633 among the Latino groups is remarkable given that it is one of the most extensively studied SNPs. Since population admixture could influence AF, ensuring greater genetic diversity in future pharmacogenomic studies is essential.

Session Title: Pharmacogenomics Poster Session III

PB2131 An association of rs7442295 in SLC2A9 with urate levels at baseline and in interaction with iloperidone

Authors:

C. Tyner, S. Smieszek; Vanda Pharmaceuticals Inc., Washington, DC

Serum urate concentrations represent a complex genetic trait influenced by genetic variation as well as environmental factors. Serum uric acid levels correlate with blood pressure, metabolic syndrome, diabetes, gout, and cardiovascular disease. Hyperuricemia, can lead to monosodium urate crystal deposition and thereby cause gout, the most common type of inflammatory arthritis among adults. Recent GWAS have found common genetic variants of *SLC2A9* to be associated with increased serum urate level and gout. The *SLC2A9* gene encodes a high-capacity urate transporter in humans which is mainly expressed in kidneys, liver and intestine. Loss of function variants were previously identified in hypouricemia. Bipolar I disorder is characterized by episodes of manic and hypomanic activity, and it has one of the highest rates of serious impairment among mood disorders. Iloperidone is a second-generation antipsychotic approved by the FDA that has anti-manic effects. Here we report genetic association of rs7442295 with urate levels at baseline and in interaction with iloperidone In a placebo-controlled trial of patients with bipolar mania, treatment with iloperidone 24mg/day resulted in a small but statistically significant increase of serum urate levels of approximately 27.2 μmol/L (0.457 mg/dL) compared to 0.1 μmol/L (0.002 mg/dL) in placebo group. This observation led to an investigation to see if genetic information could be used to predict which patients may be most susceptible to increases in serum urate following iloperidone treatment. WGS was conducted using whole blood samples obtained from the study subjects. A pronounced increase of 40.1 μmol/L (0.674 mg/dL) was seen in iloperidone-treated patients homozygous for the for the rs7442295 (G) allele at the SLC2A9 gene, compared to a decrease of -16.86 μmol/L in the corresponding GG placebo group. Similar results were observed for iloperidone in a second study in schizophrenia patients supporting generalization of the results across patient populations. The results show iloperidone-associated increases in serum urate are greatest in patients who were homozygous (GG) for the rs7442295 (G) allele at the SLC2A9 gene. Among male patients with the GG genotype, serum urate concentrations frequently shifted to above the upper limit of normal for iloperidone-treated patients in comparison to placebo group. Overall, the mechanism of this iloperidone-induced increase in serum urate levels is likely due to decrease in clearance of urate through interaction with the SLC2A9 urate transporter protein.

Session Title: Pharmacogenomics Poster Session I

PB2132 Association between Genetic Polymorphisms and New-Onset Diabetes in Korean Patients Treated with Statins

Authors:

Y-A. Park[1], D. Lee[1], J. Kim[1], T-J. Song[2], H. Gwak[1]; [1]Coll. of Pharmacy and Graduate Sch. of Pharmaceutical Sci., Ewha Womans Univ., Seoul, Korea, Republic of, [2]Dept. of Neurology, Seoul Hosp., Ewha Womans Univ. Coll. of Med., Seoul, Korea, Republic of

Introduction: Statins are lipid-lowering drugs used to prevent atherosclerotic cardiovascular disease and treat dyslipidemia. However, recent studies revealed that statin therapy could develop the risk of new-onset diabetes mellitus. The mechanisms of statin-induced diabetes, including effect of gene polymorphism, are still not precise. Therefore, we analyzed the relationship between genetic polymorphisms and new-onset diabetes mellitus in patients on statin therapy. **Methods:** Four single nucleotide polymorphisms (SNPs) of *FTO* (rs8050136, rs1121980, rs7206629, and rs9941349), three SNPS of *CDKAL1* (rs10946398, rs7756992, and rs2206734), *IGF2BP2* rs4402960, *SLC30A8* rs13266634, *HHEX* rs5015480, and *CDKN2A/2B* rs10811661 were analyzed by a TaqMan genotyping assay. Univariate and multivariable logistic regression analyses were performed to investigate the associations between genetic polymorphisms and statin-induced diabetes mellitus. **Results:** A total of 597 Korean patients on statins were included in the analysis, among whom 59 developed new-onset diabetes. Demographic factors of body mass index (≥30), smoking, and co-medications of diuretics increased the risk of new-onset diabetes mellitus in the multivariable analysis. Patients with *CDKAL1* rs10946398 variant allele (C) carriers and *FTO* rs8050136 variant allele (A) carriers developed 4.0-fold (95% CI: 1.63-9.67) and 2.2-fold (95% CI: 1.22-4.00) higher risk of diabetes than those with wild-type allele carriers, respectively, after adjusting for confounders. The Hosmer-Lemeshow test for statin-induced diabetes showed a good fit for the model ($\chi^2 = 2.902$, 5 degrees of freedom, and $p = 0.715$). The AUROC value of multivariable analysis model was 0.741 (95% CI: 0.68-0.81, $p < 0.001$).**Conclusion:** The study showed that *CDKAL1* and *FTO* polymorphisms could be associated with the risk of new-onset diabetes mellitus in patients on statin therapy.

Session Title: Pharmacogenomics Poster Session II

PB2133 Association of *TNFRSF1B*with nonresponse to infliximab in treatment of ulcerative colitis.

Authors:

L. Tessier[1,2], A-L. Gagnon[2], S. St-Amour[1,3], M. Côté[2], C. Allard[3], M. Durand[4], D. Bergeron[4], A. Lavoie[5], A. Michaud-Herbst[6], K. Tremblay[1,2,3]; [1]Pharmacology–Physiology Dept., Université de Sherbrooke, Sherbrooke, QC, Canada, [2]Res. Ctr. of Ctr. intégré universitaire de santé et de services sociaux du Saguenay–Lac-Saint-Jean (CIUSSS-SLSJ, Univ. Hosp.), Saguenay, QC, Canada, [3]Ctr. for Res. at the Ctr. Hosp.ier Univ.ire de Sherbrooke (CRCHUS), Sherbrooke, QC, Canada, [4]RNomics platform, Université de Sherbrooke, Sherbrooke, QC, Canada, [5]Pharmacy Dept., Ctr. Intégré Univ.ire de Santé et de Services Sociaux du Saguenay–Lac-Saint-Jean (Univ. Hosp.),, Saguenay, QC, Canada, [6]Gastroenterology Dept., Ctr. Intégré Univ.ire de Santé et de Services Sociaux du Saguenay–Lac-Saint-Jean (Univ. Hosp.), Saguenay, QC, Canada

Introduction: Ulcerative colitis (UC) is a chronic inflammatory bowel disease (IBD) characterized by continuous colonic inflammation. Worldwide, the disease's incidence is increasing in industrial countries. In Canada, ~120,000 people suffer from UC. For patients with severe forms of UC and for whom conventional therapies cannot induce or maintain remission, the last option before surgery is biologic therapies, including tumor necrosis factor inhibitors (anti-TNF; infliximab, adalimumab, golimumab). One of the main issues with these therapies is the high variability in treatment response. Indeed, ~20-50% anti-TNF users don't improve following drug initiation and 50-90% suffer from adverse events (AE). Multiple factors influence treatment response, such as age, sex, tobacco use and genetics. Indeed, genetic variants have already been associated with phenotypes of response to anti-TNF in Crohn's disease (another form of IBD) in genes part of the pharmacodynamic of anti-TNFs. The hypothesis was made that these variants may also affect response to anti-TNF in UC. Therefore, the goal of this study is to verify the association of these candidate variants with phenotype of response to anti-TNF in UC treatment. Method: This association study included 76 participants suffering from UC and who had taken or were taking at least one anti-TNF. Each recruited participant came to the research centre for a sole visit. Medical charts were reviewed to establish phenotype of treatment response using endoscopic scores. Blood or saliva samples were collected to extract DNA and to genotype 8 selected candidate variants in genes *TNF*, *TNFAIP3*, *TNFRSF1A* and *TNFRSF1B*. Statistical analyses were done using R software. Results: In the combined cohort of anti-TNF users, 30% of individuals were non-responders, 70% suffered from AE and no variant was found to be associated with the response's phenotype. However, a sub-group analysis in infliximab users (n=44) showed that *TNFRSF1B* rs1061622 variant was associated with nonresponse to infliximab (p-value=0.028). Other co-variables were tested, and total number of biologics taken throughout UC treatment was associated with nonresponse to infliximab (p-value=<0.0001).Conclusion: This study identified the first association between *TNFRSF1B* rs1061622 and non-response to infliximab in a cohort of UC patients. Next steps are to replicate this association in independent cohorts and to perform functional studies to elucidate by which mechanisms this associated variant may influence infliximab response, which will increase evidence level of this variant so it can eventually become a clinical marker of treatment response in UC.

Session Title: Pharmacogenomics Poster Session III

PB2134 Balancing cost and effectiveness in development of a sequencing panel for Pharmacogenomics variant identification

Authors:

M. Carter[1], Y. Fan[2], W. Wang[1], C. Goswami[1], J. Otto[2], L. Turner[1], C. Mozdzierz[2], J. Rosenfeld[3], H. Latif[2], G. Zhou[2]; [1]Azenta Life Sci., South Plainfield, NJ, [2]Azenta Life Sci., South Plainfield, NJ, [3]Tonix Pharma, Chatham, NJ

Pharmacogenomics (PGx) has become a staple in therapeutic development programs providing insights into the molecular underpinnings of drug responsiveness and clinical safety and efficacy. Advancements in genomics over the past 25 years have greatly given us a wide variety of tools for analyzing the genome of patients. Currently, costs are getting to a low enough level that we are able to sequence every individual in a clinical trial. These genomic findings are integrated with the trial results to identify PgX markers for that drug. Even with the current cost of a whole genome of less than $1,000 this is a significant expense for a large trial of thousands of patients. Here we apply recent advancements in NGS on clinical trial samples from Tonix Pharmaceuticals as well as standard human control samples to assess which methods, platforms, analysis approaches and combinations thereof offer the best outcomes from a performance, cost, and scalability perspective. Standard whole genome sequencing was performed using the latest generation of short read sequencing platforms (Illumina NovaSeq X+) and long read sequencing platforms (PacBio Revio). In addition, target capture-based approaches using PGx-specific enrichment panels from Twist were also performed on short and long read sequencing platforms. These methods were analyzed using a variety of bioinformatics tools independently and in combination to assess performance against truth data across ADMET genes including SNPs and complex genes such as CYP2D6 and HLA genes. The data generated here provide an outlook on how NGS adds value and holds promise in the next evolution of PGx technologies

Session Title: Pharmacogenomics Poster Session I

PB2135 † Breast cancer drugs that inhibit TOP2, including anthracyclines, induce cardiotoxicity through shared mechanisms in iPSC-derived cardiomyocytes

Authors:

E. Matthews[1], O. D. Johnson[2], K. J. Horn[3], J. Gutierrez[1], S. Powell[4], M. Ward[1]; [1]Univ. of Texas Med. Branch-Galveston, Dept. of Biochemistry and Molecular Biology, Galveston, TX, [2]Univ. of Texas Med. Branch, Biochemistry, Cellular and Molecular Biology Graduate Program, Galveston, TX, [3]Univ. of Texas Med. Branch, John Sealy Sch. of Med., Galveston, TX, [4]Univ. of Texas Med. Branch-Galveston, NeuroSci. Graduate Program, Galveston, TX

TOP2 inhibitors (TOP2i) used in the treatment of breast cancers, including the anthracyclines (AC) Doxorubicin (DOX), Daunorubicin (DNR), Epirubicin (EPI), and the anthraquinone Mitoxantrone (MTX) can lead to cardiotoxicity in some women. It is unclear whether women would experience the same adverse effects from all drugs in this class. To gain insight into whether specific drugs would be preferable for certain individuals based on their cardiotoxicity risk profile, we studied the effects of treatment of DNR, DOX, EPI, MTX, and an unrelated monoclonal antibody Trastuzumab (TRZ) on iPSC-derived cardiomyocytes (iPSC-CMs) from six females. While TRZ does not affect cell viability, all TOP2i induce cell death at concentrations observed in cancer patient serum. To elucidate primary responses due to the drug treatment, we chose a sub-lethal dose to measure effects over 24 hours. All TOP2i affect calcium handling, a function critical for cardiomyocyte contraction. Global gene expression data show that drug treatment and time accounts for most variation between samples, followed by the individual from which samples were derived. Thousands of genes respond after treatment with each of the TOP2i. We did not observe a gene expression response to TRZ. Response genes are associated with the p53 signaling pathway, DNA replication, and the cell cycle. Bayesian analysis across all drugs reveals four gene expression signatures, which we denote as TOP2i early, late, time-independent, and non-response genes. TOP2i early response genes are enriched in chromatin regulators, which have been implicated in mediating AC sensitivity in breast cancer patients. To investigate genetic effects on drug response, we first identified a reported set of eQTLs that are uncovered specifically following DOX treatment in iPSC-CMs. DOX response eQTLs are enriched in genes that respond to all TOP2i compared to baseline eQTLs. Of the 180 DOX response QTLs that respond to DOX in our cells, 93% also respond to DNR, 85% to EPI, and 20% to MTX. We next identified eight genes in loci associated with AC toxicity by GWAS or TWAS. Five of these genes respond to all ACs, including *RARG* and *HDDC2*. Finally, we investigated whether genes near SNPs associated with cardiovascular disease respond to TOP2i. DNR, EPI, and MTX response genes are enriched for atrial fibrillation genes, while no drug response genes are enriched for heart failure genes. Our data thus demonstrate that TOP2i induce thousands of shared gene expression changes in cardiomyocytes, including genes near SNPs associated with inter-individual variation in response to DOX treatment, AC-induced cardiotoxicity, and atrial fibrillation.

Session Title: Pharmacogenomics Poster Session II

PB2136 Cannabinoid 1 receptor (CNR1) 1359G>A polymorphism associated with psychotic adverse events in cannabis users with chronic pain

Authors:

K. Tremblay[1], W. Beauchesne[1], J. Turcotte[2], A-L. Gagnon[2], P. Mercier[2], M. Durand[3], D. Bergeron[3], G. Léonard[4], L. Gendron[1]; [1]Pharmacology-physiology Dept., Université de Sherbrooke, Sherbrooke, QC, Canada, [2]Res. Ctr. of Ctr. intégré universitaire de santé et de services sociaux du Saguenay–Lac-Saint-Jean (CIUSSS-SLSJ, Univ. Hosp.), Saguenay, QC, Canada, [3]RNomics platform, Université de Sherbrooke, Sherbrooke, QC, Canada, [4]Readaption Sch., Université de Sherbrooke, Sherbrooke, QC, Canada

Chronic pain is one of the leading causes of disability worldwide. Despite its high prevalence and the diverse pharmacologic therapies available, patients with chronic pain often remain with significant disability and inadequate pain control. Cannabis and cannabinoids are sometimes used in the treatment of chronic pain as they have been shown to be useful in some patients. However, they are associated with significant adverse effects. Notably, the occurrence of cannabis use disorder (CUD) and cannabis-induced psychosis, which have been associated with many genetic variants in the literature. However, the paucity of data on some reported variants and contradictory results limit our ability to use them as genetic markers to personalize cannabis treatment for patients tailored to their genetic background.

The aim of this genetic association study was to characterize the phenotypes of patients with chronic pain who were using or had used cannabis or cannabinoids in the past and to investigate the effects of previously reported genetic variants.

Phone interviews were conducted to document participants' characteristics, cannabis use and its effects, concurrent pharmacotherapy and comorbid conditions. CUD was screened for using the Cannabis Use Disorders Identification Test - Revised questionnaire. Blood or saliva samples were collected for genotyping and 19 variants on 12 candidate genes (*BDNF*, *CHRM3*, *CNR1*, *CNR2*, *COMT*, *CYP2C9*, *FAAH*, *GABRA2*, *HES7*, *NRG1*, *OPMR1*, *KAT2B*) were tested.

This study recruited 100 participants and DNA were obtained in 77 participants. We report one significant single nucleotide variation (rs1049353) in CNR1 (cannabinoid receptor 1) gene associated with psychotic adverse events. Homozygous patients for *CNR1* c.1359G>A presented higher rates of psychotic adverse events ($p = 0.008$). No variants were associated with CUD or pain relief. These results provide novel evidence of association between *CNR1* rs1049353 and psychotic adverse events but did not replicate many previously documented genetic associations for CUD. Our results also underscore the need for replication studies using adequate sample size and functional studies to elucidate by which mechanisms this associated variant may influence cannabis response.

Session Title: Pharmacogenomics Poster Session III

PB2137 Cardiometabolic drug response pharmacogenetics using EHRs from the UK Biobank

Authors:

M. Sadler[1], E. R. Pearson[2], R. Altman[3], Z. Kutalik[1]; [1]Univ. of Lausanne, Lausanne, Switzerland, [2]Div. of Population Hlth.and Genomics, Sch. of Med., Univ. of Dundee, Dundee, United Kingdom, [3]Stanford Univ., Menlo Park, CA

Despite the massive progress of genome-wide association studies (GWAS), pharmacogenetics (PGx) is advancing at a slower pace. Here, we extracted clinical and medication prescription data from electronic health records (EHRs) of 200,000 individuals of European ancestry in the UK Biobank and conducted PGx GWAS on the change in biomarkers following cardiometabolic drug therapy. GWAS of LDL cholesterol (LDL-C) response to statins (N = 15,356) identified genome-wide significant hits at *PCSK9* (P = 9.1e-12), *SORT1*/*CELSR2*/*PSRC1* (P = 7.5e-11), *LPA* (P = 2.8e-18) and *APOE* (P = 1.5e-81) which have been previously reported. GWAS of total cholesterol response to statins (N = 24,810) identified an additional locus at *CETP* which coincided with the HDL-C response locus (N = 21,908). No genome-wide significant signals were identified for HbA1c response to metformin (N = 2,890) and SBP response to major first-line antihypertensives, combined and individually (ACE inhibitors, calcium channel blockers, thiazide diuretics, N = 1,281-5,835). Since the lipid response hits (h^2_{SNP} = 8.2%) overlapped with variants impacting baseline lipid levels (h^2_{SNP} = 8.0%; genetic correlation of 0.29), we adjusted the LDL-C response GWAS for LDL-C baseline level and the polygenic risk score (PRS) of LDL-C, leaving only the *LPA* and *APOE* loci significant. It also revealed that despite higher baseline being associated with more reduction, 1SD increase in PRS leads to 0.11 mmol/L less LDL-C reduction (P = 1.2e-72). Similarly, SBP-response to antihypertensives was also attenuated for those with higher genetic predisposition (0.70 mmHg less reduction / SD increase in SBP PRS, P = 3.5e-4), but the trend was not present in the HbA1c response to metformin. Control experiments (using statin-naïve samples) showed a similar trend: Individuals with higher baseline PRS had less longitudinal decrease in LDL-C levels - indicating that part of the drug response signal is due to regression-to-the-mean. An LDL-C change GWAS in the same control cohort (N = 27,159) also identified *APOE* as the top hit, however, with a significantly lower effect size (b = 0.15) than in the statin response GWAS (b = 0.38), further evidencing the presence of both drug-specific and regression-to-the-mean effects. In summary, we present a comprehensive catalog of cardiometabolic PGx GWAS and showed that drug response genetics comprises both a drug-and disease-specific component. Our study demonstrates that EHRs enable new opportunities to study drug response to reveal the complex contribution of the genetic and environmental components of the baseline level of the targeted cardiometabolic biomarker.

Session Title: Pharmacogenomics Poster Session I

PB2138 Characterisation of pharmacogenetic variation relevant to anti-tuberculosis drug metabolism across African populations.

Authors:

T. Malinga[1,2], H. Othman[2], D. Twesigomwe[2,1]; [1]Div. of Human Genetics, Natl. Hlth.Lab. Service, and Sch. of Pathology, Faculty of Hlth.Sci., Univ. of the Witwatersrand, Johannesburg, South Africa, [2]Sydney Brenner Inst. for Molecular BioSci., Faculty of Hlth.Sci., Univ. of the Witwatersrand, Johannesburg, South Africa

Background and aims: Tuberculosis (TB) is a major health burden in Africa. Although TB is treatable, current medications (e.g. isoniazid) are associated with adverse drug reactions such as hepatotoxicity. These ADRs can partly be attributed to pharmacogenetic variation. The distribution of star alleles (haplotypes) that influence anti-TB drug metabolism, is unknown in many African ethnolinguistic groups. This presents challenges and limitations in implementing genotype-guided therapy in African clinical settings for purposes of anti-TB drug dosage optimisation with a view toward promoting safe and efficacious TB treatment. Therefore, this study aimed to characterise the distribution of star alleles in genes that are involved in anti-TB drug metabolism, namely *CYP2E1*, *NAT1*, *NAT2*, *GSTM1* and *GSTT1*, across diverse African populations. **Methods:** This study used 802 high-depth whole genome sequence datasets representative of eight Sub-Saharan African (SSA) population groups. Data sources included the 1000 Genomes Project and H3Africa AWI-Gen. *CYP2E1*, *NAT1*, *NAT2*, *GSTM1* and *GSTT1* star alleles were called from the WGS data using StellarPGx. Novel star allele-defining variants were annotated using the Ensembl Variant Effect Predictor. **Results:** We present both common and rare star alleles influencing anti-TB drug metabolism across various SSA populations, in comparison to other global populations. Among the SSA population, relatively common key star alleles were identified such as *NAT1*4*, *NAT1*10*, *GSTM1*0 (*deletion), and *GSTT1*0* (deletion) which had frequencies above 50%, while *NAT2*5B* had a frequency of 22%. Seventeen novel haplotypes were identified for these genes across SSA. Among these, three novel alleles were present at frequencies >1% (e.g. *GSTT1*A*+rs2266637(T)+rs8140585(T) was observed at a frequency of 10%), while others were relatively rare. Regarding anti-TB drug metaboliser phenotype predictions, NAT2 intermediate and slow acetylation phenotypes were the most frequent in the SSA populations in this study. **Conclusion:** This study provides insight into the distribution of star alleles relevant to anti-TB drug metabolism across various African populations, and in comparison, to global populations. Furthermore, it provides a foundation of implementing pharmacogenetic testing in Africa, to reduce risk of ADRs related to TB treatment. Future work entails characterizing the functional impact of novel alleles identified in this study and submitting them to the Pharmacogene Variation Consortium database for designation.

Session Title: Pharmacogenomics Poster Session III

PB2140 Clinically actionable pharmacogenomic landscape of psychotropic medications in a Middle Eastern population.

Authors:

P. Jithesh[1], D. Velayutham[1], K. Bastaki[2], S. Ghuloum[3], M. Azeem[4], Qatar Genome Program Research Consortium; [1]Hamad Bin Khalifa Univ., Doha, Qatar, [2]Qatar Univ., Doha, Qatar, [3]Hamad Med. Corp., Doha, Qatar, [4]Sidra Med., Doha, Qatar

Background: Psychotropic medications are prescribed widely worldwide. However, their clinical response is highly varied due to several factors including the individual's genetic variants. Understanding the distribution of clinically actionable genotypes and their predicted response phenotypes is essential in facilitating clinical implementation of pharmacogenomics in psychiatry. However, such reports are missing from several populations, including the Middle East. **Methods:** We used whole genome sequencing data from a cohort of 14,392 adult Qataris from the Qatar Genome Program (QGP) to study the frequency distribution of actionable diplotypes (a combination of variants in the gene on both the homologous chromosomes) and their predicted phenotypes. A total of 490 alleles in four genes coding for metabolizing enzymes that significantly affect response to antidepressants and antipsychotics, and have guidelines for clinical interpretation, namely *CYP2C19, CYP2D6, CYP2B6* and *CYP3A4*, were analyzed. Clinically actionable pharmacogenomic variance was defined based on guidelines from the Clinical Pharmacogenetic Implementation Consortium (CPIC) and the Dutch Pharmacogenetics Working Group (DPWG) suggesting a course of action, such as change in dosage or prescription of an alternate drug. **Results:** Following the DPWG guidelines, the clinical actionability of antipsychotics was in the range of 0.1 - 32%, based on *CYP3A4* and *CYP2D6* diplotypes . However, we found that more than 52% of the individuals may have actionable metabolizer phenotype associated diplotypes based on CPIC guideline in *CYP2C19, CYP2B6* and *CYP2D6* affecting response to tricyclic antidepressants, while for serotonin reuptake inhibitors, the frequency varied from ~2 - 51%. For example, CYP2C19 ultrarapid, rapid, poor and intermediate metabolizers need alteration in dosage or alternate prescription, and constituted 51% of the population. **Conclusion:** Our previous research has identified heavy prescription of antidepressants, especially escitalopram in Qatar. Therefore, our results from this study highlight the importance of implementation of pharmacogenomic testing prior to prescribing these medications to prevent unnecessary hospitalizations and inefficient treatment regimens, reducing the associated economic burden, and improving the quality of life of the patients. We will present the detailed results for several drugs and discuss their implications.

Session Title: Pharmacogenomics Poster Session I

PB2141 ClinPGx: An Integrated Single Resource for Pharmacogenomics (PGx)

Authors:

T. Klein[1], K. Caudle[2], A. Gaedigk[3], M. D. Ritchie[4], S. A. Scott[5], L. Gong[6], K. Sangkuhl[5], C. F. Thorn[1], C. Klein[1], I. M. Keseler[1], R. Whaley[5], M. Woon[5], I. Reinert[1], M. W. Wright[7], M. Whirl-Carrillo[5]; [1]Stanford Univ, Stanford, CA, [2]St. Jude Children s Res. Hosp., Memphis, TN, [3]Children s Mercy Kansas City, Washington, MO, [4]Univ. OF PENNSYLVANIA, Philadelphia, PA, [5]Stanford Univ., Stanford, CA, [6]Stanford Univ, Palo Alto, CA, [7]Stanford Univ. Sch. of Med., Palo Alto, CA

Pharmacogenomics (PGx) is the study of how a person's genetic makeup affects their response to drugs, with the goal of helping clinicians personalize medication management and improve patient outcomes. However, the current landscape of PGx resources is fragmented, resulting in reporting inconsistencies, duplication of efforts, and unintended barriers to implementation. To address these challenges, we propose a new long-term and conceptual framework for broadly integrating the available PGx resources into a single resource, ClinPGx. ClinPGx would complement ClinGen and ClinVar by providing essential PGx information to clinical providers and the research community.

Standardization is a critical need in the field of PGx, as inconsistencies between resources and laboratories pose significant challenges to implementation. Furthermore, barriers such as electronic health record (EHR) integration, provider education and awareness of PGx, insurance coverage, and communication gaps between laboratories and healthcare providers hinder the widespread adoption of PGx. ClinPGx will address these challenges by leveraging existing tools and resources (e.g., PharmGKB, CPIC, PharmCAT and PharmVar) to streamline efforts and ensure compatibility and reporting consistencies, facilitating integration of PGx into EHR systems, and increasing stakeholder buy-in and confidence in the scientific foundation of PGx testing and interpretation. This unified resource may also catalyze expansion of implementation and insurance coverage, increasing access to PGx testing and potentially reducing health disparities. One key aspect of this integration involves expanding the ClinGen Community Curation efforts to include PGx data, enabling the collective expertise of the community to enhance the resource's comprehensiveness and accuracy.

In conclusion, the development of a unified PGx resource represents a transformative step towards realizing the full potential of PGx, which would span from discoveries to clinical implementation. By engaging the community, leveraging existing tools, and addressing the challenges of standardization and implementation, ClinPGx will enable clinicians to make informed decisions based on individual genetic profiles, and improve patient outcomes. We will actively solicit feedback from the genomics community on the interest and development of ClinPGx to facilitate the incorporation of PGx into standard of care.

We would like to acknowledge support from the following funding agencies: NIH, NHGRI, NICHD and NIDA (PharmGKB (U24 HG010615); CPIC (U24HG010135); PharmCAT (U24HG010862), ClinGen (Baylor/Stanford U24HG009649)).

Session Title: Pharmacogenomics Poster Session III

PB2143 DDINER: Drug-Drug Interactions classification Neural network model using Random walk with restart algorithm.

Authors:

J. Seo, Y. Ko; Hankuk Univ. of Foreign Studies, Seoul, Korea, Republic of

Drug-Drug interactions (DDIs) occurs due to the unexpected pharmacological effects when multiple drugs are prescribed simultaneously. While short-term administration of multiple drugs can enhance efficacy, it can lead to severe adverse reactions. The prescription of multiple drugs relies on the clinician's experience or knowledge, lacking a standardized database for safe co-prescription. It is crucial to identify potential DDIs in advance for patient safety and treatment strategies. However, it is not feasible to experimentally identify all existing drug interactions. Instead, drug-drug interaction can be identified with computational methods by utilizing various drug features such as chemical structure, target genes, or associated pathological pathways. However, DDIs usually result from unexpected associations of target proteins or shared mechanisms and such indirectly hidden associations should be considered to predict DDIs.In this study, we proposed DDINER a novel deep learning model to predict 80 types of DDIs by utilizing various drug features such as chemical structures and drug-binding proteins. The model's training and testing were conducted using interaction data of FDA-approved drugs obtained from Drugbank, resulting in 163,574 DDIs involving 1,436 drugs. We used chemical structure and binding protein of drugs as an representation of drug. We applied the random walk with restart (RWR) algorithm to propagate drug-binding proteins across a PPI network. Our experimental results demonstrated the effectiveness of DDINER for the prediction of clinically proven epilepsy drug combinations as well as the unknown drug pairs. DDINER would be very useful for the identification of DDIs and the pharmacological mechanisms associated with DDIs. Our method could apply to all possible drug pairs and the predicted potential DDI can be used clinically with experimental validations, leading to the way to drug repurposing.

Session Title: Pharmacogenomics Poster Session I

PB2144 Determining the role of TOP2B and p53 in mediating Doxorubicin-induced cardiotoxicity in iPSC derived-cardiomyocytes

Authors:

S. Paul, J. Gutierrez, M. Ward; Univ. of Texas Med. Branch Galveston, Galveston, TX

The Anthracycline Doxorubicin (Dox), which is prescribed in 32% of breast cancer patients, can cause irreversible left ventricular dysfunction and heart failure. Dox intercalates into DNA forming a ternary complex with TOP2B. This complex induces DNA double-strand breaks, mitochondrial dysfunction and apoptosis by activating the p53 transcription factor. TOP2B is essential for the cardiotoxicity observed in mice. Under physiological conditions, TOP2B regulates DNA topology and interacts with CTCF and cohesin at topologically associated domain boundaries. However, it is unknown how these functions are affected by Dox in human hearts. To gain insight into the mechanisms behind Dox-induced cardiotoxicity, we have developed a human iPSC-derived cardiomyocyte (iPSC-CM) model to study the binding of these transcriptional regulators in response to Dox treatment. We first performed chromatin immunoprecipitation followed by sequencing (ChIP-seq) for p53, TOP2B, CTCF, and RAD21 (a cohesin subunit) in iPSC-CMs. We find thousands of genomic regions that are bound by each of these factors. 32% of TOP2B-enriched regions are co-bound by CTCF and RAD21 suggesting roles for TOP2B in gene regulation and chromatin organization in cardiomyocytes as has been observed in other cell types. To determine transcriptional regulator binding in response to Dox, we exposed iPSC-CMs to a non-lethal dose of Dox that induces thousands of gene expression changes in 24 hours compared to the vehicle control. ChIP-qPCR reveals that p53 binding is increased at the CDKN1A and MDM2 promoters in Dox-treated iPSC-CMs compared to vehicle, which suggests that Dox treatment leads to the activation of the p53 stress response pathway in these cells. Using this model, we are currently performing ChIP-seq experiments for all factors in Dox-treated iPSC-CMs to determine the gene regulatory networks activated in response to Dox that may explain the cardiotoxic effects of drug treatment.

Session Title: Pharmacogenomics Poster Session II

PB2145 Differential gene expression and regulation at single nucleus resolution in muscle of statin users

Authors:

B. Wolford[1], S. C. Hanks[2], L. Taylor[3], A. U. Jackson[2], A. Varshney[2], N. Manickam[2], A. Ramirez[3], A. T. Beck[2], H. M. Stringham[2], P. Orchard[2], N. Narisu[3], L. Bonnycastle[3], M. Erdos[3], M. Sweeney[2], J. Tuomilehto[4], M. Melbye[5], T. Lakka[6], M. Laakso[6], H. Koistinen[4], F. Collins[3], M. Boehnke[2], S. C. J. Parker[2], L. J. Scott[2]; [1]Norwegian Univ. of Sci. and Technology, Trondheim, Norway, [2]Univ. of Michigan, Ann Arbor, MI, [3]NHGRI/NIH, Bethesda, MD, [4]Finnish Inst. for Hlth.and Welfare, Helsinki, Finland, [5]Danish Cancer Inst., Copenhagen, Denmark, [6]Univ. of Eastern Finland, Kuopio, Finland

Statins are prescription drugs which reduce plasma LDL cholesterol and are important for preventing and treating cardiovascular disease and preventing premature mortality. Based on 2013 American Heart Association and American College of Cardiology guidelines, 48.6% of US adults are eligible for statins. Adverse effects such as muscle pain and weakness and even rhabdomyolysis can occur in statin users, often leading to poor adherence. The mechanism for these adverse effects is poorly understood. We hypothesize that identification of genes differentially expressed or regulated in skeletal muscle of statin users may help explain these effects.

We performed transcriptome and epigenome data analyses of muscle biopsies from 281 participants of the FUSION Tissue Biopsy Study to characterize the effects of statin use on skeletal muscle. Using bulk RNA-seq combined with single nucleus RNA-seq and ATAC-seq, we tested for differences in gene expression and open chromatin status between 48 statin (3-hydroxy-3-methyglutaryl-coenzyme A reductase inhibitors) users and 233 non-users, controlling for oral glucose tolerance test status, age, sex, and technical covariates.

Using snRNA-seq, we tested for associations between gene expression and statin use for each of 12 observed muscle cell types and subcellular components with DESeq. At 5% FDR, we observed higher expression of TRIT1 in statin users in mesenchymal stem cells and lower expression of PGGT1B in statin users in smooth muscle. Notably, PGGT1B has been associated with quantitative traits related to muscle performance in elite athletes. Gene Ontology term enrichment analysis of differentially expressed genes across cell types showed that Type 1 muscle fibers have upregulated pathways in statin users (e.g., proteasome, endopeptidase, and peptidase complexes). We observed significant differential expression in bulk RNA-seq in ABCA1 and ASPA.

Composition of muscle cell types was also significantly associated with statin use. Combining snRNA-seq and snATAC-seq reads, the number of nuclei was higher in statin users in both the neuromuscular junction (fold change 1.5 [95% CI 1.2-1.9], p-value 1E-4) and smooth muscle (fold change 1.2 [95% CI 1.1-1.4], p-value 1E-3). At 5% FDR, smooth muscle and mesenchymal stem cells had regions of differentially accessible chromatin. In mesenchymal stem cells, a differentially open region was 50 kilobases downstream of ACTR3, a gene previously associated with myopathy.

Ultimately, characterization of the impact of statins on skeletal muscle may lead to preventive measures for adverse effects, increased pharmaceutical adherence, and reduced cardiovascular events.

Session Title: Pharmacogenomics Poster Session III

PB2146 Effects of Genetic Polymorphisms of Drug Metabolizing Enzymes and co-Medications on Tamoxifen Metabolism in Black South African Women with Breast Cancer

Authors:

S. Chiwambutsa[1,2], O. Ayeni[3,4], N. Kapungu[5], C. Kanji[5], R. Thelingwani[5], W. Chen[6], D. Mokone[7], D. O'Neil[8], A. Neugut[9,10,11], J. Jacobson[10,11], P. Ruff[3,12,13], H. Cubasch[3,12,14], M. Joffe[3,12,15], C. Masimirembwa[6,5]; [1]Div. of Human Genetics, Natl. Hlth.Lab. Service, and Sch. of Pathology, Faculty of Hlth.Sci., Univ. of the Witwatersrand, Johannesburg, South Africa, [2]Sydney Brenner Inst. for Molecular BioSci., Johannesburg, South Africa, [3]Strengthening Oncology Services Res. Unit, Faculty of Hlth.Sci., Univ. of the Witwatersrand, Johannesburg, South Africa, [4]Dept. of Radiation Oncology, Dept. of Med., Faculty of Hlth.Sci., Univ. of the Witwatersrand, Johannesburg, South Africa, [5]African Inst. of BioMed. Sci. and Technology (AiBST), Harare, Zimbabwe, [6]Sydney Brenner Inst. for Molecular BioSci., Faculty of Hlth.Sci., Univ. of the Witwatersrand, Johannesburg, South Africa, [7]Dept. of Surgery, Sefako Makgatho Hlth.Sci. Univ., Dr George Mukhari Academic Hosp., Ga-Rankuwa, Johannesburg, South Africa, [8]Comprehensive Cancer Ctr. and Dept. of Med., Miller Sch. of Med., Univ. of Miami, Miami, FL, [9]Dept. of Med., Vagelos Coll. of Physicians and Surgeons, Columbia Univ., Columbia, NY, [10]Herbert Irving Comprehensive Cancer Ctr., Vagelos Coll. of Physicians and Surgeons, Columbia Univ., Columbia, NY, [11]Dept. of Epidemiology, Mailman Sch. of Publ. Hlth., Columbia Univ., Columbia, NY, [12]South Africa Med. Res. Council Common Epithelial Cancers Res. Ctr., Faculty of Hlth.Sci., Univ. of the Witwatersrand, Johannesburg, South Africa, [13]Div. of Med. Oncology, Univ. of the Witwatersrand, Johannesburg, South Africa, [14]Dept. of Surgery, Faculty of Hlth. Sci., Univ. of the Witwatersrand, Johannesburg, South Africa, [15]SAMRC/Wits Dev.al Pathways to Hlth.Res. Unit, Dept. of Pediatrics, Faculty of Hlth.Sci., Univ. of the Witwatersrand, Johannesburg, South Africa

Clinical outcomes of tamoxifen treatment show wide inter-individual variability. Co-medications and genetic polymorphisms of enzymes involved in tamoxifen metabolism contribute to this variability. Drug-drug and drug-gene interactions have seldom been studied in African black populations. We evaluated the effects of commonly co-administered medicines on tamoxifen pharmacokinetics in a cohort of 229 South African black female hormone-receptor positive breast cancer patients. We also investigated the pharmacokinetic effects of genetic polymorphism in enzymes involved in tamoxifen metabolism including the variants *CYP2D6*17* and **29* which have been mainly reported in people of African ancestry. Tamoxifen (TAM) and its major metabolites, N-desmethyltamoxifen (NDM), 4-OH-tamoxifen (4-OHTAM), and endoxifen (ENDO), were quantified in plasma using the liquid chromatography-mass spectrometry (LC/MS/MS). The GenoPharm® open array was used to genotype *CYP2D6*, *CYP3A5*, *CYP3A4*, *CYP2B6*, *CYP2C9*, and *CYP2C19*. Results showed that *CYP2D6* diplotype and *CYP2D6* phenotype significantly affected endoxifen concentration ($p<0.001$ and $p<0.001$). *CYP2D6*17* and *CYP2D6*29* significantly reduced the metabolism of NDM to ENDO. Antiretroviral therapy (ART) had a significant effect on NDM levels and the TAM/NDM and NDM/ENDO Metabolic Ratios (MRs) but did not result in significant effects on ENDO levels. In conclusion, *CYP2D6* polymorphisms affected endoxifen concentration and the variants *CYP2D6*17* and *CYP2D6*29* significantly contributed to low exposure levels of endoxifen. This study also suggests a low risk of drug-drug interaction in breast cancer patients on tamoxifen.

Session Title: Pharmacogenomics Poster Session I

PB2147 Enhancing Asthma Pharmacogenetics through Endotype-Specific Associations

Authors:

S. Piparia[1], K. Tantisira[2]; [1]Univ. of California at San Diego, San Diego, CA, [2]Univ. of California, San Diego, La jolla, CA

Background:Asthma is a complex disease with various subtypes, known as endotypes, characterized by distinct underlying pathophysiological mechanisms. Inhaled corticosteroids (ICS) are a primary treatment for asthma, but individual response to this medication varies significantly. Although genetic variations in T-genes have been associated with differences in ICS response, they are not sufficient as predictive biomarkers. This study aims to investigate whether endotype-based pharmacogenetics can enhance the potential of precision medicine for asthma.Methods:We conducted candidate gene analysis to examine whether the pharmacogenetic association is stronger when stratifying the Childhood Asthma Management Program (CAMP) cohort by eosinophil counts (threshold of 300 cells/uL), a commonly used delineation for type 2 asthma. We dichotomized the cohort based on ICS usage into high eosinophils (n=82) and low eosinophils (n=50) and performed genetic association analysis for rs3127412, a T-gene variant previously associated with ICS response, with change in lung function on ICS. Additionally, we computed the Area Under the Curve (AUC) for the upper and lower quartiles within both EOS groups and the entire cohort.Results:Using an additive model, we observed a 5.7% increase in percent predicted FEV1 (forced expiratory volume at one second) for the entire cohort (p = 0.003), while the FEV1 change (p-values) were 7.1% (p = 0.008) for high and -0.2% (p = 0.912) for low eosinophils groups respectively, with a significant interaction (p = 0.038). The AUC of the high Eos group demonstrated good predictive value (AUC 0.71) while there was no ability of this variant to predict outcomes in the low Eos group (AUC of 0.48). Replication analysis is currently ongoing. Our findings indicate that endotypes enhance the prognostic ability of medication response. These results suggest that these genetic variants may play a role in the response to changes in lung function in pediatric patients treated with ICS.

Session Title: Pharmacogenomics Poster Session II

PB2148 Evaluating cardiometabolic risk factors and treatment efficacy in diverse British ancestries: Insights from a novel application of drug target Mendelian randomisation

Authors:

A. Carter[1], C. Maroteau[1], J. Yu[1], J. Bowden[1,2], J. Howson[1], Y. Jamshidi[1]; [1]Novo Nordisk Res. Ctr. Oxford, Oxford, United Kingdom, [2]Univ. of Exeter, Exeter, United Kingdom

Coronary artery disease (CAD) poses a significant health burden on individuals across ancestries. Individuals of South Asian ancestry living in England have over a 70% increased risk of CAD compared with the White British population and the onset of disease tends to occur at earlier ages. Given these disparities, tailored early intervention and treatment strategies for diverse ancestries need to be identified.

Using a novel Mendelian randomization (MR) approach in different ancestries, we compared the effects of three CAD risk factors, triglycerides (TG), low density lipoprotein cholesterol (LDL-C), high density lipoprotein cholesterol (HDL-C) and two gene regions reflecting drug targets (*HMGCoA* [statins], *PCSK9*) on CAD in two British cohorts: UK Biobank (UKB) and Genes and Health (G&H).

Genome wide significant polygenic scores per standard deviation increase in exposure were derived from Global Lipids Genetics Consortium ancestry specific GWAS summary statistics. Genetic variants in the *HMGCoA* and *PCSK9* gene regions were weighted by their effects on LDL-C.

UKB analyses were conducted in multi-ancestry participants (N=420,3112; CAD=42,751) and stratified by European (N=396,198; CAD=40,004) and South Asian ancestry (N=8772; CAD=1428). All G&H participants were of British Bangladeshi and British Pakistani ancestry and analysed together (N=44,396; CAD=4037).

Two-stage least squares regression, adjusted for genetic principal components, was used to estimate MR causal effects. Estimates for South Asian individuals in UKB and G&H were meta-analysed.

Analyses suggested an increased odds ratio (OR) of TG and LDL-C on CAD in Europeans (TG = 1.55; 1.45-1.66; LDL-C = 1.79; 1.64-1.96) but not South Asians (TG = 1.11; 0.86-1.36; LDL-C = 1.03; 0.77-1.29). HDL-C trended towards a greater protective effect in Europeans (0.70; 0.66-0.75) than South Asians (0.87; 0.60-1.14).

The OR for the effect of *HMGCoA* on CAD (European = 1.29; 0.99-1.68; South Asian = 0.65; 0.12-1.19) was smaller than *PCSK9* (European = 1.94; 1.66-2.27; South Asian = 1.59; 0.13-3.07).

As ancestrally diverse GWAS become available, novel applications of MR to test causal effects across ancestries can be conducted. This analysis suggests the increased burden of CAD in individuals of South Asian ancestry in England is unlikely to be due to lipids. The wide confidence intervals in South Asian analyses indicate low power for drug-target MR and the need for ever larger ancestrally diverse samples. By using genetics approaches to identify differences in cardiometabolic risk profiles and treatment effectiveness, treatment strategies can be personalised, improving cardiovascular health worldwide.

Session Title: Pharmacogenomics Poster Session III

PB2149 Exploring pharmacogenetic relationships in Malaria and Tuberculosis: A Text-Mining analysis of biomedical literature and gene diversity assessment in African Populations

Authors:

H. Othman; Sydney Brenner Inst. for Molecular BioSci., Univ. of the Witwatersrand, Johannesburg, South Africa

Malaria and Tuberculosis continue to persist as significant health challenges in Africa, claiming over 500,000 lives annually for each of these diseases. The abundant genetic diversity present in Africa necessitates the acquisition of extensive knowledge about genes associated with drug-processing mechanisms. This, in turn, requires harnessing information from various data sources, including both public and non-public repositories such as biomedical literature and chemical structures of ligands/drugs. The detection of the drug-gene relationship is a crucial piece of information that can provide insights into potential pharmacogenetic associations. In order to explore this area, we conducted a text-mining analysis of the biomedical literature focused on Malaria and TB. Initially, we constructed a predictive model to identify articles containing information about pharmacogenetic relationships. To train the model, we utilized over 20,000 abstracts from PharmGKB and more than 40,000 randomly sampled abstracts from PubMed. Additionally, we created a comprehensive list of 36 anti-TB and anti-malaria drugs, along with candidate drugs. Using Biotransformer 3.0, we predicted the phase I metabolizers for these drugs. Our preliminary analysis of the biomedical literature yielded 5,194 sentences associated with drug-gene relationships for TB, and 5,714 sentences for malaria. These sentences include information on 1,370 unique drugs/ligands and 1,794 unique genes for malaria, as well as 1,459 unique drugs/ligands and 1,868 unique genes for TB. Within the identified unique genes, we pinpointed 85 genes for malaria and 96 genes for TB, out of a total of 340 pharmacogenes. Moreover, we successfully identified 9 candidate genes involved in the metabolization of 31 out of the 36 anti-malaria and anti-TB drugs/ligands. These candidate genes include CYP2C19, CYP3A4, CYP2D6, CYP2C9, CYP1A2, CYP2B6, CYP2A6, CYP2E1, and CYP2C8. Presently, we are conducting a gene diversity assessment using whole genome sequencing and DNA genotyping array datasets of 900 and 22,000 African subjects respectively.

Session Title: Pharmacogenomics Poster Session I

PB2150 Exploring the association between polygenic score of adverse drug reaction genes and disease risk

Authors:

J. Kim[1,2], S. Lee[1], J. Sung[1], S-S. Byun[3]; [1]Seoul Natl. Univ., Seoul, Korea, Republic of, [2]Procagen, Seongnam, Korea, Republic of, [3]Seoul Natl. Univ. Bundang Hosp., Seongnam, Korea, Republic of

Background: Identification of genetic variants associated with adverse drug reaction (ADR) is important for optimizing treatment efficacy and reducing the side effects. While many studies have focused on pharmacokinetics and pharmacodynamics genes for their effects on adverse reactions it is not clear whether those genes play a role in the risk of developing common diseases as well. Our study aimed to explore the diseases that may increase the risk in individuals with general genetic tendency to adverse drug reactions (ADR).Methods: We calculate a polygenic score for ADR (PGS-ADR) based on the group 1&2 pharmacogenes in the Pharmacogenomics Knowledge Base (PharmGKB). The effect size reported in the PharmGKB (about 80% of genes), or the functional deterioration score (CADD) was used as the weighting. A phenome-wide search was conducted for the PGS-ADR.Results: We identified an association between PGS-ADR and the drug toxicity phenotype in the European ancestry population. (odds ratio (OR)=1.21; 95% confidence interval (CI)=1.08-1.39) From the phenome-wide association search, we observed a suggestive associations between PGS-ADR and schizophrenia (p=0.0007), chronic obstructive pulmonary disease (COPD, p=0.0009). Furthermore, sex-specific analyses demonstrated some diseases with discordant association between sex; men showed associations with endocrine, urological, and respiratory diseases, whereas women with mental disorders and cardiovascular diseases.Conclusion: Our findings suggest that so-called pharmacogenes may be linked with some disease susceptibilities, and require a further research whether the apparent associations imply any clues in the pathogenesis of the diseases. Moreover, the effects of pharmacogenes may exert differential effects by sex. It suggests the need for further investigation to develop precise treatment approaches based on an individual's genetic predispositions.

Session Title: Pharmacogenomics Poster Session II

PB2151 Exploring the role of local ancestry in the distribution of pharmacogenomics variants in an admixed population

Authors:

T. De Oliveira[1], R. Secolin[2], M. Naslavsky[3], T. Hünemeier[4], I. Lopes-Cendes[5]; [1]Univ. of Campinas, São Paulo, Brazil, [2]Univ. of Campinas - UNICAMP, Campinas, Brazil, [3]Univ. of Sao Paulo, Sao Paulo, Brazil, [4]Univ. of Sao Paulo, São Paulo, Brazil, [5]Univ. of Campinas, Campinas, Brazil

Introduction: Pharmacogenomics (PGx) studies in populations with multi-ethnic ancestry indicate that ancestry-informative markers were enriched in pharmacogenetic *loci*, suggesting that trans-ancestry differentiation must be carefully considered in population PGx studies. We aim to explore this issue further by investigating haplotype blocks (HB) of different ancestries in the distribution of pharmacogenomics variants in an admixed population. **Materials and Methods**: Data were collected from 1,349 admixed individuals from two Brazilian population cohorts (BRA): The Brazilian Initiative on Precision Medicine (www.bipmed.org) and the Health, Well-Being, and Aging Project (abraom.ib.usp.br). We compared these with 95 Brazilian Native-American individuals and 1,525 samples from the 1KGP database representing different ancestral origins. Data analysis included: (i) local ancestry inference using RFMix2; (ii) identification of PGx variants based on the PharmaGKB database; and (iii) exploration of the relationship between PGx variants and local ancestry in the BRA samples. **Results**: Global ancestry shows that BRA was predominantly of European ancestry, with 75% of samples having HBs with >=70% European ancestry, followed by African ancestry at 3%. None of the samples had HBs with >=40% Native American ancestry. A total of 458 PharmaGKB SNPs with an allele frequency (AF) >5% were identified in BRA. Among these, 10 SNPs were exclusively associated with European HBs, mainly related to drug efficacy. More than 81% of the PharmaGKB SNPs were found in both European and African haplotype blocks, with many of them associated with drug efficacy in phenotypes such as schizophrenia and hepatitis C and toxicity related to weight gain. Interestingly, we found PharmaGKB SNPs with AF<=5% exclusively present in the African HBs, although we observed only 3% of global African ancestry in BRA samples. In addition, we observed that while less than 9.8% of the SNPs were found in individuals with both European and Native American haplotype blocks in the same *locus* (EUR-NAT_HB), 23% of these SNPs were associated with drug efficacy in hypertension-related medications. **Conclusion**: Our findings indicate that local ancestry in admixed individuals may be linked to different patterns of PharmaGKB SNPs, raising important questions regarding the application of pharmacogenomics in individuals displaying different proportions of ancestral HBs.

Session Title: Pharmacogenomics Poster Session III

PB2152 Feasibility and utility of multi-gene, pharmacogenetic testing among patients in primary care, community-based hospital setting.

Authors:

S. Patel[1], J. Allen[2], J. Grund[2], G. Miller[2], C. Brueckner[2], A. Crider[2], J. Wallace[2], B. Phillips[2], T. Koepp[3], J. England[3], A. Desai-Naik[4], J. Chiang[4], B. Wendt[2], N. El Rouby[1]; [1]Univ. of Cincinnati, Cincinnati, OH, [2]St. Elizabeth, Edgewood, KY, [3]OneOme, Minneapolis, MN, [4]Humana, Louisville, KY

Pharmacogenetic-guided prescribing is a key component of precision medicine. Multi-gene panel testing is especially beneficial for patients who are on medications that are likely impacted by multiple Pharmacogenes. However, the data on the feasibility and utility of pharmacogenetic testing across community-based Healthcare systems are fewer compared to data from large academic hospitals.

We sought to determine the value of a multi-gene panel, pharmacogenetic testing by identifying the prevalence of drugs with pharmacogenetic evidence, and genotype-drug interaction among patients from primary care at St Elizabeth Healthcare system. We focused on primary care for two reasons: 1) this is a common setting for prescribing chronic medications with pharmacogenetics evidence; 2) patients often have co-morbidities, requiring polypharmacy.

We identified 6000 patients as candidates for pharmacogenetic testing based on recent pharmacogenetic prescription(s) from insurance claims data. Patients were contacted using letters from their payer system and Epic generated messages from St. Elizabeth. A total of 1043 patients were tested using a 27-gene, pharmacogenetic panel that captures *CYP2C19*, *CYP2D6*, *CYP2C9*, *CYP3A5*, *SLCO1B1*, *VKORC1*, *DPYD*, *UGT1A1*, *TPMT*, *NUDT15*, among others. Genotypes were collected and current pharmacogenetic medications were recorded from the electronic health records. We calculated the number and percentage of patients with 1) one or more pharmacogenetic medications at the time of testing; 2) one or more genotype-drug interactions.

Among patients with returned results, 86% have one or more pharmacogenetic medications, at the time of testing. The most commonly prescribed drugs or drug classes were statins, proton pump inhibitors (PPIs), metoprolol, Selective Serotonin Reuptake Inhibitors (SSRIs), and Non-Steroidal Anti-inflammatory Drugs (NSAIDs), representing 58%, 40%, 22%, 15%, and 13% of patients with these pharmacogenetic medications, respectively, whose responses would be influenced by variants in *SLCO1B1* (statins), *CYP2D6* (metoprolol), *CYP2C19* (SSRIs, PPIs), and *CYP2C9* (NSAIDs). Among patients with genotype-drug interactions, 31.3% had one gene-drug interactions, 19.5% had two gene-drug interactions, and 9.5% had three or more gene-drug interactions, with 60.1% of the patients having one or more gene-drug interactions.

In conclusion, the data demonstrate a high prevalence of genotype-drug interactions among primary care patients, justifying a multi-gene, panel-based, pharmacogenetic testing approach. Further research is needed to demonstrate the economic value of panel-based testing.

Session Title: Pharmacogenomics Poster Session I

PB2153 † Further Personalizing Medicine: Is There a Role for Comprehensive Genomic Evaluations for Hematopoietic Cell Transplantation Recipients?

Authors:

M. Similuk[1], J. Yan[1], T. Sissung[2], E. Damskey[1], B. Beers[1], H. Orellana[1], S. Byers[1], M. Kamen[1], C. Jodarski[1], R. Gore[1], N. Reynolds-Lallement[1], K. Lewis[1], R. Ghosh[1], B. Seifert[1], D. Dimitrova[3], J. Kanakry[3], D. Hickstein[4], T. Hughes[5], G. Uzel[1], C. Zerbe[1], H. C. Su[1], A. Freeman[1], S. De Ravin[1], H. Malech[1], L. Notarangelo[1], S. M. Holland[1], M. Walkiewicz[1]; [1]Div. of Intramural Res., Natl. Inst. of Allergy and Infectious Diseases, NIH, Bethesda, MD, [2]Genitourinary Malignancies Branch, Ctr. for Cancer Res., Natl. Cancer Inst., NIH, Bethesda, MD, [3]Ctr. for Immuno-Oncology, Natl. Cancer Inst., NIH, Bethesda, MD, [4]Immune Deficiency Cellular Therapy Program, Natl. Cancer Inst., NIH, Bethesda, MD, [5]Clinical Ctr. Pharmacy Dept., NIH, Bethesda, MD

Background: Allogeneic hematopoietic cell transplantation (HCT) provides patients with severe immune disorders the potential for effective long-term management. Improved methods are increasing the chance, duration, & quality of survival. Focused genetic testing for the underlying defect is common and informs risk assessment, conditioning, donor selection, & graft-versus-host prophylaxis. Concurrently, the accessibility of broad genetic testing is improving, as are the approaches for complex loci such as those for pharmacogenomics (PGx). More comprehensive genomic evaluations for HCT have uncertain clinical utility. We seek to address this gap through a review of HCT patients referred to our center.
Methods: We performed exome and/or genome sequencing on patients with immune disorders from 2017-2023. We analyzed the data for primary & secondary findings (SF). *Post-hoc*, we analyzed NGS-derived PGx alleles & medication history.
Results: We analyzed exome (n=144) and/or genome (n=124) data for 210 patients (202 families) who received, or had been evaluated for, HCT (85.7% & 14.3%, respectively). Participants were 41.9% female, mean age 26.8y, & 36.2% had non-European genomic ancestry. Overall, 153/210 (72.9%) patients had an identified molecular diagnosis underlying their HCT indication, most commonly with defects in *GATA2* (n=57) & *CYBB* (n=31), reflecting our center's expertise. Additionally, 10/210 (4.8%) patients had SF related to cancer (n=6: *BRCA1*, 2; *RET*, 1; *PMS2*, *MLH1*, *MSH2*, 1 each), cardiovascular disease (n=2: *TTR*, *KCNQ1*), or malignant hyperthermia (n=2: *RYR1*) conferring risks that may be exacerbated by HCT's genotoxic stress or use of succinylcholine. Among the 4 HCT recipients with cancer syndromes, no malignancies have occurred with cumulative follow up of 6.1y (2d-3.5y). Lastly, PGx analysis of *CYP2C19*, *CYP2D6*, *CYP3A5*, *NUDT15*, & *TMPT* identified ≥1 actionable allele(s) in 124/210 (59.0%) patients. Among those with at-risk genotypes, 109/124 patients (87.9%; or 109/210, 51.9% of total) were given ≥1 medication(s) that may have benefited from PGx-informed personalization including tacrolimus, azathioprine, ondansetron, opioids, SSRIs, tricyclic antidepressants, voriconazole, & proton pump inhibitors.
Conclusions: Comprehensive genomic evaluations for HCT are increasingly accessible & may be an emerging avenue for optimizing HCT approaches & outcomes. Specifically, identifying cancer or cardiovascular predisposition in a critical minority of patients may inform conditioning & post-HCT risk assessment/management. Prospective PGx may also be informative, particularly given HCT's inherent polypharmacy.

Session Title: Pharmacogenomics Poster Session II

PB2154 Genetic Diversity of variants involved in drug response among Tunisian and Italian populations: implication for personalized medcine

Authors:

H. JMEL[1,2,3], S. Sarno[4], C. Guilani[4], W. Boukhalfa,[1], S. Abdelhak[1,2], D. Luiselli[5], R. KEFI[1,2,3]; [1]Lab. of BioMed. Genomics and Oncogenetics, Inst. Pasteur de Tunis, Tunis, Tunisia, Tunis, Tunisia, [2]Univ. of Tunis El Manar, 2092 El Manar I Tunis, Tunisia, Tunis, Tunisia, [3]Genetic typing service, Inst. Pasteur de Tunis, BP 74, 13 Place Pasteur, Tunis 1002, Tunis, Tunisia, Tunis, Tunisia, [4]Lab. of Molecular Anthropology & Ctr. for Genome Biology, Dept. of Biological, Geological and Environmental Sci. (BiGeA), Univ. of Bologna, Bologna, Italy., Italy, Italy, [5]Lab. of Ancient DNA (aDNALab), Dept. of Cultural Heritage (DBC), Univ. of Bologna, Ravenna, Italy, Italy, Italy

Adverse drug reactions are a cause of significant morbidity and mortality to patients and a source of financial burden to the healthcare system. Several studies highlighted that ethnicity is one of the main factors influencing individual genetic patterns in drug response. The aim of the study is to characterize the genetic variability of some pharmacogenes invoved in drug biotransformation and ADR in Tunisian (ADR), Italian population and to compare our results to the worldwide populations. A set of 135 healthy Tunisians and 690 Italians were genotyped using array chip. Variants located in 25 Very Important Pharmacogenes involved in drug response variability and ADR were extracted from the genotyping data. Analysis of variant distribution in Tunisian and Italian populations compared to 24 worldwide populations publicly available was performed using plink and R software. Common variants between Tunisians, Italian and the 24 investigated populations were extracted from genotyping data. Results of fixation index (Fst), Principle Component Analysis and ADMIXTURE analyses showed that there is high similarity among Mediterranean populations which are genetically divergent from South African populations. Moreover, the Fst comparaive analysis highlighted 27 variants with high level of differentiation between the Tunisian, Italian and other studied populations. Among these variants, there are four SNPs rs622342, rs3846662, rs7294, rs5215 located respectively in S*LC22A1*, *HMGCR*, *VKORC1*, *KCNJ11*, involved in ADR showed genotypic frequency differences. In conclusion, our study showed that Tunisian and Italian population are genetically homogenous regarding the studied pharmacogenes. The Correlation of the genotype and allelic frequencies of risk variants with their associated adverse drug reactions would enhance the drug outcomes and have an important impact in the implementation of personalized medicine in worldwide populations. Similar studies need to be reproduced to identify populations that require attention when taking a particular drug.

Session Title: Pharmacogenomics Poster Session III

PB2155 Genetic prediction of medication side effects: a genome-wide investigation of abnormal dreams on varenicline.

Authors:

M. Chenoweth[1,2], Y. Kim[2], N. Nollen[3], C. Lerman[4], J. Knight[5], R. Tyndale[2,1]; [1]Ctr. for Addiction and Mental Hlth., Toronto, ON, Canada, [2]Univ. of Toronto, Toronto, ON, Canada, [3]Univ. of Kansas Sch. of Med., Kansas City, KS, [4]Univ. of Southern California, Los Angeles, CA, [5]Lancaster Univ., Bailrigg, United Kingdom

Introduction: Varenicline, the most efficacious smoking cessation aid, produces sleep disturbances including abnormal dreams. While genetic contributors to varenicline-associated nausea and cessation have been identified, the role of genetics in varenicline-associated abnormal dreams is unknown. **Methods:** We conducted a GWAS of abnormal dreams in n=188 European ancestry (EA) participants treated with varenicline for smoking cessation (NCT01314001). Additive genetic models examined the likelihood of experiencing abnormal dreams two weeks following varenicline initiation, controlling for population substructure, sex, and age. We also examined associations between the top variant and abnormal dreams in EA participants treated with nicotine patch (n=181) and placebo (n=168) and tested for replication in a second varenicline trial (NCT01836276). Finally, we tested for an association between the top variant and abnormal dreams in n=137 varenicline-treated participants of African ancestry (AA) (NCT01314001). **Results:** Thirty-eight percent of varenicline-treated EA participants experienced abnormal dreams. The top variant associated with abnormal dreams was rs901886 (Odds ratio (OR)=2.94 for T vs. C, 95% CI=1.92-4.55, P=2.03e-7; T allele frequency=52%), mapping to intron 2 of *ICAM5* on chromosome 19. The prevalence of abnormal dreams in those with rs901886 TT, TC, and CC genotypes was 62%, 36%, and 15%, respectively. Conversely, there was no association between rs901886 and abnormal dreams on nicotine patch (P=0.519) or placebo (P=0.909). In the second trial of varenicline-treated EA participants (n=184), a similar direction of effect of rs901886 on abnormal dreams was observed: the prevalence of abnormal dreams was 39.2%, 29.2%, and 27.3% for those with rs901886 TT, TC, and CC genotypes, respectively (P=0.203). In contrast to EA, rs901886 was not associated with abnormal dreams in varenicline-treated AA participants (OR=0.74 for T vs. C allele, 95% CI=0.32-1.69; P=0.407; T allele frequency=88%). However, there was some support for a region located ~74.4kb 5' of *ICAM5* in AA (top variant: rs113359011: OR=0.41 for T vs. C allele, 95% CI=0.20-0.84, P=0.00256; T allele frequency=29%). **Conclusions:** Genetic variation in *ICAM5* influences the risk for experiencing abnormal dreams on varenicline. This work improves our understanding of individual differences in smoking cessation medication response.

Session Title: Pharmacogenomics Poster Session I

PB2156 Genetic risk of immune related adverse events in melanoma patients receiving ipilimumab.

Authors:

J. Pluta[1], P. Kanetsky[2], C. Horak[3], K. Nathanson[4]; [1]Univ. of Pennsylvania, Philadelphia, PA, [2]Moffitt Cancer Ctr., Tampa, FL, [3]Bristol Myers Squibb, Princeton, NJ, [4]Univ. of Pennsylvania Sch. of Med., Philadelphia, PA

Background: A side effect of ipilimumab, an immune checkpoint inhibitor targeting CTLA-4, is increased risk of serious immune related adverse events (irAE). We performed a genome-wide association study (GWAS) on patients with melanoma treated with ipilimumab monotherapy to explore inherited genetic variation associated with the development of irAE. We also examined the impact of prior treatment, i.e., surgery, chemotherapy, and/or radiotherapy, on genetic risk of irAEs. **Methods:** Genotype data and clinical information were obtained on 706 patients with metastatic melanoma enrolled on CA184-169 (NCT01515189) and treated with ipilimumab monotherapy. Trial participants were genotyped on the Affymetrix 6 array. After removing individuals failing genetic quality control and those with missing clinical data, a total of 668 patients were analyzed. SNP associations were assessed by additive logistic regression models using irAE grade 3 or greater (yes, no; n=167) as the dependent variable and SNP genotype as the independent variable; all models were adjusted for ECOG status (0, 1+), dosage (3, 10 mg/kg), number of doses (<4, 4, 4+), genetic ancestry (3 principle components), and an interaction term between SNP and any prior cancer treatment (yes, no; n= 374). The proportion of patients experiencing an irAE was similar between those with and without prior treatment ($p = 0.37$). A likelihood-ratio test was used to jointly test for an association attributable to the SNP main effect or the interaction between the SNP and prior treatment; and we used a relaxed p-value threshold of 1×10^{-5} due to the limited sample size. **Results:** The most significant marker (rs1030480; $p = 1.42 \times 10^{-7}$) is intergenic between *CDH9* and *CDH10.* Five markers mapped to genes with known or suspected roles in autoimmune disease: *NELL1* (rs7947369, $p = 4.2 \times 10^{-6}$), *MRGBP* (rs33992134, $p = 8.03 \times 10^{-6}$), *RBMS3* (rs9844665, $p = 5.34 \times 10^{-6}$), *CACNA1B*(rs7021325, $p = 3.07 \times 10^{-6}$), and ZNF479 (rs3887628, 2.54×10^{-6}). Two markers, rs1111665 ($p = 8.24 \times 10^{-6}$) and rs1857772 ($p = 4.24 \times 10^{-6}$), map to the super-enhancer regions of *FGL2* and *SEMA5A*, respectively, both of which are involved in autoimmune processes. Of the 14 associated markers, 13 statistically interacted ($p < 0.05$) with prior treatment. **Conclusions**: We found broad involvement of autoimmune markers with the development of irAE, consistent with other studies. Our results indicate prior treatment for cancer may alter genetic susceptibility to irAE. Replication of results across studies remains a challenge, likely due to limited sample sizes and heterogeneity of subjects.

Session Title: Pharmacogenomics Poster Session II

PB2157 Genetic Variants Associated with Iloperidone Response in Acute and Mixed Mania Associated with Bipolar I Disorder.

Authors:

H. Bai, S. P. Smieszek, R. Torres, C. Xiao, C. M. Polymeropoulos, G. Birznieks, M. H. Polymeropoulos; Vanda Pharmaceuticals Inc., Washington, DC

Bipolar I disorder is characterized by episodes of manic and hypomanic activity, and it has one of the highest rates of serious impairment among mood disorders. Iloperidone is a second-generation antipsychotic approved by the FDA that has anti-manic effects. Here we report genetic associations on iloperidone efficacy for treatment of bipolar manic episodes from a phase 3, randomized, placebo-controlled study. Whole genome sequencing was conducted with 30× read depth using whole blood samples obtained from the study subjects. The efficacy of iloperidone was evaluated using the Young Mania Rating Scale (YMRS) -calculated as the change in YMRS measurement from baseline to the end of study (EOS). First, we found the omega3 total levels are associated with decreased YMRS at baseline (b=-5.34±1.91, p=0.006), which is concordant with previous findings and indicates the validity of the data. In order to investigate the genetic factors that impacts the iloperidone efficacy, we then conducted GWAS in the Iloperidone group (n=167) using linear regression on YMRS change with adjustment of AGE, SEX, BMI and the first 2 PCs. Multiple variants were found to be associated with YMRS change at 1e-5 level. One of the top hits was RELN (rs55837573, b=-4.61±0.94, p=2.48e-6). RELN encodes a secretory glycoprotein critical for brain development and synaptic plasticity. Lower Reelin protein levels were reported in bipolar disorder patient brains. NAV2 (rs1118464, b=-4.15±0.89, p=6.98e-6) is a neuron navigator gene and may play a role in cell growth and migration. FAT3 (rs76876307, b=15.44±3.28, p=5.27e-6) is an atypical cadherin that is predicted to be involved in cell-cell adhesion and act upstream of dendrite development and neuron migration. SHROOM3 (rs62300864, b=12.08±2.43, p=1.64e-6) may play a role in regulating cell shapes. These variants are not associated with YMRS at baseline, which suggests potential drug mechanisms. We then defined the drug responders as subjects that having EOS YMRS decreased >=50% compared to baseline. GWAS with logistic regression were conducted, and the association of RELN (rs55837573) was robustly detected in this analysis. Over representation analysis in disease gene sets showed bipolar I disorder was the strongest result (p=1.12e-5, FDR=0.044). This was driven by the contribution of variants within BCL2, EGR4, and GADL1. In summary, we report genetic factors associated with the efficacy of iloperidone in treating bipolar manic episodes. This result suggests that the treatment with iloperidone potentially leads to improvement as captured by YMRS through acting on the identified gene pathways, setting a foundation for further investigation.

Session Title: Pharmacogenomics Poster Session III

PB2158 Genetic Variants in *GC* and *CYP24A1* are associated with Bleeding complications for Atrial Fibrillation Patients treated with Direct Oral Anticoagulants.

Authors:

D. Lee[1], Y-A. Park[1], J. Kim[1], J. Park[2], H. Gwak[1]; [1]Ewha Womans Univ., Seoul, Korea, Republic of, [2]Dept. of Cardiology, Mokdong Hosp., Ewha Womans Univ. Coll. of Med., Seoul, Korea, Republic of

The aim of this study was to investigate vitamin D-related genetic factors that are associated with bleeding complications in atrial fibrillation patients taking direct oral anticoagulants (DOACs). This study was a retrospective analysis of prospectively collected samples from June 2018 to May 2022. In order to examine the associations between vitamin D-related genetic factors and bleedings, we selected 12 single nucleotide polymorphisms (SNPs) from five specific genes: *CYP24A1* (rs1570669, rs2296241, rs6068016), *CYP27B1* (rs4646537), *CYP2R1* (rs1993116), *VDR* (rs11568820, rs2228570, rs731236, rs7975232), and *GC* (rs4588, rs7041). Multivariable logistic regression analysis was utilized to identify independent risk factors for bleeding complications. A total of 516 patients were included in the analysis, and 109 experienced bleeding complications. Both severe renal impairment (CrCl < 30 mL/min) and concurrent use of proton pump inhibitors increased the risk of bleedings (adjusted odds ratio (AOR): 2.54; 95% confidence interval (CI): 1.05-6.13, AOR: 1.82; 95% CI: 1.01-3.27, respectively). Among the vitamin D-related genetic variations, patients carrying the TT genotype of rs4588 and the CC genotype of rs6068816 exhibited 2.32- (95% CI: 1.06-5.05) and 1.59-fold (95% CI: 1.01-2.50) higher risk of bleedings. This is the first study to suggest vitamin D-related genetic markers, including both clinical and genetic factors, for bleeding complications in patients with atrial fibrillation receiving DOAC treatment. These findings would be useful for personalized DOAC therapy.

Session Title: Pharmacogenomics Poster Session I

PB2159 Genome-wide association studies categorized by class of anti-hypertensive drugs reveal complex pathogenesis of hypertension with drug resistance.

Authors:

K. Yamazaki[1,2,3], C. Terao[4], A. Takahashi[5], Y. Kamatani[6,3], K. Matsuda[7], S. Asai[2], Y. Takahashi[2]; [1]Chiba Univ., Chiba, Japan, [2]Nihon Univ. Sch. of Med., Tokyo, Japan, [3]RIKEN Ctr. for Integrative Med. Sci., Yokohama, Japan, [4]RIKEN IMS, Yokohama, Japan, [5]RIKEN, Yokohama, Japan, [6]The Univ. of Tokyo, Minato-ku, Japan, [7]The Univ. of Tokyo, Tokyo, Japan

Resistant hypertension is defined as uncontrolled blood pressure despite the use of three or more anti-hypertensive drugs of different classes. Though genetic factors may greatly contribute to hypertension with resistance to multiple drug classes, more than for general hypertension, its pathogenesis remains unknown. To reveal the genetic background of resistant hypertension, we categorized 32,239 patients whose data were obtained from the BioBank Japan Project, by prescription of seven classes of anti-hypertensive drugs and performed genome-wide association studies. Our genome-wide association studies identified four loci with significant association ($P < 5\times10^{-8}$): rs6445583 in *CACNA1D* ($P = 4.99\times10^{-8}$, odds ratio (OR) (95% confidence interval (CI)) = 1.13 (1.08-1.18)) and rs12308051 in the intergenic region on chromosome 12 ($P = 1.34\times10^{-8}$, OR (95% CI) = 1.15 (1.09-1.20)) for ARBs, rs35497065 in *FOXA3* ($P = 2.28\times10^{-8}$, OR (95% CI) = 0.89 (0.86-0.93)), for CCBs, and rs11066280 in *HECTD4* ($P = 2.71\times10^{-8}$, OR (95% CI) = 0.76 (0.67-0.85)) for αβ-blockers. Since these loci are known to be susceptibility loci for hypertension and/or blood pressure, our results indicate that resistant hypertension is caused by a combination of excessive blood pressure and drug resistance to each anti-hypertensive pharmacological class. Furthermore, to investigate the genetic difference between BP traits and the treatment effectiveness of anti-hypertensive drugs, we performed gene-set analysis and calculated the genetic correlation continuously with GWAS summary statistics of blood pressure from BioBank Japan. Most of the genetic factors were in common between BP traits and anti-hypertensive effectiveness, but it seems that the genetic architecture of the drug response to anti-hypertensive treatment is more complicated than BP traits. This corresponds to the well-known mosaic theory of hypertension. Our findings reveal the complex pathogenesis of hypertension with resistance to multiple classes of anti-hypertensive drugs.

Session Title: Pharmacogenomics Poster Session II

PB2160 Global pharmacogenetics: An analysis of the 1000 Genomes Project.

Authors:

C. Sherman[1], K. Claw[1], S-b. Lee[2]; [1]Univ. of Colorado, Aurora, CO, [2]Macrogen Inc, Seoul, Korea, Republic of

Pharmacogenetics (PGx) examines how individual genetic variations impact drug responses and has the potential to optimize medication dosage, selection, and efficacy. While there have been significant advances in our knowledge of PGx and drug-gene interactions, population-level PGx variation is still not well characterized, which could exacerbate health disparities. Previous studies have primarily focused on populations of European descent, and to ensure equitable implementation of PGx findings, it is crucial to enhance the inclusion of historically underrepresented groups. The goal of this study was to uniformly characterize population-level star alleles (haplotype pattern) and phenotype frequencies for 58 pharmacogenes within the 1000 Genomes Project (N=2,504; 26 global populations) using high-coverage whole-genome sequencing data and incorporating structural variant (SV) analyses. These pharmacogenes involve genes that have a role in drug absorption, distribution, metabolism, and excretion. PyPGx identifies star alleles from sequencing data by detecting single nucleotide variants (SNVs), insertion-deletion variants, and SVs. With previous sequencing methods, SVs in drug metabolizing enzymes have been difficult to detect and interpret in part due to high sequence homology, especially among functional genes and nonfunctional pseudogenes. PyPGx uses a machine learning-based approach with the WGS to detect SVs, including gene deletions, duplications, and hybrids. We found that while many of our star allele and phenotype frequencies were consistent with the literature, our approach was able to detect several population-specific frequencies that had at least a 2-fold difference. This suggests that previous works may have mischaracterized or inaccurately measured these frequencies. We validated known SVs and identified several novel SVs that are not present in public PGx databases. Using in-silico tools, we also identified 210 SNVs and insertion-deletion variants with severe functional consequences, 8 of these variants have been previously identified as pathogenic, likely pathogenic, or drug-responsive. Our findings provide a valuable reference resource that can be used to improve clinical care across diverse populations.

Session Title: Pharmacogenomics Poster Session III

PB2161 Human genetics of COVID-19 mRNA vaccination efficacy.

Authors:

A. Bolze[1], C. M. Clinton[1], A. Mackie[1], C-H. Chan[2], J. Grzymski[3], D. T. Harris[4], P. A. Pawloski[5], E. Cirulli[1]; [1]Helix, San Mateo, CA, [2]Sanford Res., Sioux Falls, SD, [3]Desert Res. Inst., Reno, NV, [4]The Univ. of Arizona, Tucson, AZ, [5]Hlth.Partners, Minneapolis, MN

Introduction: COVID-19 mRNA vaccines are safe and effective. Real-world data estimate that they have saved millions of lives. However, some fully vaccinated individuals have experienced SARS-CoV-2 infections with severe symptoms, prompting additional medical care and even resulting in hospitalizations. There are few studies on the genetics of vaccine failure, when people acquire severe infectious disease despite being vaccinated against it. As a result, the genetic and immunological basis of vaccine-induced protection against infections and severe disease is not well understood.
Objective: The goal of this work is to identify human genetic variants, which may be associated with (i) COVID-19 severity in fully vaccinated individuals, and (ii) the rate of declining immunity in fully vaccinated individuals.
Methods: Adults aged 18 to 75 years old living in the United States and who received at least two doses of COVID-19 mRNA vaccines were included. Cases had documented symptomatic infection while controls had no record of symptomatic COVID-19 since vaccination. Three types of human genetic analyses will be conducted: a GWAS, an HLA-WAS and rare variant gene-based collapsing tests. In addition to genome-wide discoveries, we will also test several hypotheses including whether the known genetic associations (n=49) with severe COVID-19 in the unvaccinated are also associated with severe COVID-19 disease despite vaccination.
Results: In the project's first six months, we established the research protocol and collaborations with four health systems. At least 2,000 cases have been identified and are being sequenced. The analysis will be completed in September 2023, and the results will be presented.
Conclusion: The results of this study may provide insights that could inform public health decisions regarding vaccine administration including potentially refining the definition of at-risk populations, and the timing of booster schedules.

Session Title: Pharmacogenomics Poster Session I

PB2162 Identification of African-specific*NOS3*variants associated with direct oral anticoagulant response.

Authors:

G. Yang[1], C. Parker[1], T. Priya[1], C. Alarcon[1], P. Friedman[1], E. Nutescu[2], M. Tuck[3], T. O'Brien[4], L. Gong[5], T. Klein[5], D. Meltzer[6], M. Perera[1]; [1]Northwestern Univ., Chicago, IL, [2]Univ. of Illinois Chicago, Chicago, IL, [3]Washington DC VA Med. Ctr. and The George Washington Univ., Washington, DC, [4]The George Washington Univ., Washington, DC, [5]Stanford Univ, Stanford, CA, [6]Univ. of Chicago, Chicago, IL

BACKGROUND: Direct oral anticoagulants (DOACs) are widely used for thromboembolism prevention and treatment by directly inhibiting factor Xa, a key clotting factor involved in the formation of thrombin. Clinical trials and adverse event reports suggest variability among individuals in response to DOACs. Unravelling the genetic factors influencing DOAC response is important for better clinical management. We aimed to identify pharmacogenomic biomarkers specifically related to DOAC response in African American patients, a population that suffers disproportionally from thromboembolism.**METHODS:** Samples were obtained from self-identified African American patients who had received treatment with DOACs for a minimum of 3 days. Anti-Xa levels, a key indicator of DOAC response, were collected for these patients. A genome-wide association study (GWAS) was performed to identify variants associated with alterations in anti-Xa levels, while adjusting for age, gender, and principal components. *In silico* analysis was conducted to identify potential functional variants. Nitric Oxide (NO) assays were performed to investigate the impact of these potential functional variants on NO production (surrogate marker of eNOS activity) in the plasma of African American subjects carrying the associated *NOS3* haplotype.**Results:** Our GWAS identified one locus on chromosome 7 associated with a decreased anti-Xa level. The top association, rs114982053, is located upstream of Nitric Oxide Synthase 3 **(***NOS3* - encoding eNOS) ($p=9.29x10^{-7}$,β=-0.68, 95% confidence interval=-0.95--0.42). *In silico* analyses showed three SNPs with high LD located at the potential enhancer/promoter region of *NOS3* gene. The plasma NO assay showed an increase in nitrite concentration in subjects carrying the effect alleles compared with the subjects with the reference haplotype (p=0.003).**Conclusion:** Our data suggests that a 3 SNP haplotype, located on chromosome 7 contributes to increasing NO production, resulting from the elevated *NOS3* expression and eNOS activity and decrease DOAC drug concetrations. Notably, this haplotype is African-specific. Given the pivotal role of *NOS3* in regulating vascular function and thrombosis, our results suggest a potential role of this enzyme in modulating response to DOACs and thrombosis risk. Further investigations into the precise mechanisms underlying this association may provide valuable insights for personalized treatment strategies in individuals at risk of thromboembolic events.

Session Title: Pharmacogenomics Poster Session II

PB2163 Identification of Novel Genetic Markers for Antipsychotic Response in Schizophrenia Patients: A Comprehensive Genome-Wide Association Study and Development of an Automated Treatment Decision System

Authors:

L. Chen[1], C. Huai[1], W. Shaochang[2], C. Song[3], Y. Xu[4], Z. Yi[5,6], S. Gao[7], J. Tang[8], C. Liu[9], S. Weng[10], G. Wang[11], X. Zhao[12], X. Zhang[13], H. Huang[14], S. Qin[1]; [1]Shanghai Jiao Tong Univ., Shanghai, China, [2]The Second People's Hosp. of Lishui, Lishui, China, [3]The Fourth People's Hosp. of Wuhu, Wuhu, China, [4]Shanxi Key Lab. of Artificial Intelligence Assisted Diagnosis and Treatment for Mental Disorder, First Hosp. of Shanxi Med. Univ., Taiyuan, China, [5]Shanghai Mental Hlth.Ctr., Shanghai Jiao Tong Univ. Sch. of Med., Shanghai, China, [6]Dept. of Psychiatry, First Hosp./First Clinical Med. Coll. of Shanxi Med. Univ., Taiyuan, China, [7]Zhumadian Psychiatric Hosp., Henan, China, [8]Dept. of Psychiatry, Sir Run-Run Shaw Hosp., Sch. of Med., Zhejiang Univ., Hangzhou, China, [9]Jining Med. Univ., Dept. of Psychiaty, Sch. of Mental Hlth., Psychiat Genet Lab PSYG Lab, Jining, China, [10]Fuzhou Neuro-psychiatric hosptial affilted to Fujian Med. Univ., Fuzhou, China, [11]Wuxi Mental Hlth.Ctr. Affiliated to Nanjing Med. Univ., Wuxi, China, [12]Shanghai Pudong New Area Mental Hlth.Ctr., Tongji Univ. Sch. of Med., Shanghai, China, [13]Jingzhou Mental Hlth.Ctr., Jingzhou, China, [14]The Broad Inst. of MIT and Harvard, Boston, MA

Schizophrenia is a debilitating mental disorder that affects approximately 1% of the global population. Despite the availability of various antipsychotic medications, a significant number of patients experience inadequate response, leading to substantial suffering and economic burden. The emergence of personalized medicine offers a promising approach to address this issue by tailoring treatments based on individual genetic profiles. Thus, this study aimed to discover novel genetic variants that could serve as biomarkers for predicting the response to antipsychotic drugs in Han Chinese patients with schizophrenia. A two-stage genome-wide association study (GWAS) was conducted on a total of 3069 patients who received mono-antipsychotic therapy for a minimum of 8 weeks, including olanzapine, risperidone, quetiapine, aripiprazole, and/or amisulpride. The categorization of antipsychotic drug response groups was performed using the Positive and Negative Syndrome Scale (PANSS). Binary logistic regression identified significant single nucleotide polymorphisms (SNPs) distinguishing high responders (≥60% PANSS reduction) from low responders (≤40% PANSS reduction) for these five drugs. The GWAS analysis successfully revealed three novel genetic loci within brain-related genes and identified several previously associated SNPs related to antipsychotic drug response. Moreover, we used ordinal logistic regression to examine the SNPs associated with three drug response categories (high responders, medium responders, and non-responders) in the entire cohort. Medium responders were defined as individuals with PANSS reduction between 40% and 60%. This approach identified two novel loci that were significantly associated with antipsychotic drug response. Furthermore, a Gene Ontology (GO) enrichment analysis demonstrated that the identified genetic loci were significantly enriched in neuronal signaling pathways within the brain. Additionally, a novel five-drug scoring system was developed, incorporating both newly discovered and previously identified SNPs from the GWAS. This scoring system utilized a polygenic risk score (PRS) and exhibited relatively high prediction accuracy in both the discovery and replicate cohorts. It will enabled clinicians to predict drug response and make informed decisions regarding optimal treatment selection. The findings of this study, along with the developed drug scoring method, will contribute to a deeper understanding of the underlying mechanisms of antipsychotic drug response and the advancement of personalized treatment strategies for individual patients with schizophrenia.

Session Title: Pharmacogenomics Poster Session III

PB2164 Identification of sex-related pharmacogenetic predictors of warfarin-related hemorrhage.

Authors:

B. Davis, C. Dillon, N. Limdi; Univ. of Alabama at Birmingham, Birmingham, AL

Anticoagulants are used for prophylaxis and treatment of thromboembolism. Hemorrhage is a major adverse effect of therapy and is the target of a national action plan to prevent adverse drug reactions (ADRs). Although warfarin use has declined since the introduction of newer agents, it remains among the top prescribed drugs (#58). Warfarin exhibits female-biased bleeding ADRs, but sex-related genetic predictors of warfarin-related hemorrhage (WRH) have not been evaluated. Given sex-related differences in gene expression, genetic variant effects, and drug pharmacokinetics, we evaluated sex-related genetic predictors of WRH among African Americans (AA) and European Americans (EA).

Warfarin users (N=1,786; 48% female; 47% AA) were enrolled at UAB and followed for 2 years. Demographic and clinical data were collected at baseline and changes documented during follow-up. Samples were genotyped and imputed to the Trans-Omics for Precision Medicine (TOPMed) reference panel. Variants in ~1,200 drug-related genes were obtained from the imputed data. The influence of genetic variants on WRH were analyzed using logistic regression, stratified by sex and ancestry. We also evaluated only females >55 years old to identify differences related to hormonal variation. Among EA women, NEGR1 rs12130762, encoding a neuronal cell adhesion molecule, was associated with ~4-fold higher odds of WRH (p=9.31x10-7). Among AA women, PLCB1 rs6055904, encoding a phospholipase involved in endothelial barrier regulation, was associated with 3-fold higher WRH odds (p=3.51x10-7). These associations remained when limited to women >55 years. Additionally, UGT2B4 rs1051752 (African-specific), encoding an enzyme involved in estrogen glucuronidation, was associated with 5-fold higher odds of WRH in this age group (p=5.88x10-6). LAIR2 rs35821889, encoding a collagen receptor inhibiting collagen-induced platelet aggregation, was associated with 4-fold higher odds of WRH (p=8.36x10-6) among AA men. MOV10L1 rs138244, which influences expression of PANX2, a structural component of gap junctions, was associated with 2-fold higher odds of WRH among EA men (OR 2.37; p=8.98x10-6).

We identified age, ancestry, and sex-specific variants associated with WRH, demonstrating the importance of conducting stratified analyses. Additionally, a comprehensive evaluation of drug-related genes not commonly included in candidate gene studies, implicated pathways involved in cell adhesion, endothelial barrier integrity, and platelet aggregation that contribute to WRH. This offers insight into underlying biological processes/pathways that contribute to altered drug response.

Session Title: Pharmacogenomics Poster Session I

PB2165 iGenoMed-MTT: A prospective multi'omic study of response to molecularly targeted therapies in IBD.

Authors:

G. Boucher[1], J. Thompson Legault[1], V. Mercier[1], R. Battat[2,3], A. Bitton[4,5], J. Rioux[1,3]; [1]Montreal Heart Inst., Montreal, QC, Canada, [2]Ctr. Hosp.ier de l'Université de Montréal, Montreal, QC, Canada, [3]Université de Montréal, Montreal, QC, Canada, [4]McGill Univ. Hlth.Ctr., Montreal, QC, Canada, [5]McGill Univ., Montreal, QC, Canada

Inflammatory bowel diseases (IBD) result from chronic inflammation of the gastrointestinal tract. These are heterogenous, lifelong and debilitating conditions affecting over a million Americans. With several drugs now available to treat the disease, there is a need for physicians to identify which drug is more likely to be effective for a given patient, instead of the current trial and error approach. Over 200 genetic loci have been identified for IBD, implicated in many biological pathways. We hypothesize that measuring activity of these pathways in patients could provide information to link the underlying biology to response to therapy, with the objective to help physicians make informed decisions about treatment course. This prospective study focuses on the most recent class of molecularly targeted therapy (MTT), which can be highly effective, but also costly and potentially risky. By recruiting all patients before initiation of a new MTT, this real-life prospective approach has the potential to identify predictive biomarkers of response, and directly compare different therapies. We collected serum samples from 166 patients before initiation of treatment, and at 1st assessment visit (~16 weeks) for a subset (24) of these patients. Serum samples were analyzed for >100 protein biomarkers and >1500 lipid metabolites. Whole exome sequencing and GWAS data was generated for all participants.We found elevated levels of TNF at baseline in patients who reported prior use of anti-TNF, more than a month after anti-TNF cessation. Elevated levels of TNF were also found post treatment for patients under anti-TNF in the context of this study. Notably, these elevated levels didn't correlate with other cytokines. As an example, CXCL10, which was positively correlated with TNF levels in bio naive patients, was not found elevated in patients who reported prior use of anti-TNF. This suggests an accumulation of inactive (complexed) TNF during treatment. A similar pattern was observed for IL-12 and IL-12B after anti-IL12/IL23 treatment. Corticosteroids are often prescribed to control inflammation in IBD patients. We identified many cytokines to be associated with current use of corticosteroids. Further analyses of baseline proteomics identified promising proteomics biomarkers for prediction of remission after biologic treatment, some of which are treatment specific. This study provides promising insight into the biologic pathways implicated in response to therapy in IBD, worth investigating in larger cohorts. For this purpose, additional recruitment is ongoing in a multicentric effort, with a target of additional 600 patients.

Session Title: Pharmacogenomics Poster Session II

PB2166 Impact of global and local ancestry on pharmacogenetic analyses in African American multiple sclerosis patients

Authors:

W. Brugger, L. C. Silva, M. F. Davis; Brigham Young Univ., Provo, UT

Multiple sclerosis (MS) is a disease of the central nervous system and is treated with disease-modifying therapies such as interferon-β (IFN-β). Up to 60% of patients treated with IFN-β develop abnormal biochemical liver test results, and 1 in 50 experiences drug-induced liver injury (DILI). Single nucleotide polymorphism (SNP) rs2205986 has demonstrated a statistical association with DILI in patients of European ancestry (EP), though this association has not been explored in other patient populations. A differential risk for MS development, progression, comorbidity, and severity has been documented extensively among African American patients (AAP) compared to EP; poorer treatment responses have also been observed. We aim to evaluate the association of rs2205986 with IFN-β-induced liver injury in AAP in the context of both global and local genomic ancestry. We utilized de-identified electronic health records (EHRs) for 70 AAP from Vanderbilt University's BioVU database. We plan to expand our dataset with approximately 361 AAP whose genomic and EHR data are available through the All of Us Research Program. Patient genomes were genotyped on the Illumina MEGA^{EX} platform and imputed using the TOPMed imputation server. We estimated global ancestry via ADMIXTURE using CEU and YRI datasets from the 1000 Genomes Project as European and African controls, respectively. We estimated local ancestry values for all AAP via RFMix with the same reference genomes, highlighting regions that exhibit local ancestry differing substantially from the genome average. We will follow up on this preliminary work through 1) regression-based analyses to identify correlations between treatment response and ancestry peaks, and 2) fine-mapping analyses to identify novel or previously reported SNPs in these regions that may drive the associative signals to establish ancestry-driven discrepancies in MS treatment response. In the global ancestry analysis, we discovered that the majority of our samples had predominantly European ancestry. Local ancestry calculations highlighted multiple regions as possessing higher-than-average African genomic ancestry. A preliminary linear regression analysis revealed an association between rs2205986 and abnormal levels of alanine aminotransferase ($p = 0.02$). Our preliminary results support that AAP are susceptible to DILI resulting from IFN-β treatment, although differ from EP in whom rs2205986 was associated with abnormal levels of aspartate aminotransferase and alkaline phosphatase. Our results represent information that may be used to guide DMT treatment strategies for AAP and reduce unnecessary risks to improve patient outcomes.

Session Title: Pharmacogenomics Poster Session III

PB2167 Impaired lysosome by bafilomycin A1 induced TFEB pathway in colorectal cancer cells

Authors:

J-Y. Lee; Korea Univ, Seoul, Korea, Republic of

Lysosomes is an important cellular organelle which plays a key role in cellular homeostasis by controlling both cellular clearance and energy production to respond to environmental cues, which is dependent on lysosomal acidification. Inability to maintain lysosomal acidic pH is associated with lysosomal dysfunction as well as lysosome dependent cell death. Bafilomycin A1 (BaMA1), a vacuolar type H^+-ATPase (v-ATPase) inhibitor reducing lysosomal pH, applied to investigate its anticancer effect on colon cancer cells *in vitro* and *in vivo* zebrafish model in this study. The results showed colon cancer cells were sensitive to BaMA1, but not in normal colon fibroblast cell, CCD-18Co. Lysosomal dysregulation upon BaMA1 treatment in colon cancer cell was verified by the release of cathepsin B into the cytoplasm with enlarged lysosomes and massively increased acidic compartments, indicating impaired lysosomal membrane stability. To elucidate target gene(s) or pathway of BaMA1, RNA sequencing was performed, and activated TFEB pathway was identified. Activated TFBE further confirmed by nuclear translocation of TFBE after treatment of BaMA1. The TFEB, the master regulator of lysosomal biogenesis and autophagy through targeting the mTORC1 pathway. This regulation in turn increased gene expressions in the autophagy lysosome pathway and increased autophagosomal and lysosomal biogenesis. However, autophagy flux was blocked in our study, which further investigation of deeper mechanism is ongoing.

Session Title: Pharmacogenomics Poster Session I

PB2168 Improving *G6PD* variant interpretation through multiplexed functional assessment.

Authors:

R. Geck, M. Dunham; Univ. of Washington, Seattle, WA

Glucose-6-phosphate dehydrogenase (G6PD) deficiency affects over 500 million people. Individuals with G6PD deficiency are often asymptomatic but can experience hemolysis after infections or exposure to oxidants, including many drugs. However, diagnosis by activity-based assays give false negatives during hemolytic crises, and since *G6PD* is on the X chromosome, most heterozygotes have normal activity but must be identified to avoid triggers during pregnancy. Therefore, accurate interpretation of *G6PD* genetic testing is valuable for preventing adverse drug reactions.

Recent updates to *G6PD* variant classification systems and guidelines for medication use make now the optimal time to study *G6PD* variant function for maximal clinical actionability. Therefore, we sought to collect reports of *G6PD* variant function and to conduct a deep mutational scan (DMS) of *G6PD* to measure function of additional variants.

We previously published a collection of 1,341 *G6PD* variants from literature and databases, including 118 newly identified through All of Us. By applying ACMG guidelines for interpretation of sequence variants, we interpreted 186 VUS, bringing the total number of potentially actionable *G6PD* variants to 400. However, hundreds of variants remain of uncertain significance.

We are expanding these functional data by conducting a DMS of *G6PD*. We transformed *S. cerevisiae* lacking the *G6PD* homolog *ZWF1* with a library of all single missense *G6PD* variants. We are growing the yeast library under oxidative stress to select for G6PD activity, and deeply sequencing over time to calculate variant frequency. From this we can infer variant function by ability to rescue growth. We will benchmark our functional assay with variants of known clinical effect, compute evidence for levels of support, and incorporate this functional data into *G6PD* variant interpretations.

We are also introducing SNVs of common background haplotypes onto our variant library to conduct DMS on multiple genetic backgrounds. *G6PD* variants have historically been studied on a Caucasian background even though that does not represent the populations most affected by G6PD deficiency. We measured the function of select multiple missense variants in our yeast system and observed that the function is not always predicted from additive effects. Given this genetic complexity, our DMS on multiple backgrounds addresses an unmet need to systematically investigate the effects of common variants on rare G6PD variant activity. Altogether, our study will enable clinical interpretation of a larger number of *G6PD* alleles and ensure that these interpretations are applicable to the affected populations.

Session Title: Pharmacogenomics Poster Session II

PB2169 Inference and characterization of *CYP2D6* whole gene deletions from a SNP array in the Michigan Genomics Initiative Biobank.

Authors:

B. Vanderwerff[1], A. Pasternak[2,3], L. Fritsche[1], E. Bertucci-Richter[1], S. Patil[1,4], M. Boehnke[1], X. Zhou[1], S. Zollner[1,5], D. Hertz[2], M. Zawistowski[1]; [1]Dept. of Biostatistics and Ctr. for Statistical Genetics, Univ. of Michigan Sch. of Publ. Hlth., Ann Arbor, MI, [2]Dept. of Clinical Pharmacy, Univ. of Michigan Coll. of Pharmacy, Ann Arbor, MI, [3]Michigan Med., Ann Arbor, MI, [4]Dept. of Computational Med. and Bioinformatics, Ann Arbor, MI, [5]Dept. of Psychiatry, Ann Arbor, MI

Background: The Michigan Genomics Initiative (MGI) Biobank links genetic and electronic health record (EHR) data for >70K Michigan Medicine patients in a central data set for investigators with diverse research applications. The ability to assess variation in genes that interact with pharmaceuticals (pharmacogenes) to EHR-based phenotypes makes biobanks an attractive platform for large-scale pharmacogenomics (PGx) studies. PGx studies often use star allele nomenclature, which maps genetic variation in pharmacogenes to diplotypes which are annotated for overall protein function level (activity phenotypes). Many star alleles are called based on SNPs/indels, but for some pharmacogenes, this call depends on one or more structural variants (SV). *CYP2D6* is an important pharmacogene that metabolizes ~25% of current drugs. Accurate *CYP2D6* phenotyping requires calling *5 (whole gene deletion), one of the most common SV-based loss-of-function alleles. In this study we used a support vector machine (SVM) to classify *5 carriers among MGI participants genotyped on the CoreExome (CE) SNP array, which lacks designed content for *5 inference. **Methods:** We trained a SVM on CE array signal intensity (LRR) from 10 intragenic *CYP2D6* SNPs from 449 multi-ancestry samples with truth-known *5 calls. We evaluated SVM accuracy in a cross-validation (CV) of the training data and in a multi-ancestry test set of 42 samples. **Results:** PCA and tSNE visualizations of LRR show clusters that differentiate the known carriers and non-carriers. The SVM accurately classified all training samples in the CV analysis and the test set. The SVM classified 4,108 *5 carriers among 59,776 participants genotyped on the CE, we additionally called 613 *5 carriers among 9,506 participants genotyped on the Global Screening Array, which has explicit content for calling *5. The frequency of *5 was ~3.5% among all participants and ~3.4% among European ancestry participants (n=60,358), closely aligned with the 2.9% European frequency in PharmGKB (n=59,826). Incorporating *5 calls either reduced or introduced uncertainty for the activity phenotype of 6.4% of participants, including ~3,200 estimated to have received a prescription with a CPIC guideline for reduced *CYP2D6* function. **Conclusions:** We demonstrate an accurate, scalable method for *5 carrier classification using an array without explicit *5 content. While our method doesn't replace comprehensive PGx testing, it enables biobanks to capture *5-based drug-gene-interactions with existing array data and minimal training samples. Moreover, we envision this method may classify carriers of known deletion alleles of other pharmacogenes.

Session Title: Pharmacogenomics Poster Session III

PB2170 Leveraging Colorado Center for Personalized Medicine and electronic health records to perform HLA typing with an emphasis on potential pharmacogenetic alleles

Authors:

S. Chavan, H. Smith, M. Fisher, J. Shortt, N. Pozdeyev, L. Meng, K. Crooks, C. Gignoux, N. Rafaels, Colorado Center for Personalized Medicine (CCPM); Univ. of Colorado Anschutz Med. Campus, Aurora, CO

<Background: The Human Leukocyte Antigen (HLA) is a highly polymorphic region in the genome and plays a critical role in immunogenetics and disease research. Genotyping HLA alleles is essential for analysis of diseases and adverse drug reactions. Given the large and diverse sample size and access to clinical information through electronic health record (EHR), the data from Colorado Center for Personalized Medicine (CCPM) biobank is an excellent resource to study HLA and its role in health and disease.
Methods: HLA calling was performed using HLA-TAPAS (SNP2HLA) from the Michigan Imputation server on 73,346 CCPM Biobank samples comprising 49,372 samples from whole exome sequencing (WES) and 24,079 samples from Multi-Ethnic Genotyping Array (MEGA). Imputation was based on a global reference panel with two-field resolution comprising of 21,546 individuals with whole genome sequencing samples from five global populations including admixed African, European, East Asian, Latino and South Asian populations. Leveraging the imputed HLA alleles, we performed Phenome-wide association study (PheWAS) on the HLA alleles for 1563 phecodes available in the CCPM biobank using firth logistic regression as implemented in REGENIE. To study HLA and its role in drug reaction, we specifically homed in on potential pharmacogenetic HLA alleles namely, HLA_A*31:01 (carbamazepine), HLA_B*57:01 (abacavir), HLA_B*58:01 (allopurinol) and HLA_B*15:02 (carbamazepine, oxcarbazepine, phenytoin).
Results: To perform sensitivity analysis of the HLA genotype calls, we observed 16/16 patients positive for HLA_B*57:01 confirmed by EHR provided HLA. PheWAS analyses replicated association with psoriasis (P=7.6e-22) with HLA_B*57:01. Adding any use of allopurinol (N=3639) or epilepsy drugs (N=5510; carbamazepine, oxcarbazepine, phenytoin) as a covariate revealed associations with kidney diseases and epilepsy, respectively. Interestingly, HLA_B*15:02 revealed association with adverse drug events despite fewer patients taking medications and being positive for HLA_B*15:02 than expected, whereas HLA_B*58:01 showed no association with adverse drug events even though the number of patients taking epilepsy drugs was roughly similar.
Conclusions: HLA genotyping data from CCPM biobank is a robust resource for research-based HLA and disease association in multi-ancestry population and requiring additional validation also provides a promising resource for drug testing recommendation based on HLA calls.>

Session Title: Pharmacogenomics Poster Session I

PB2171 Leveraging donor variability in primary human hepatocytes to identify variation in drug metabolism in African Americans: CYP1A2 and glucoronidation

Authors:

C. Clark, T. De, G. Yang, C. Alarcon, J. Avitia, M. Perera; Northwestern Univ., Chicago, IL

Background: Genetic variation in drug metabolizing enzymes (DMEs), their regulatory regions, or within regulatory proteins can contribute to interindividual variability of drug response and adverse drug reactions. Genetic studies have underrepresented minority populations, such as African Americans (AA), resulting in poor prediction of drug response in these groups. This study aims to identify variation affecting drug metabolism using a clinically relevant phenotype in an AA cohort. **Method:** Hepatocytes were extracted from 75 AA cadaveric livers, assayed with probe substrates phenacetin (PHE) and acetaminophen (APAP), and underwent genome-wide genotyping. These drugs were used to examine variability of *CYP1A2* and glucuronidation through the measurement of metabolite formation rate (MFR). A genome-wide association study was conducted for each MFR with age, sex, and principal components as covariates. Additionally, a candidate gene study was conducted using genes from each probe's PharmGKB drug metabolism pathway. Significant results were prioritized based off publicly available functional data. **Results:** Loci reaching nominal significance were identified. rs10046278 (PHE, $p=9.27x10^{-6}$) is an eQTL for *TPMT*, a known phase II DME. rs166641 (APAP, $p= 1.83x10^{-6}$) is located within the 3'UTR of *SFXN1* and is an eQTL and sQTL for *SFXN1*. rs8191167 (APAP, $p= 1.03x10^{-6}$) is located within an enhancer region of *HSD17B2.* Both of these SNPs show large allele frequency difference between populations. Through candidate gene analysis, we identified a novel association of rs10203853 (APAP, $p= 5.7x10^{-5}$) near the *UGT1A* gene family and is a pQTL for *UGT1A6*. This SNP has been linked to differences in nicotine glucuronidation in AAs and shows associations to UGT1A-linked phenotypes (i.e., Gilbert's Disease, Bilirubin excretion) in a pheWAS of the combined FinnGen, UKBiobank and GWAS catalog datasets.**Conclusion:** We identified known and novel variants linked to drug metabolism. Variation in *TPMT* has been associated to increased risk of adverse events with thiopurine drugs used in auto-immune diseases and cancer. The UGT1A family is responsible for over half of glucuronidation reactions, a process which 12% of all drugs undergo. Other results have mechanistic links to DME regulation but have not been previously identified. For example, SFXN1 is a transporter involved in a-ketoglutarate metabolism, and HSD17B2 is a known regulator of steroid hormones, both of which have known links to DME gene regulation via their steroid hormone transcription factors. These findings represent some of the potential genetic variants contributing to known differences in drug response.

Session Title: Pharmacogenomics Poster Session II

PB2172 Leveraging the All of Us biobank to build multi-ancestry polygenic scores for NSAID-induced gastrointestinal bleeding.

Authors:

K. Keat, E. Ricciotti, K. Barekat, K. Cardone, G. A. FitzGerald, D. Kim, M. D. Ritchie; Univ. of Pennsylvania, Philadelphia, PA

Non-steroidal anti-inflammatory drugs (NSAIDs) are a class of widely used drugs for the management of pain, fever, and inflammation in a broad spectrum of diseases. NSAID use has been associated with both mild and life-threatening adverse drug reactions (ADRs), including gastrointestinal bleeding and acute coronary syndrome. Given the severity of these outcomes, broad use of NSAIDs, and the large number of patients affected, there is a significant need to predict individual risk of ADR from NSAIDs. The Clinical Pharmacogenetics Implementation Consortium (CPIC) has published guidelines for clinicians to modify NSAID treatment in the presence of *CYP2C9* loss-of-function variants; these variants result in reduced clearance of NSAIDs and increased risk of ADRs. However, *CYP2C9* alone explains a relatively small proportion of risk of ADR, which is currently better predicted using clinical covariates such as age, sex, concomitant drugs, and comorbidities. To gain a better understanding of the heritable risk of NSAID ADRs, we have performed the first large-scale genome-wide association study for gastrointestinal bleeding following NSAID use in a diverse population using the All of Us cohort; identifying several significant associations. We also validate these associations in other medical biobanks including the Penn Medicine BioBank and the UK Biobank and subsequently use them to evaluate the performance of a polygenic risk score (PRS) derived from the All of Us GWAS summary statistics. Furthermore, to augment the cross-ancestry performance of our PRS, we are integrating it with a transcriptomic risk score (TRS) based on imputed transcriptomes. Discovery of novel genomic risk factors for ADRs improves both our biological understanding and ability to predict NSAID ADRs. Subsequent work will be conducted to integrate pharmacogenetic, pharmacogenomic, and clinical factors such as concomitant drugs in order to build a more comprehensive and clinically actionable predictor for NSAID ADRs.

Session Title: Pharmacogenomics Poster Session III

PB2173 Long-read sequencing for *CYP2D6* genotyping of Korean: Preliminary result

Authors:

B. Kim; SMC, Seoul, Korea, Republic of

CYP2D6 is a drug metabolizing enzyme responsible for about 20% of drugs. The *CYP2D6* gene that determines the enzyme activity of CYP2D6 is about 4.3 kilobases (kb), located in the long arm of chromosome 22. *CYP2D6* is a polymorphic gene, with about 133 known alleles and different enzyme activities depending on the combination of alleles. Meanwhile, pseudogenes called *CYP2D7* and *CYP2D8* are located upstream of the *CYP2D6* gene, which makes the exact genotyping of *CYP2D6* harder. In this study, long-read sequencing and CRISPR/Cas9 method were applied for CYP2D6 genotyping to obtain the adjacent area, and the structural variations of CYP2D6 are revealed. With the samples of which CYP2D6 genotypes were determined previously, CRISPR/Cas9 target enrichment for CYP2D6 regions was done. In addition, long-range PCR for CYP2D6 genes and the adjacent area was amplified for long-read sequencing with Nanopore. Flongle and MinION flow cell (R9.4.1) was used for the sequencing by MinION Mk1B. After long-range PCR for the *CYP2D6* regions, long-read sequencing was done with Nanopore MinION, but the read sequence was not aligned at the target area. However, after CRISPR/Cas9 target enrichment for CYP2D6, the read sequence was aligned at the target area around the *CYP2D6* gene. CRISPR/Cas9 target enrichment for long-read sequencing was appropriate for the *CYP2D6* regions rather than long-range PCR.

Session Title: Pharmacogenomics Poster Session I

PB2174 Lymphocyte count-derived polygenic score and interindividual variability in CD4 T-cell recovery in response to antiretroviral therapy

Authors:

K. Cardone[1], S. Dudek[1], K. Keat[1], Y. Bradford[1], Z. Cindi[1], E. Daar[2], R. Gulick[3], S. Riddler[4], P. Sinxadi[5], D. Haas[6], M. Ritchie[1]; [1]Univ. of Pennsylvania, Philadelphia, PA, [2]Lundquist Inst. at Harbor-UCLA Med. Ctr., Torrance, CA, [3]Weill Cornell Med. New York, New York, NY, [4]Univ. of Pittsburgh, Pittsburgh, PA, [5]Univ. of Cape Town, Cape Town, South Africa, [6]Vanderbilt Univ. Med. Ctr., Nashville, TN

Access to safe and effective antiretroviral therapy (ART) is a cornerstone in the global response to the HIV pandemic. Among people living with HIV, there is considerable interindividual variability in absolute CD4 T-cell recovery following initiation of virally-suppressive ART. The contribution of host genetics to this variability is not well understood. We explored the contribution of a polygenic score which was derived from large, publicly available summary statistics for absolute lymphocyte count from individuals in the general population (PGS_{lymph}). There are not available summary statistics to derive a CD4 T-cell PGS. We explored associations with baseline CD4 T-cell count prior to ART initiation (n=4959), and change from baseline to week 48 on ART (n=3274) among treatment-naïve participants in prospective, randomized ART studies of the AIDS Clinical Trials Group. We separately examined African and European PGS_{lymph}, and applied to all participants, and to African and European ancestral groups separately. Multivariate models that included PGS_{lymph}, baseline plasma HIV-1 RNA, age, sex, and 15 principal components (PCs) of genetic similarity explained ~26-27% of variability in baseline CD4 T-cell count, but PGS_{lymph} accounted for <1% of this variability. Models that also included baseline CD4 T-cell count explained ~7-9% of variability in CD4 T-cell count increase on ART, but PGS_{lymph} accounted for <1% of this variability. In univariate analyses, PGS_{lymph} was not significantly associated with baseline or change in CD4 T-cell count. Among individuals of African ancestry, African PGS_{lymph} was significantly associated with change in CD4 T-cell count in the multivariate model but not the univariate model. When applied to lymphocyte count in the general population (Penn Medicine BioBank), PGS_{lymph} explained ~6-10% of variability in multivariate models (including age, sex, and PCs) but only ~1% in univariate models. In summary, a lymphocyte count PGS derived from the general population was not consistently associated with CD4 T-cell recovery on ART. Adjusting for clinical covariates is important when estimating polygenic effects.

Session Title: Pharmacogenomics Poster Session II

PB2175 † Machine Learning Prediction on Ketamine Treatment Response in Posttraumatic Stress Disorder Based on Methylation Capture Sequencing Data

Authors:

X. Zhang[1], Y. Hu[2], C. Abdallah[3], J. Roache[4], L. Averill[3], R. Vandenhoudt[1], D. Williamson[5], A. Peterson[4], J. Krystal[6], K. Xu[7]; [1]Yale Univ., New Haven, CT, [2]Natl. Cancer Inst., Rockville, MD, [3]Michael E. DeBakey VA Med. Ctr., Houston, TX, [4]Univ. of Texas Hlth.Sci. Ctr. at San Antonio, San Antonio, TX, [5]Duke Hlth., Durham, NC, [6]Yale Sch. of Midicine, New Haven, CT, [7]Yale Sch. Med., New Haven, CT

Background: Post-Traumatic Stress Disorder (PTSD) is a psychological disorder that occurs after exposure to a traumatic event. Ketamine, primarily used for anesthesia and pain management, has shown potential as a treatment for PTSD. However, predicting individual response to ketamine treatment for PTSD symptoms remains challenging. In this study, we aimed to identify DNA methylation sites that could serve as biomarkers for ketamine response in veterans with PTSD using a machine learning approach. **Methods**: A total of 161 individuals diagnosed with PTSD took part in the study and were divided into two groups: Ketamine (n=106) and Placebo (n=55). The severity of PTSD symptoms was assessed using the PTSD Check List for DSM-5 (PCL-5) questionnaire. A Linear Mixed Model (LMM) was used to analyze the trajectory of PTSD symptoms over time. Based on the median LMM slope of PCL-5, responders (R) and non-responders (NR) were identified, resulting in four response patterns: Ketamine-R (KR), Placebo-R (PR), Ketamine-NR (KNR), and Placebo-NR (PNR). DNA methylation in blood was profiled using SureSelectXT Methyl-Seq for MC-seq. We first conducted a Linear Interaction Model (LMM) analysis to select candidate CpG sites in the genome for treatment response. The samples were randomly divided into a training set (N= 107) and a testing set (N=54), and the Support Vector Machine (SVM) method was employed to identify the best-performing model for ketamine treatment. **Results**: The top 5,000 CpGs identified by the LMM analysis were chosen for subsequent machine learning analysis. A model consisting of 1,553 CpGs exhibited the best performance in the testing set. Notably, these predictive CpGs were located in genes previously associated with PTSD, such as *ELFN1*, *MAD1L1*, and *WNT5A*. Hierarchical cluster analysis using DNA methylation beta values of the 1,553 CpGs successfully divided the samples into two distinct clusters. One cluster primarily included ketamine responders, while the other cluster mainly included placebo responders. The distribution of individuals with KR, KNR, PR, and PNR in these two clusters showed significant differences (p=4.9E-012). Therefore, the machine learning-derived methylation clusters effectively differentiated responders and non-responders to ketamine and placebo treatments. **Conclusions**: Our findings demonstrate the potential of DNA methylation as a biomarker for predicting response to ketamine and placebo treatments for PTSD.

Session Title: Pharmacogenomics Poster Session III

PB2176 Metabolomic Signature of Canagliflozin

Authors:

M. Montasser, Z. S. Yazdi, E. A. Streeten, H. B. Whitlatch, S. I. Taylor, A. L. Beitelshees; Univ. of Maryland Sch. of Med., Baltimore, MD

Sodium glucose cotransporter-2 inhibitors (SGLT2i) as second line of diabetes treatments provide many clinical benefits: HbA1C-lowering, weight loss, as well as cardiovascular (CV) and renal benefits. However, like all drugs, SGLT2i can cause rare but severe side effects. The most common are: genital mycotic infections, urinary tract infections and volume-depletion related side effects. Less common, but more severe side effects include fractures, decrease in bone mineral density, urosepsis, and ketoacidosis. The purpose of The Genetics of Response to Canagliflozin (GRC) study is to use multiomics approach to better understand the underlying biological mechanisms of these different effects and identify determinants of variation in response that can distinguish those who will be benefited most and harmed least by the use of SGLT2i, which will pave the road to reaching the precision medicine goals. Biomarker analysis and Metabolomic profiling was performed for the first 216 generally healthy Amish participants in the GRC at 3 time points: baseline before drug exposure (day 1) and after 2 and 5 doses of Canagliflozin (300 mg/d) (day 3 and day 6). 24-hour glucosuria was measured at day 3. Phenotypic and metabolomic signatures of response were determined using repeated measure analyses adjusting for age, age^2, sex, BMI, and familial correlation. There was about 5-fold variation (mean 33 ± 9 g_{gluc}/g_{cr}) in glucosuria between the participants. The response of bone biomarkers was consistent with a negative impact on bone health as presented by increase in PTH and C-terminal telopeptide (CTX) at day 3 and 6, initial increase of phosphate (P) and FGF23 at day 3 and then decrease at day 6, while $1,25(OH)_2D$ was decreased at day 3 and then returned to baseline at day 6. Consistent with beneficial CV effects, we observed a significant decrease in systolic and diastolic blood pressure and uric acid. There was highly significant increase in all ketone bodies on day 3 but the levels returned to baseline at day 6. Out of the 249 profiled metabolites, 189 were significantly altered at day 3 but only 85 were altered at day 6 compared to day 1. The large increase at day 3 reflects the acute response to the drug before the compensatory homeostatic mechanisms brings many metabolites back to normal. The most significantly changed metabolite was increased alanine (ALA), which is consistent with previously reported interactions between renal handling of alanine, glucose, and phosphate. Baseline higher ALA was also associated with better SBP response to canagliflozin with more dramatic decrease, which could represent a novel biomarker for BP lowering or CVD protection in response to SGLT2i.

Session Title: Pharmacogenomics Poster Session I

PB2177 † Methods to prevent hepatitis B vaccine escape elucidated through a genetic association study of antibody responses in Bangladeshi infants

Authors:

G. Butler-Laporte[1,2,3], K. Auckland[4], T. Carstensen[4], G. Wojcik[5], A. Y. Chong[6], C. Pomilla[4], J. Noble[7], G. Smits[8], S. Wareing[9], F. R. van der Klis[10], K. Jeffery[9], B. Kirkpatrick[11], S. Sirima[12], A. Elliott[13], J. Richards[14], A. V. Hill[15], P. Duggal[5], M. S. Sandhu[16], R. Haque[17], W. A. Petri, Jr[18], A. J. Mentzer[1]; [1]Wellcome Ctr. for Human Genetics, Oxford, United Kingdom, [2]Lady Davis Inst., Montréal, QC, Canada, [3]McGill Univ. Hlth.Ctr., Montréal, QC, Canada, [4]Wellcome Trust Sanger Inst, Hinxton, Cambridgeshire, United Kingdom, [5]Johns Hopkins Bloomberg Sch. of Publ. Hlth., Baltimore, MD, [6]Univ. of Oxford, Oxford, United Kingdom, [7]Children's Hosp. Oakland Res. Inst., Oakland, CA, [8]Natl. Inst. for Publ. Hlth.and the Environment, Bilthoven, Netherlands, [9]Microbiol. Dept., John Radcliffe Hosp., Oxford, United Kingdom, [10]Natl. Inst. for Publ. Hlth.and the Environment, Bilthoven, United Kingdom, [11]Univ. of Vermont Coll. of Med. and Vaccine Testing Ctr., Burlington, VT, [12]Groupe de Recherche Action en Santé, Ouagadougou, Burkina Faso, [13]London Sch. of Hygiene & Tropical Med., London, United Kingdom, [14]McGill Univ., Montreal, QC, Canada, [15]The Jenner Inst., Oxford, United Kingdom, [16]Imperial Coll. London, London, United Kingdom, [17]Intl. Ctr. for Diarrhoeal Disease Res., Dhaka, Bangladesh, [18]Univ. of Virginia Sch. of Med., Charlottesville, VA

Human leukocyte antigen (HLA) variants influence vaccine responses, but multiple challenges limit clinical translation. These include over-reliance on European-derived imputation panels, complex linkage disequilibrium, high pleiotropy, and difficulties interpreting biological mechanisms assigned to allelic associations. We sought to overcome these challenges to deliver a public health intervention, specifically in relation to hepatitis B virus (HBV) vaccination, currently delivered worldwide during infancy as an S, *adw* subtype. Our findings offer an opportunity to move towards universal vaccine responsiveness and halt HBV vaccine escape mutants.

We studied the genetic determinants of antibody responses against eight childhood vaccines in 1,096 children from Bangladesh, a region with well recognized HBV vaccine escape mutants. Using genome-wide associations, we found that the HLA locus was associated with responses against four vaccine antigens. We then constructed and validated the largest ever published south Asian ancestry HLA imputation reference panel (n = 9,448) to fine-map our signals. Specifically, we identified DPB1*04:01 as a determinant of higher antibody response against the HBV surface antigen, HBsAg ($p=4.5x10^{-30}$).

We then interrogated the binding affinity HBsAg peptides to HLA receptors *in silico*. We found that binding of HBsAg peptides to DPB1 dimers was proportional to antibody levels observed genetically. This was replicated in a cohort of African children, and at the *HLA-DQB1* and *HLA-DRB1* genes where independent associations were found in the more genetically heterogeneous African cohort. This has two clinically significant findings. First, individuals carrying HLA alleles with reduced HBsAg vaccine response, as defined by genetics and binding affinity, would benefit the most from vaccines containing HBsAg pre-S isoforms rather than the currently used *adw* types. This are consistent with vaccination trials showing that pre-S vaccines perform better than the contemporary S vaccines. Second, the difference in binding affinity between protective HLA alleles compared to other alleles was specifically at its highest in the a-determinant locus of HBsAg. This matches well reported orthogonal epidemiological observations of mutations at the a-determinant leading to HBV vaccine escape, particularly in Bangladesh. Hence, individuals with genetic predisposition to reduced response to HBV S isoform vaccines may provide a simple evolution opportunity for HBV to escape host immunity with a single amino acid mutation. Thus, pre-S containing HBV vaccines should be prioritized to help decrease the burden of these HBV mutants.

Session Title: Pharmacogenomics Poster Session II

PB2178 Methotrexate Drug Responses in Rheumatoid Arthritis are influenced by Methylation Mapping Genes

Authors:

K. ELAYAKOVILAKOM[1], M. Banerjee[2], S. S. Vishnu[3], S. sreenath[4]; [1]Rajiv Gandhi Ctr. for Biotechnology, Thiruvananthapuram, India, [2]Rajiv Gandhi Ctr Biotech, Trivandrum Kerala, India, [3]Government Med. Coll., Thiruvananthapuram, India, [4]PRS Hosp., Thiruvananthapuram, India

Rheumatoid arthritis (RA) is a crucial auto-immune and inflammatory joint disease characterized by loss of self-tolerance, severe cartilage loss, and subchondral bone erosions. The pathogenesis of rheumatoid arthritis remains unclear and this poses challenges in the effective management of disease-modifying anti-rheumatic drugs (DMARD) which also results in an unpredictable spectrum of side effects. Methotrexate (MTX) is a folic acid analog commonly used as DMARD for the treatment of rheumatoid arthritis. The variation in the responses in the MTX drug in each diseased individual may be due to the genetic variation that occurs in the drug transporters and metabolizing genes in one carbon metabolic pathway. With this background, our present study focuses on genetic variation liable for folate metabolic genes for methotrexate drug responses in rheumatoid arthritis patients**.** In the present study, the drug metabolizer gene *MTRR* a methylation mapping gene shows a significant association with MTX drug responses in RA individuals. The functional variants of the folate pathway gene *MTRR* rs1532268 CC/CT show greater efficacy to MTX drug in individuals with RA. None of the other folate genes show an association with drug responses. The study concludes that the genetic variants of the folate pathway genes are real predictive markers for methotrexate drug responses in RA.

Session Title: Pharmacogenomics Poster Session III

PB2179 Multi-omics and Functional Analysis Identifies Novel Genetic and Methylation Markers Contributing to Antipsychotic-Induced Metabolic Syndromes

Authors:

S. Qin[1], W. Zhou[1], J. Sun[2], C. Huai[1], H. Huang[3], L. He[1]; [1]Shanghai Jiao Tong Univ., Shanghai, China, [2]Jiangsu Univ., Zhenjiang, China, [3]MGH, Boston, MA

Antipsychotic-induced metabolic syndrome (APs-induced Mets) is the most common adverse drug reaction, affecting over 60% of psychiatric patients. Although extensive research has been conducted on the etiology of APs-induced Mets, there is a lack of integrated analysis of genetic and epigenetic factors. In this study, we performed genome-wide association study (GWAS), whole-exome sequencing (WES) and epigenome-wide association studies (EWAS) in schizophrenia (SCZ) patients with or without APs-induced Mets to elucidate the underlying mechanisms. Subsequently, we conducted *in vitro* and *in vivo* functional validations. Our case-control based genome-wide association analysis of 444 individuals with APs-induced Mets and 546 controls that identified six loci (P < 5E-7). Recode the Mets severity z-score (MetS-z) found 1 genome-wide-significant loci(P < 5E-8). Furthermore, set-based omics analysis revealed an imbalance of rare functional variants across the leptin and peroxisome proliferator-activated receptor (PPAR) gene sets in APs-induced Mets compared to Non-Mets cohort. Additionally, we discovered hypermethylation of *ABCG1* (chr21:43642166-43642366, adjusted P < 0.05) in APs-induced Mets compared to Non-Mets, and this hypermethylation was associated with elevated levels of total cholesterol (TC) and triglycerides (TG) in HepG2 cells. Candidate genes identified from the omics studies were further screened in *C. elegans*, and we validated associations of 17 genes with olanzapine (OLA)-induced fat deposition. Several of these genes exhibited differential expression in the Mets cohort and APs-induced in vitro/in vivo models compared to controls, thereby confirming the validity of the omics study. Notably, overexpression of one of the most significant genes, *PTPN11*, led to compromised glucose responses and insulin resistance. Pharmacologic inhibition of PTPN11 protected HepG2 cells and C57/BL mouse from APs-induced insulin resistance and disorder lipid deposit. These findings provide important insights into the mechanisms underlying APs-induced Mets.

Session Title: Pharmacogenomics Poster Session I

PB2180 New drug repositioning method using multi-layer omics data based on transformer-based network method

Authors:

s. Lee[1], J-H. Ji[1], M-Y. Song[2], H. Shin[3], B. Kim[1]; [1]ajou Univ., Suwon-si, Korea, Republic of, [2]Kyung Hee Univ., seoul, Korea, Republic of, [3]Ajou Univ., Suwon-si, Korea, Republic of

Recently, deep-learning approaches have been applied to discover new drug efficacies for various cancer types using big multi-omics data of cancer patients. However, multi-omics data typically have high dimensions compared to relatively few patient samples, this imbalance is a recognized bottleneck to apply integrated characteristics of multi-omics in cancer research. To address this issue, Deep learning-based approaches including autoencoder among the dimensionality reduction techniques that are known to have strength in handling high dimensional data with few samples have been applied to devise drug repositioning methods based on deducting representative signatures of multi-omics data. However, this black box model makes it difficult to explain which genes are essential. In this study, we developed a transformer-based drug repositioning model to extract informative features and the potential anti-breast cancer therapeutic target genes to construct the gene network for finding new anti-breast cancer drug efficacy. Unlike previous approaches, our model allows for better interpretability of the essential genes, which can lead to more targeted drug development. It is based on a unified framework of applying the compressed representation learned through transformer from multi-omics data of breast cancer patients in The Cancer Genome Atlas (TCGA) and measuring network similarity between gene and drug networks using graph kernels to predict candidate anti-breast cancer drugs. We validated the performance of our algorithm using various perspectives of data. Our results showed not only an inverse correlation between drugs and cancer multi-omic signatures as the graph similarity score increased but also a pronounced increasing trend in the retention rate of breast cancer marker genes and pathways. These findings demonstrate the effectiveness of our algorithm in identifying potential anti-cancer drug candidates by leveraging the relationship between drug and cancer signatures. From this point of view, the proposed model will pave the way for a new direction to discover new anti-breast cancer efficacy due to representing the biological anti-cancer phenomena by multi-omics data.

Session Title: Pharmacogenomics Poster Session II

PB2181 Novel gene polymorphisms for stable warfarin dose in a Korean population: Genome-wide association study

Authors:

J. Kim[1], S. Lee[2], J. Yee[1], K. Park[3], E. Jang[1], B. Chang[4], H. Gwak[1]; [1]Ewha Womans Univ., Seoul, Korea, Republic of, [2]Yonsei Univ. Coll. of Med., Seoul, Korea, Republic of, [3]Ulsan Natl. Inst. of Sci. and Technology, Ulsan, Korea, Republic of, [4]Bundang CHA Med. Ctr., Seongnam, Korea, Republic of

Background: Warfarin exhibits a narrow therapeutic window and considerable intra- and inter-individual variability in dose requirements. Given that most published studies on genotype-guided dosing originate from European populations, this study aimed to explore novel genetic variants associated with the variability of stable warfarin doses in a Korean population with cardiac valve replacement, utilizing the genome-wide association study (GWAS) approach. **Methods:** A retrospective cohort study using prospectively collected data was conducted from January 1982 to December 2020 at the Severance Cardiovascular Hospital of Yonsei University. The GWAS method was employed to identify associations between genotypes and the warfarin stable dose by comparing the allele frequency of genetic variants among individuals. Subsequently, the influence of genetic and non-genetic factors on dose variability was assessed. **Results:** A total of 229 participants were enrolled, and the strongest signal cluster was identified on chromosome 16 around *VKORC*1 from the GWAS. In addition to *VKORC1* rs9934438, which is a well-known polymorphism, three novel variants (*NKX2-6* rs310279, *FRAS1* rs4386623, and *FAM95C* rs1890109) demonstrated an association with stable warfarin dose requirements in univariate analysis. A predictive algorithm was developed using multivariable analysis, incorporating both genetic and non-genetic factors, accounting for 58.5% of the variations in stable warfarin doses. Within this variability, *VKORC1* rs9934438 and *FRAS1* rs4386623 contributed to 33.0% and 9.9%, respectively. **Conclusion:** This GWAS analysis revealed three novel variants (*NKX2*-6 rs310279, *FRAS1* rs4386623, and *FAM95C* rs1890109) associated with stable warfarin doses. Further research is required to validate these findings and develop personalized treatment strategies for the Korean population.

Session Title: Pharmacogenomics Poster Session III

PB2182 Novel Systematic Congenital Anomaly Case Collection Method and Preliminary Analyses Identifying Causes of Congenital Anomalies

Authors:

E. Brokamp; Vanderbilt Univ. Med. Ctr., Nashville, TN

Congenital anomalies (CAs) affect ~3% of live births, are the leading cause of infant death, and cost health care systems billions of dollars annually. Despite their frequency and consequences, the cause of ~80% CAs is unknown. Further, for the 20% with a potentially identified cause, variability in penetrance suggests additional drivers of risk exist. The interplay between environmental teratogens, such as medication use in pregnancy, and genetic predisposition impacts CA risk, but such interactions are poorly understood. Research to better understand the causes of CAs is hampered by the lack of a uniform CA identification method in electronic health records (EHRs) and biobank databases. We aim to **1.** Demonstrate the first large-scale effort to define and characterize CAs in the EHR and **2.** Test the hypothesis that pharmacogenomic variation impacts CA risk. Previous researcher lacked an accurate and universal method for CA case collection in the EHR. Phecodes, a widely used and easily adaptable phenotyping tool, provide a useful approach for CA identification. Earlier versions of phecodes lacked granularity and specificity in the CA chapter. Therefore, to identify CAs we used the new PhecodeX nomenclature that includes 5.8 times more codes for CAs compared with the previous v1.2 (365 vs 56). We also created a novel quantitative manor for assessing potential causal disease codes, such as known genomic variations, that should be considered in study design. Out of all genetic disease billing codes, 85 (19.7%) have a known association with CAs in the literature. Phenome-wide association studies revealed an additional 16 (3.7%) with one or more significant association with CA-related phecodes (Bonferroni corrected p< 2.75 x10-5). Determining CA-associated billing codes allows researchers to differentiate between idiopathic CAs and those that have a known cause, providing a scalable approach to increase power in studies designed to evaluate the epidemiologic and genetic drivers of CAs. Using phecodeX CA phecodes, we identified 3,278 CA cases and 24,224 controls of genetically determined European ancestry in BioVU. We performed a drug absorption, distribution, metabolism, and excretion (ADME)-wide association study including 5,398 independent SNPs from 292 genes. We adjusted the models for median age, gender, principle components of ancestry, and presence of a genetic disease billing code associated with CAs. Despite the relatively small sample size numerous significant associations at both the gene and SNP levels were identified. The most significant result was for *PPARA*, rs4253795 (OR=1.5, p= 6.33E-05). Future work to expand this cohort is underway.

Session Title: Pharmacogenomics Poster Session I

PB2183 Pharmacoepigenetics of antiepileptic drugs

Authors:

N. Mohan, M. Banerjee; Rajiv Gandhi Ctr. for Biotechnology, Thiruvananthapuram, India

Background: Antiepileptic drugs are the mainstay for the treatment of epilepsy. Studies in human have been shown that these AEDs tend to have teratogenic potential. And it is becoming evident that these pharmaceutical drugs can cause changes in gene expression that persist long after the exposure has ceased. Pharmacoepigenetic studies explain the role of epigenomics in intrapersonal and interpersonal variations in response of individuals to drugs, in the effects of drugs on gene-expression profiles, in the mechanism of action of drugs and adverse drug reactions. **Method:** The present study involves assessing the epigenetic modifications induced by antiepileptic drugs in cell culture model. The effect of AEDs on global DNA methylation and underlying gene expression was studied. The experimental cell line was treated with varying concentration of AEDs at different time intervals, following global DNA methylation assay and the gene expression of epigenetic modulators including DNA methyltransferases (*DNMTs*) and Ten-eleven translocases (*TETs*) was assessed. Along with the antiepileptic drugs, we have supplemented folic acid and assessed the expression pattern of *DNMT*s and *TET*s. **Results:** Data suggests decreased global DNA methylation for AED treated cells and the global DNA methylation changes show similar pattern with respect to *DNMT*s gene expression .As treatment duration increased from 12 to 48 hours, we observed a decrease in *DNMT*s (methylation) and increase in *TETs* (demethylation) expression. The gene expression of *DNMT*s and *TET*s got increased significantly with folic acid supplementation along with AEDs. And this can be attributed by the variability in the one -carbon metabolism genes -*MTHFR* and *MTRR*
Conclusion: The antiepileptic drug induced changes in global DNA methylation were observed in HEK293 and the modulation of global DNA methylation status in cell line can be attributed to variability in underlying *DNMT* and *TET* gene expression.

Session Title: Pharmacogenomics Poster Session II

PB2184 Pharmacogenetic study identifies IL6R as a genetic risk factor for imatinib-induced severe skin rash in gastrointestinal stromal tumor

Authors:

X. Liu, Y. Xia, S. Chen, X. Zhang; First Affiliated Hosp., Sun Yat-sen Univ., Guangzhou, China

Skin rash is one of the most common adverse drug reactions (ADRs) in gastrointestinal stromal tumor (GIST) patients treated with imatinib. Severe skin rash can reduce medication adherence, affect treatment efficacy, reduce quality of life and even lead to death. However, little is known about the mechanism of imatinib-induced skin rash. To investigate genetic polymorphisms associated with imatinib-induced skin rash in a Chinese population, a total of 475 patients (≥18 years) treated with imatinib for unresectable or metastatic GIST were recruited in the study. We conducted a pharmacogenetic investigation for imatinib-induced severe skin rash (31 ADR cases and 444 controls). Genetic loci in IL6R were strongly associated with imatinib-induced severe skin rash. The strongest finding (*P*=0.039, OR=0.533, 95% CI=0.292-0.970) identified as conferring ADR risk was with rs4129267, located in intron 8 of IL6R and is in linkage disequilibrium with the IL6R coding SNP rs2228145 that modifies the IL-6R peptide structure and shown to affect IL-6 serum levels. The findings could contribute to a better understanding of imatinib-induced severe skin rash and shed light on targets for new therapeutics.

Session Title: Pharmacogenomics Poster Session III

PB2185 Pharmacogenetic variability of tuberculosis biomarkers in native and mestizo peruvian populations.

Authors:

L. Jaramillo Valverde[1,2,3], K. Levano[4,2], D. Tarazona[2], S. Capristano[2], C. Sanchez[2], J. Poterico[3], E. Tarazona-Santos[5,6], H. Guio[2,3]; [1]Sch. of Med., Univ. Continental, Lima, Peru, [2]Laboratorio de Biotecnología y Biología Molecular, Inst. Natl. de Salud, Lima, Peru, [3]INBIOMEDIC Res. and Technological Ctr., Lima, Peru, [4]Sci. Dept., Helene Fuld Coll. of Nursing, New York, NY, [5]Departamento de Biologia Geral, Univ.e Federal de Minas Gerais, Belo Horizonte, Brazil, [6]RIBEF Ibero-American Network of Pharmacogenetics and Pharmacogenomics, Belo Horizonte, Brazil

BACKGROUND: In Peru, 29292 people were diagnosed with tuberculosis in 2022. Although TB treatments are effective, 3.4-13% are associated with significant adverse drug reactions, with drug-induced liver injury (DILI) considered the most predominant. Among the first-line antituberculosis drugs, isoniazid (INH) is the main drug responsible for the appearance of DILI. In liver, INH is metabolized by N-acetyltransferase-2 (NAT2) and cytochrome P450 2E1 (CYP2E1). Information regarding genetic risk factors associated with the presence of DILI to anti-tuberculosis drugs is limited in Latin American and even more so in native and mestizo Peruvian population. The objective of the study was to determine the prevalence of NAT2 and CYP2E1 genotypes in native and mestizo population.METHODS: An analytical cross-sectional analysis was performed using genetic data from mestizo patients with pulmonary tuberculosis in Lima and data from native participants Quechua, Ashaninka, Shimaa and Aymara of the EPIGEN - Brazil project. NAT2 acetylator genotype was determined as fast, intermediate and slow, and CYP2E1 genotypes were classified as c1/c1, c1/c2 and c2/c2, from molecular tests and bioinformatic analyses. RESULTS: Of the 472 participants, 36 haplotypes were identified in the mestizo population and 6 haplotypes in native population paired with NAT2. In mestizo population, the most frequent NAT2*5B and NAT2*7B haplotypes were associated with DILI risk; while in natives, NAT2*5G and NAT2*13A haplotypes were the most frequent and associated with decreased risk of DILI. In addition, the SNPs with the highest frequency was 282T in both populations. Differences were observed in the genotypes and allele frequencies of SNPs NAT2 282C>T, 481C>T, 590 G>A and SNP CYP2E1 1053 C>T between the two populations ($p<0.001$). In univariate analyzes for NAT2, non-wild genotypes of mestizo population reported a higher prevalence than wild genotypes in SNPs 282C>T, 341T>C, 481C>T, 590G>A ($p< 0.001$), while, non-wild genotypes of native population reported a higher prevalence than wild genotypes in SNPs, 341T>C, 481C>T, 590G>A and 830A>G ($p<0.001$). For CYP2E1, c1/c1 and c1/c2 allele are the most frequent in native and mestizo population, respectively. In both, non-wild genotypes reported a higher prevalence than wild-type genotypes in SNP 1053C>T ($p<0.001$). CONCLUSION: Despite the limitations of a secondary study, it was possible to report associations between NAT2 and CYP2E alleles with Peruvian native and mestizo population by prevalence ratios. The results of this study will help the development of new therapeutic strategies for a Tuberculosis efficient control between populations.

Session Title: Pharmacogenomics Poster Session I

PB2186 Pharmacogenetics of cardiotoxicity in the treatment of black Zimbabwean breast cancer patients on doxorubicin

Authors:

V. Nyangwara[1], M. Tinashe[1], Z. Chikwambi[1], M. Collen[1], T. Campbell[2], M. Borok[3], N. Ndlovu[3]; [1]AiBST, Harare, Zimbabwe, [2]Univ. of Colorado Anschutz Med. Campus, Aurora, Colorado, CO, [3]Univ. of Zimbabwe, Harare, Zimbabwe

Introduction: Doxorubicin induced cardiotoxicity (DIC) is a common nonmalignant treatment-related cause of death in cancer patients. Doxorubicin is well established as an efficacious drug for the treatment of a variety of cancers. It is administered as a single agent or more commonly in combination for breast cancer patients. This study was conducted to establish the frequency of DIC in Zimbabwean breast cancer patients on doxorubicin and to test the DIC predictive power of genetic biomarkers. **Methods:** All the 50 participants were women aged ≥ 18 years with breast cancer who were on doxorubicin. Non-black women and those with prior chest wall radiotherapy were excluded. To determine DIC, cardiologists performed echocardiographic examinations and echocardiographic recordings. The patients were evaluated for cardiovascular disease at entry (baseline), 3, 6 and 12 months during the study. Additional information, such as BMI, cumulative dose of doxorubicin, co-medications and physical performance of the patients were recorded using a questionnaire. To determine the DIC predictive power, patients were genotyped for UGTA1A6*4, SLC28A3 and RARG variants. To evaluate the impact of the clinical and genetic factors on left ventricular ejection fraction (LVEF) and cardiac failure, a uni- and multivariable logistic regression analysis was carried out. **Results:** Out of the 50 patients who were recruited into the study only 35 (54%) completed the study. The median age was 48 years (IQR 44.5 - 59.0). Twenty-eight (80%) had advanced stage breast cancer (stage III/IV) with Karnofsky performance status between 80-90%. The prevalence of HIV, diabetes and hypertension were 3%, 11% and 46% respectively. The median LVEF before treatment was 64.7%. The median cumulative doxorubicin dose was 238.89 mg/m2. Five, 14% of the patients experienced acute DCI associated with LVEF reduction of ≥ 10% from baseline (<60%). The three patients who developed cardiotoxicity were hypertensive. The final LVEF measurements of the five patients after the 12-month follow up ranged from 18.2% to 39%. The frequencies of SLC28A3 (rs7853758), UGT1A6*4 (rs17863783) and RARG (rs2229774) were 60.7%, 17.9% and 14.3%, respectively. No association between DIC and the three variants was observed. **Conclusions:** This is the first study on the prevalence of DIC and associated genetic biomarker predictive evaluation in Zimbabwean breast cancer patients. There is need for incorporation of both clinical and genetic determinants to refine toxicity risk assessment of DIC. A larger sample size with a longer follow-up time will be necessary in future studies.

Session Title: Pharmacogenomics Poster Session II

PB2187 Pharmacogenetics of Tenofovir Diphosphate and Lamivudine Triphosphate Concentrations and Hepatitis B Virus Suppression During Antiretroviral Therapy in Patients with HIV/HBV Coinfection

Authors:

J. Bae[1], M. Tantawy[1], Y. Gong[1], T. Langaee[1], M. Lartey[2], V. Ganu[2], K. Tachi[2], O. Ojewale[1], A. Obo-Akwa[3], I. Boamah[3], J. Kiser[4], L. Bushman[4], L. Ellison[4], H. Yang[5], P. L. Anderson[4], A. Kwara[1]; [1]Univ. of Florida, Gainesville, FL, [2]Univ. of Ghana Med. Sch., Korle-Bu Teaching Hosp., Accra, Ghana, [3]Univ. of Ghana Med. Sch., Accra, Ghana, [4]Univ. of Colorado-Anschutz Med. Campus, Aurora, CO, [5]Univ. of Rochester Sch. of Med. and Dentistry, Rochester, NY

Background: In Africa, up to 25% of persons living with human immunodeficiency virus (HIV) are coinfected with hepatitis B virus (HBV). Suboptimal suppression of HBV DNA during antiretroviral therapy (ART) in HIV/HBV-coinfected patients is common, ranging from 13.5% to 65%. We hypothesize that variations in genes that encode transporters and phosphodiesterase enzymes will be associated with nucleoside analogs intracellular pharmacokinetics (PK) and incomplete HBV treatment response. **Methods:** In this study, 138 HIV/HBV-coinfected patients on tenofovir (TDF)/lamivudine (3TC)-containing ART at the Korle-Bu Teaching Hospital in Ghana were enrolled. Selected single nucleotide polymorphisms (SNPs) in ABCB1, ABCC2, ABCC4, ABCC10, PDE1C, PDE3A, PDE4D, PDE11A, and SLC28A2 with potential effect on nucleoside PK were genotyped using TaqMan assays. Relationships between SNPs and steady-state intracellular concentrations of the active phosphate anabolites of TFV and 3TC called TFV diphosphate (TFVdp) and 3TC triphosphate (3TCtp), respectively, were analyzed by using multivariable linear regression adjusted for age, gender, BMI, and CD4 count. Concentrations of TFVdp and 3TCtp in peripheral blood mononuclear cells (PBMCs) were used as the reference tissue for in-vivo nucleoside reverse transcriptase studies, while that in red blood cells (dried blood spots (DBS)) were used as a measure for long-term adherence of TDF/3TC. **Results:** Of the 136 patients (mean age 45 years; 63% females) on TDF/3TC-containing ART for a median (IQR) of 6 (4.08 to 7.5) years, 9 (6.6%) had unsuppressed HBV DNA, 10 (7.4%) had unsuppressed HIV RNA, and 6 (4.4%) had suppressed HIV RNA but unsuppressed HBV DNA. The missense rs2273697 variant allele A in the ATP-binding cassette sub-family C member 2 (ABCC2) was associated with higher intracellular 3TCtp concentration in PBMC, with mean concentrations of 10.06, 10.98, and 19.31 pmol/10^6 cells in GG, AG, and AA, respectively (P=0.015). In unadjusted analysis, PDE1C rs30561 (P=0.038) and ABCC2 rs3740066 (P=0.003) were associated with TFVdp in DBS, while PDE4D rs6889641 (P=0.028) and ABCC2 rs17216177 (P=0.049) were associated with TFVdp and 3TCtp in PBMCs, respectively. None of the evaluated SNPs were associated with unsuppressed HBV DNA. **Conclusion:** This exploratory study identified a significant association between ABCC2 rs2273697 and intracellular 3TC concentration in HIV/HBV-affected African patients. Further studies with larger sample size are warranted to confirm our findings as well as elucidate the underlying molecular mechanism of ABCC2 rs2273697 and its contribution to the elimination of 3TC.

Session Title: Pharmacogenomics Poster Session III

PB2188 Pharmacogenomic Landscape of TNF Inhibitors in the Qatari Population

Authors:

Z. Jan, D. Velayutham, B. Mifsud, P. Jithesh; Coll. of Hlth.& Life Sci., Hamad Bin Khalifa Univ., Doha, Qatar

Autoimmune diseases are increasing globally, causing mortality and morbidity worldwide. Over the past few decades, TNF inhibitors that target tumor necrosis factor (TNF-α) have been widely used for the treatment of autoimmune diseases. However, 40% of the patients do not respond to TNF inhibitors. Several pharmacogenetic variants affecting TNF inhibitor response have been identified. However, the prevalence of these variants in the Qatari population is still not known. In this study, we identified the distribution of known and novel variants in 102 genes associated with response to TNF inhibitors in 14,392 Qataris from whole genome sequencing data and calculated the cumulative risk probability score. Out of the 119 known pharmacogenomic variants associated with response to TNF inhibitors, ~90% had different allele frequency distribution from other world populations present in the gnomAD dataset. High frequency of rs1143634 (*IL1B)*and rs1800896 (*IL10)* variants was observed, that are known to be associated with both negative and positive responses to infliximab and etanercept respectively. Moreover, we identified that the allele frequency of rs1295686 (*TH2LCRR*) associated with better response to Adalimumab is higher in the Qatari population. Furthermore, higher distribution of AA genotype on *HLA-E* (rs1264457) gene was observed which showed a better response to all TNF inhibitors. Additionally, we identified a novel missense deleterious variant in *TNF* (p. Pro88Ser) that increases the binding of *TNF* with *TNFR1*, which suggests the potential reduced response of TNF inhibitors in such patients. The findings of this study could be the basis for developing and implementing pharmacogenetic testing in autoimmune disease patients being considered for TNF inhibitor therapy in Qatar and beyond.

Session Title: Pharmacogenomics Poster Session I

PB2189 Pharmacogenomics Applied to Chronic Pain Treatment in Primary Care (PGx-ACT) trial; A Largely Virtual Randomized Trial

Authors:

D. Smith[1,2], R. Beyene[1], P. Kolm[1], T. A. Young[1], S. Zifa[1], V. Natividad[1], A. Licata[1], T. Moore[3], R. Walsh[1], D. Shikha[1], A. D. Walker[1], A. Turabi[1], M. B. Jacobs[1], B. N. Peshkin[4], S. M. Swain[1,2,4]; [1]MedStar Hlth., Columbia, MD, [2]Georgetown Univ. Med. Ctr., Washington, DC, [3]Kailos Genetics, Huntsville, AL, [4]Georgetown Univ., Washington, DC

Background *CYP2D6* variation is associated with reduced bioactivation of tramadol, codeine, and hydrocodone. The primary objective was to identify effects of providing pharmacogenomic (PGx) recommendations for patients with chronic pain treated in primary care clinics compared to standard care (SC). **Methods** An open-label, prospective, largely virtual trial randomized participants to a PGx-guided care (PC) or SC arm. Adults with chronic pain (≥ 3 months) treated with either tramadol, codeine, or hydrocodone enrolled from 20 primary care practice sites 1/2021 - 12/2022. Targeted next-generation sequencing was used for *CYP2D6*. CYP2D6 phenotype was assigned per genotype and drug interactions. PGx-trained pharmacists sent recommendations to provider in electronic health record (EHR). Alternative analgesic therapy was recommended for CYP2D6 intermediate or poor metabolizers (IM/PMs) for risk of ineffectiveness. Prescribing decisions were at provider discretion. Data were collected at baseline and 3 months. Analysis was in CYP2D6 IM/PMs. Primary outcome was change in pain intensity (PI) T-score in PC vs. SC arms. Secondary outcomes were number with PGx-aligned care, change in morphine milliequivalents (MME) prescribed, number with ≥ 30% improvement in PI. Analyses used t-test, Wilcoxon, Fisher's exact test. **Results** Eligibility was assessed for 4573 patients. 315 patients enrolled, 253 randomized, and 214 completed the trial (80% virtual). 104 (49%) were IM/PMs with 49 (PC) and 55 (SC). No difference was found in PI (-1.1 ± 5.6 vs. -1.5 ± 7.4; P = 0.74), ≥ 30% improvement (16% vs. 18%; P = 1), MME prescribed (-1.7 ± 13 vs. -1.1 ± 13; P = 0.82), or PGx-aligned care (69% vs. 65%; P = 0.68) between PC and SC arms, respectively. Post hoc analyses reassessed outcomes regardless of study arm. PI was -1.8 ± 6.4 vs. -0.24 ± 6.8 (P = 0.085) between those with (n=70) vs. without (n=34) PGx-aligned care. Patients with an analgesic change, with (n= 31) vs. without (n=34) PGx-aligned care had a reduced MME prescribed (-8.5 ± 16 vs. 2.8 ± 13; P < 0.001) and marginal change in PI (-2.7 ± 7.1 vs. -0.24 ± 6.8; P = 0.061). PC arm, 16% of provider notes mentioned PGx. Lessons learned 1) Since providers did not order test it may impact use of PGx 2) Providers saw participants weeks-months after PGx availability, which reduced awareness of PGx in EHR. **Conclusion** This PGx-guided strategy for patients with chronic pain prescribed select opioids was not associated with different prescribing decisions or clinical outcomes. Providers did not apply PGx for several potential reasons. Favorable post hoc analyses suggest future efforts should identify effective methods to implement and optimize PGx-aligned care.

Session Title: Pharmacogenomics Poster Session II

PB2190 Pharmacogenomics in the pediatric intensive care unit: an exome sequencing study

Authors:

N. Lynch[1], N. T. Elefant[2], J. Beckett[1], A. Geneslaw[1], V. Jobanputra[3], Z. Ren[1], D. S. Hughes[1], A. Revah-Politi[1], J. E. Motelow[2]; [1]Columbia Univ. Med. Ctr., New York, NY, [2]Columbia Univ., New York, NY, [3]New York Genome Ctr., New York, NY

Pharmacogenomics (PGx) refers to the effect of the patient's genetic architecture on medication response. Limited data in pediatric populations such as oncology and cardiology patients indicate that PGx drives changes in management by optimizing therapeutic drug levels and avoiding adverse drug events (ADEs). ADEs are more common in the pediatric intensive care unit (PICU) than the general pediatric ward, yet no prior studies have investigated the impact of PGx in the PICU. Genetic variants may govern medications commonly used in the PICU such as warfarin, fosphenytoin or proton pump inhibitors. While exome sequencing (ES) is not optimal to identify PGx variants, it is a common diagnostic tool in the PICU and a limited scope of PGx variants and phenotypes can still be identified. The objective of our study was to assess the overlap of clinically actionable metabolizer phenotypes with the administration of medication with PGx guidelines recommending modified dosing. We performed a retrospective analysis of ES from 45 children from a single PICU (the NewYork-Presbyterian/Columbia University Irving Medical Center). Exomes were aligned to human reference GRCh38 using DRAGEN and variants were called using Genome Analysis Toolkit (GATK). Star alleles and PGx phenotypes were generated using Pharmacogenomics Clinical Annotation Tool (PharmCAT). Medication administration data was collected from the electronic medical record. We found that 23 individuals received 8 unique medications for which we could identify a metabolizer phenotype from ES defined by 4 unique genes. Of these 23 individuals, seven were poor or intermediate metabolizers for medications administered to them in the PICU. Tacrolimus was the administered drug implicated in all three poor metabolizer phenotypes. In future directions, we will extend this study to 244 additional children in the PICU for a total cohort of 267 and determine whether geographic ancestry affects the identification of PGx phenotypes. To support future implementation of PGx testing for children who are at risk of frequent visits to the PICU, such as those with genetic and chronic conditions, we must lay the groundwork with data describing the expected yield and utility of PGx findings in this critical yet understudied population. Demonstrating the overlap between PGx variants and medications received in the PICU will establish preliminary data for clinical trials testing whether PGx-guided dosing improves efficacy while avoiding ADEs for children with critical illness.

Session Title: Pharmacogenomics Poster Session III

PB2191 Pharmacogenomics: Targeted Sequencing, HPRC Genome Annotation and Clinical Utilization

Authors:

Y-S. Liu[1], K-Y. Li[2], I-H. Chiu[3], C-Y. Chen[2], L-J. Shen[4], C-F. Huang[5], P-L. Chen[1,3]; [1]Graduate Inst. of Med. Genomics and Proteomics, Natl. Taiwan Univ., Taipei, Taiwan, [2]Dept. of Biomechatronics Engineering, Natl. Taiwan Univ., Taipei, Taiwan, [3]Dept. of Med. Genetics, Natl. Taiwan Univ. Hosp., Taipei, Taiwan, [4]Graduate Inst. of Clinical Pharmacy, Sch. of Pharmacy, Natl. Taiwan Univ., Taipei, Taiwan, [5]Dept. of Pharmacy, Natl. Taiwan Univ. Hosp., Taipei, Taiwan

Pharmacogenomics is a field that investigates the impact of genetic variations on drug response. Its significance lies in optimizing drug selection and dosage. Despite its potential, the complexity of pharmacogenes has limited its widespread application. This study aims to annotate the genotypes of 24 pharmacogenes to Human Pangenome Reference Consortium (HPRC) assemblies, which comprise 47 phased, diploid genomes from a genetically diverse cohort. The HPRC phased genome assemblies were aligned to the GRCh38 reference genome using Minimap2. To facilitate detection and recognition of single nucleotide variants and indels, the alignment file in PAF format was converted to the variant call format (VCF) using paftools.js. Additionally, we designed a target enrichment next-generation sequencing (NGS) panel comprising 60 important pharmacogenes selected from Clinical Pharmacogenetics Implementation Consortium (CPIC), Pharmacogene Variation (PharmVar) Consortium, and Pharmacogenomics Knowledge Base (PharmGKB). The HPRC DNA samples were sequenced using this panel, and the resulting sequencing data were compared with the HPRC VCF files. By comparing the sequencing results with the HPRC VCF files, we demonstrated the accuracy of our laboratory methodology for clinical use. Furthermore, approximately 300 participants from the National Taiwan University Hospital (NTUH) were enrolled in the trial, enabling the construction of a clinical decision support system (CDSS) for improved utilization of pharmacogenomics in clinical settings. Integrating genetic information into clinical decision-making could enhance patient care, optimize drug selection and dosage, and improve treatment outcomes. Through the comprehensive profiling of pharmacogenes and the development of a CDSS, this study offers a promising approach for advancing the integration of precision medicine into routine clinical practice.

Session Title: Pharmacogenomics Poster Session I

PB2192 Predicting Precision Inhaled Corticosteroids Response in Asthma Patients

Authors:

N. Tahmin[1], A. Mitoubsi[1], B. Bakir Batu[2], A. Mohammed[2], L. Chinthala[2], A. Carlisle[2], T. Glauser[3], I. Krantz[4], K. Mandl[5], R. Davis[2], A. Wu[5], A. Khojandi[1], and members of the Genomic Information Commons Consortium; [1]Univ. Of Tennessee, Knoxville, Knoxville, TN, [2]Univ. of Tennessee Hlth.Sci. Ctr., Memphis, TN, [3]Cincinnati Children's Hosp. Med. Ctr., Cincinnati, OH, [4]Children's Hosp. of Philadelphia, Philadelphia, PA, [5]Harvard Med. Sch., Boston, MA

Although most patients with asthma respond to inhaled corticosteroids (ICS), approximately 1/3 respond minimally or not at all. Evidence for an important role of genetics is emerging. However, current asthma risk assessment techniques are not fully capable of producing holistic and reliable risk scores that can guide precision treatment planning. Our study aims to develop machine learning models to predict patient-specific genome-based ICS response in asthma patients.

We leveraged the data collected through the Biorepository and Integrative Genomics (BIG) Initiative, University of Tennessee Health Science Center. We used whole exome sequence (WES) data from 290 patients (age: mean 16 years (IQR: 14-18), 52.4% Black, 47.6% non-Hispanic white) aged >6 years old, and a history of chronic ICS use. Cases were defined as patients with emergency department visit or hospitalization due to asthma within the 24-month follow-up window after the first ICS prescription. Out of all patients, there were 145 (50%) cases. We developed Random Forest (RF) models to examine the extent to which WES data can predict ICS response in patients. Data preprocessing included using MAF $\geq$ 5%, PLINK for linkage disequilibrium pruning (window size=50kb, step size=5, r^2=0.2), SNP variant ordinal encoding, and K-nearest neighbors (KNN) imputation (K=5). We performed two analyses: in analysis I, we used all remaining SNPs after feature selection using least absolute shrinkage and selection operator (LASSO); in analysis II, we used only SNPs from 17 previously identified genes affecting ICS response. Finally, models were evaluated using 10-fold cross-validation, and the mean and standard deviation of the area under the receiver operating characteristic curves (AUCs) were calculated.

Analyses I and II resulted in AUCs of 0.926±0.0861 and 0.904±0.0942 (paired t-test p-value 0.0887), respectively. Analysis I using all SNPs produced higher recall and precision. In addition, analysis I (using all SNPs) was more accurate for Black patients (11.7% misclassification for analysis I, using all SNPs, versus 22.4% misclassification for analysis II, using SNPs from previously identified genes, respectively).

Our results show that RF can accurately stratify ICS non-responders from responders, paving the way for producing patient-specific risk scores that can guide precision therapies. Furthermore, our results indicate that SNPs from genes not previously reported to affect ICS response could be relevant to improving predictions for Black patients. Future work includes external evaluation of the RF models and identifying the most important SNPs affecting ICS response.

Session Title: Pharmacogenomics Poster Session II

PB2193 Proteome-wide Mendelian randomization and conventional observational analyses reveal increased Alzheimer's disease risk associated with angiotensin-converting enzyme inhibitor use

Authors:

P. Huang[1], H. Tang[2], Z. Feng[2], E. Stergiakouli[1], T. Gaunt[1], J. Zheng[3,1], L. Yao[2,4]; [1]Univ. of Bristol, Bristol, United Kingdom, [2]Central South Univ., Changsha, China, [3]Shanghai Jiao Tong Univ., Shanghai, China, [4]King's Coll. London, London, United Kingdom

Introduction: Integrating evidence from real-world drug use data can enhance the translational value of findings from omics-based approaches to improve clinical practice. We aimed to identify plasma protein targets to inform drug discovery, repurposing, and vigilance for preventing Alzheimer's disease (AD). **Methods:** We first performed proteome-wide Mendelian randomization (MR) using protein quantitative trait locus (pQTL) data from Zheng et al. and genome-wide association data for AD from Bellenguez et al., followed by multi-omic validation, to explore potential druggable targets related to AD risk. For the top finding, we used multivariable MR (MVMR) to address possible pleiotropy and incorporated observational analyses of UK Biobank data with confounder-adjusted Cox proportional hazards models to further support the finding. **Results:** Of 593 proteins with single-*cis*, 124 proteins with multiple-*cis*, and 73 proteins with *cis* + *trans* pQTL instruments, MR identified 15 proteins (GRN, ACE, CD55, TREM2, TMEM106B, LILRB1, SIRPA, SIGLEC9, IDUA, LILRB2, TCN2, CTSH, GPC5, LILRA4 and IL6ST) with putative causal effects on AD after Bonferroni correction. After being validated by summary-data-based MR (SMR) with expression (eQTL), splicing (sQTL), and methylation quantitative trait locus (mQTL) data in blood and brain, angiotensin-converting enzyme (ACE) emerged as the most reliable target, exhibiting a protective effect against AD (OR: 0.91 [95% CI: 0.88-0.93]), and its eQTL, sQTL, mQTL SMR results had p-values < 0.05, passed the heterogeneity in dependent instruments (HEIDI) test ($P_{HEIDI} < 0.05$), and had plausible effect directions (gene expression had same effect directions and methylation in promoter areas had opposite effect directions with the protein). Despite its high relevance to blood pressure (BP) regulation, MVMR suggested the effect of ACE on AD was independent of BP. Observational evidence from UK Biobank supported our MR finding for ACE by showing a consistently increased incidence of AD among ACE inhibitor users, compared with: a) angiotensin II receptor blocker users (HR: 1.37 [95% CI: 1.02-1.84]), b) all other antihypertensive users (1.29 [1.08-1.54]), c) all participants without antihypertensive use (1.48 [1.28-1.72]), d) participants with hypertension but without antihypertensive use (1.26 [1.07-1.48]), and e) all participants without ACEI use (1.41 [1.22-1.63]). **Conclusions:** Our findings expand the understanding of potential targets for AD prevention and highlight the need for further investigation of AD risk associated with the use of ACE inhibitors, a commonly prescribed first-line antihypertensive in clinical practice.

Session Title: Pharmacogenomics Poster Session III

PB2194 RAB8A as a novel regulator of MED16 and its effect on 5'-deoxy-5-fluorouridine response.

Authors:

J. Campbell, C. Murtagh, A. Stark; Univ. of Notre Dame, Notre Dame, IN

Breast cancer is a leading cause of death among women in the United States and traditional treatment regimens commonly utilize trial and error methodologies. Individualized treatment regimens can optimize drug response in cancer patients. However, these plans depend upon a clear understanding of how genetics influences protein expression and drug response. Our objective was to identify functional protein candidates for optimizing capecitabine therapy. 5'-deoxy-5-fluorouridine is a metabolite of capecitabine, an oral prodrug and antimetabolite used for breast cancer treatment, and is the activated form of the drug used for in vitro studies. Three SNPS localized to the short arm of chromosome 19 in RAB8A that were associated with MED16 protein expression ($p = 0.00002$) and 5'-deoxy-5-fluorouridine response ($p < 0.0005$) in lymphoblastoid cells were initially identified. Importantly, MED16 expression was also correlated with 5'-deoxy-5-fluorouridine response in lymphoblastoid cells ($p = 0.01$). RAB8A is a member of the RAS superfamily, highly conserved, and implicated in tumorigenesis. Past studies have revealed that MED16 is a coactivator involved in the transcriptional regulation of many RNA-polymerase II dependent genes and plays a role in Vitamin D reception, which may affect calcium homeostasis, cell proliferation, and cell differentiation. Thus, expression of MED16 may significantly alter transcription levels of RNA-polymerase II dependent genes and affect functions such as cell proliferation which are abnormal in cancerous cell cycles. No current literature links MED16 and RAB8A. We proposed a pathway through which MED16 might operate utilizing the three SNPs identified due to their significant correlation with MED16 and location with RAB8A. We then knocked down RAB8A via siRNA in MCF7 breast cancer cells utilizing Lipofectamine as a transfection reagent and assessed RAB8A and MED16 expression levels. Utilizing qPCR analysis, our results displayed significant knockdown of RAB8A and MED16 and therefore support our hypothesized connection of RAB8A and MED16. Further evaluation will be done on down stream targets and the potential implication to capecitabine treatment.

Session Title: Pharmacogenomics Poster Session I

PB2195 Racial differences in *UGT1A1* allele frequencies and its potential impact in pharmacogenetic testing for cancer chemotherapy drugs.

Authors:

K. choi, G. Selkov, R. Huether, B. Rhead, F. De La Vega; Tempus Lab Inc., Chicago, IL

UGT1A1 pharmacogenetic (PGx) tests evaluate the risk of adverse events associated with the chemotherapy drugs irinotecan and belinostat by genotyping the promoter TA-repeat polymorphisms (*28, *36, and *37) and two exonic SNVs (*6 and *27). The prevalence of these variants differs significantly by race. In particular, the frequencies of *36 and *37 are 5% and 7% in African-descent populations, respectively, whereas in other populations they are less than 0.2% (cf. gnomAD database). *37 is known to reduce UGT1A1 enzyme activity to a greater extent than *28; however, it is not included in the current FDA drug labels. Given the prevalence of *UGT1A1* TA-repeat alleles, compound heterozygotes carrying a TA-repeat variant and either *6 or *27 are not uncommon. We analyzed racial differences in the prevalence of *UGT1A1* variants and the potential for misassignment of metabolizer phenotype in the absence of phasing. Normal tissues from 14,580 de-identified cancer patients were sequenced for *UGT1A1* using the Tempus xT 648-gene panel NGS test. TA repeat length was determined using a bespoke Bayesian repeat calling algorithm. Four race and ethnicity categories — Non-Hispanic (NH) Asian, Hispanic or Latino, NH Black, and NH White — were imputed from continental genetic ancestry derived from 654 ancestry informative markers, as described previously. We identified 121 unique exonic variants and four promoter TA-repeat polymorphisms ([TA]5-8) in *UGT1A1*. *In silico* analysis predicted 68 of the exonic variants as potentially deleterious. Notably, 17.6% of NH Black patients were carriers of either *36 or *37. Ninety-eight compound heterozygotes of potential clinical significance were identified. However, without phasing information, accurate assignment of *UGT1A1* metabolizer phenotype in compound heterozygotes is challenging. This is particularly relevant in the NH Asian population, where 6% of the population in the cohort are compound heterozygotes, a rate 6.4-fold higher than the overall cohort prevalence. Our research highlights racial differences in *UGT1A1* allele frequencies, particularly among NH Black and NH Asian populations, which could influence test results and their interpretation in some patients. Further investigation in diverse populations is crucial to ensure equality in pharmacogenetic testing related to cancer drug side effects.

Session Title: Pharmacogenomics Poster Session II

PB2196 Response to anti-IL17 therapy in inflammatory disease is not strongly impacted by genetic background

Authors:

L. Jostins-Dean[1], C. Zhang[2], K. Shestopaloff[1], B. Hollis[1], C. Kwok[1], C. Hon[3], N. Hartmann[4], C. Tian[2], M. Wozniak[5], L. Santos[6], D. West[1], S. Gardiner[1], A-M. Mallon[6], A. Readie[7], R. Martin[7], T. Nichols[1], M. Beste[3], J. Zierer[8], E. Ferrero[8], M. Vandemeulebroecke[4]; [1]Univ. of Oxford, Oxford, United Kingdom, [2]China Novartis Inst.s for Bio-Med. Res., Shanghai, China, [3]Novartis Inst.s for BioMed. Res., Cambridge, MA, [4]Novartis Pharma AG, Basel, Switzerland, [5]Novartis Ireland Limited, Dublin, Ireland, [6]The Alan Turing Inst., London, United Kingdom, [7]Novartis Pharmaceuticals Corp., East Hanover, NJ, [8]Novartis Inst.s for BioMed. Res., Basel, Switzerland

Response to the anti-IL17 monoclonal antibody secukinumab is heterogeneous, and not all patients respond to treatment. Understanding whether this heterogeneity is driven by genetic variation is a key aim of pharmacogenetics, and could influence precision medicine approaches in inflammatory diseases. Using changes in disease activity scores across 5218 genotyped individuals from 19 clinical trials across four indications (psoriatic arthritis, psoriasis, ankylosing spondylitis and rheumatoid arthritis), we tested whether genetics predicted response to secukinumab. We did not find any evidence of association between treatment response and common variants, imputed HLA alleles, polygenic risk scores of disease susceptibility, or cross-disease components of shared genetic risk. This suggests that anti-IL17 therapy is equally effective regardless of the genetic background of the patient, which has important implications for future genetic studies of biological therapy response in inflammatory diseases.

Session Title: Pharmacogenomics Poster Session III

PB2197 Role of Pharmacogenomics in Oncology

Authors:

C. Lee, J. Leow, S. Chew; Natl. Cancer Ctr. Singapore, Singapore, Singapore

Introduction: The uptake of pre-emptive genotype testing in clinical practice has been relatively low in Singapore. There is increasing interest in the implementation of pharmacogenomics as a mainstream service in National Cancer Centre Singapore (NCCS). The gene-drug pair of CYP 2D6-tamoxifen is regarded as the most relevant. This study aims to determine the acceptance of patients on tamoxifen for genotype testing as well as the facilitators and barriers for acceptance, and to make recommendations for the establishment of a pharmacogenomics service.
Methods: This was a cross-sectional survey of patients collecting tamoxifen from 1st September 2022 to 28th February 2023 at NCCS. Patients were evaluated on their acceptance of pre-emptive genotype testing using a questionnaire derived from the Theory of Planned Behaviour. The patients' demographics, social history, knowledge of and prior experience with genetic testing were recorded.
Results: A total of 202 patients responded, all of whom were female. Seventy percent of patients responded that they intend to undergo pre-emptive genotype testing for CYP 2D6 if offered, 76.1% felt that pre-emptive testing for CYP 2D6 should be offered up front to all patients taking tamoxifen while 42% indicated they were willing to pay for testing. High overall scores for attitude, subjective norms and perceived behavioural control were independent predictors of acceptance. Potential facilitators for acceptance include adequate data security, strong family/ social support, convenient testing procedures, 3rd-party payer support, and adequate patient education. Lack of physician support, high cost of testing and lack of information and guidance from healthcare providers would discourage patients from testing.
Conclusion: There is a high level of acceptance among patients in this study for pre-emptive genotype testing. Further research is required to demonstrate the cost-effectiveness of testing. Training on pharmacogenomics should be provided to healthcare professionals to increase competency which would help in the establishment of pharmacogenomics services.

Session Title: Pharmacogenomics Poster Session I

PB2198 † Sex-biased gene expression and gene-regulatory networks of sex-biased adverse event drug targets and drug metabolism genes

Authors:

B. Lasseigne, J. Fisher, A. Clark, E. Jones; The Univ. of Alabama at Birmingham - SOM, Birmingham, AL

Previous pharmacovigilance studies and a retroactive review of cancer clinical trial studies identified that women were more likely to experience adverse drug events (i.e., any unintended effects of medication), and men were more likely to experience adverse events that resulted in hospitalization or death. These sex-biased adverse events (SBAEs) are due to many factors not entirely understood, including differences in body mass, hormones, pharmacokinetics, and liver drug metabolism enzymes and transporters. We first identified 416 drugs associated with SBAEs from the FDA Adverse Event Reporting System (FAERS) database. Next, we evaluated sex-specific gene expression of the known drug targets and metabolism enzymes for those SBAE-associated drugs and found 32 known drug metabolism enzymes and 84 drug targets enriched. We also constructed sex-specific tissue gene-regulatory networks to determine if these known drug targets and metabolism enzymes from the SBAE-associated drugs had sex-specific gene-regulatory network properties and predicted regulatory relationships. We identified liver-specific gene-regulatory differences for drug metabolism genes between males and females (e.g. immediate node neighbors), which could potentially explain observed sex differences in pharmacokinetics and pharmacodynamics. In addition, we found that ~85% of SBAE-associated drug targets had sex-biased gene expression or were core genes of sex- and tissue-specific network communities, significantly higher than randomly selected drug targets. Also, the number of SBAE-associated drug targets that were core genes of sex- and tissue-specific network communities were significantly higher than randomly selected drug targets. Lastly, we provide the sex-biased drug-adverse event pairs, drug targets, and drug metabolism enzymes as a resource for the research community. Overall, we provide evidence that many SBAEs are associated with drug targets and drug metabolism genes that are differentially expressed and regulated between males and females. These SBAE-associated drug metabolism enzymes and drug targets may be useful for future studies seeking to explain or predict SBAEs.

Session Title: Pharmacogenomics Poster Session II

PB2199 Targeting and comprehensive characterization of pharmacogenes using Oxford Nanopore Technologies' adaptive sampling

Authors:

S. Hickey[1], **C. Tyer**[1], L. Ly[1], J. Beaulaurier[1], S. Juul[2]; [1]Oxford Nanopore Technologies, Inc, Alameda, CA, [2]Oxford Nanopore Technologies, Inc, New York, NY

Variants in pharmacogenes influence pharmaceutical metabolism by changing the efficacy of uptake, signal transduction, and breakdown of certain molecules. Here, we demonstrate enrichment and comprehensive characterization of a pharmacogenomics (PGx) gene panel including 278 unique pharmacogenetic targets sourced from the FDA, PharmGKB, and the Clinical Pharmacogenetics Implementation Consortium (CPIC). The panel was enriched and sequenced using Oxford Nanopore Technologies adaptive sampling (AS), a software-based approach to enrich regions of interest by uploading a file of genomic coordinates. During sequencing, DNA strands are basecalled and mapped to a reference genome in real time. Strands that align outside of the target regions are ejected from the nanopore before sequencing is complete, while strands that are on-target are allowed to sequence completely.Patient-derived cell lines with known allelic variants were barcoded and sequenced on PromethION flow cells. AS enabled all PGx genes from the panel to be enriched to sufficient coverage for variant calling, with no drop-out for any target genes. Samples were analyzed using the wf-human-variation Epi2Me Labs pipeline from Oxford Nanopore Technologies, which called small variants with Clair3, structural variants with Sniffles2, and phased reads with whatshap. In particular, challenging structural variants in *CYP2D6*, including tandem *CYP2D6-CYP2D7* hybrids, were able to be haplotyped and fully resolved.Star (*) allele genotypes were called using PharmCAT. Calls were 89% concordant with Genetic Testing Reference Materials Coordination Program (GeT-RM) consensus calls; discrepancies were observed for certain genes due to changing definitions in the GeT-RM reference database. Additionally, using native nanopore sequencing CpG methylation sites were called using dorado, revealing haplotype-specific methylation at genes such as *VKORC1*. While the biological significance of the methylation status of PGx genes has received comparably less attention than genetic variants, this is an interesting area for future research.Overall, we demonstrate targeted nanopore sequencing with AS as a useful tool for the enrichment and phasing of PGx gene panels, resulting in high coverage for haplotyped variant calling including structural variations and methylation. The strategy allows for comprehensive PGx analysis that can be easily modified to encompass additional genes of interest as novel pharmacogenomic variants are discovered.

Session Title: Pharmacogenomics Poster Session III

PB2200 The combined effect of rare and common variants in *GSDMB/ORMDL3* is associated with response to inhaled corticosteroids among asthmatic children

Authors:

A. Wu[1], K. Voorhies[2], M. McGeachie[1], A. Mohammed[3], L. Chinthala[3], K. Mandl[1], B. Raby[1], R. Davis[4], S. Lutz[1]; [1]Harvard Med. Sch., Boston, MA, [2]Harvard Pilgrim Hlth.Care Inst., Boston, MA, [3]Univ. of Tennessee Hlth.Sci. Ctr., Memphis, TN, [4]Univ. of Tennessee Hlth.Sci. Ctr., Memphis, TN

Rationale: Inhaled corticosteroids (ICS) are efficacious in the prevention of asthma which affects more than 300 million people in the world. While genome-wide association studies identified genes involved in differential treatment responses to ICS in asthma, few studies have evaluated the potential effects of combined rare and common variants and the contribution of rare vs common variants. **Methods:** Among children with asthma on ICS with whole exome sequencing (WES) in the PrecisionLink Biobank for Health Discovery at Boston Children's Hospital, we examined if the combined effect of rare and common variants are associated with ICS response as defined by hospitalization or emergency department visits. We examined rare and common variants associated with ICS response for 12 regions previously associated with asthma and ICS response (*DPP10, FBXL7, NDFIP1, TBXT, GLCCI1, HDAC9, TBXAS1, STAT6, GSDMB/ ORMDL3, CRHR1, GNGT2, FCER2*). We used the sequence kernel association test (SKAT) to examine the combined effect of rare and common variants on ICS response using the combined sum test. We adjusted for age, sex, and BMI and stratified by race for 91 White and 20 Black children. We tested for replication in the Biorepository and Integrative Genomics (BIG) Initiative, University of Tennessee Health Science Center among 83 White and 134 Black children with asthma, ICS use, and WES data. **Results:** Using a Bonferroni threshold of 0.05/12=0.004, *GSDMB/ ORMDL3* was significantly associated with ICS response for the combined effect of both rare and common variants (p-value=0.003 for both rare and common variants, p-value=0.008 for only rare variants, p-value=0.03 for only common variants) among White children in the PrecisionLink biobank. This signal replicated in the BIG Initiative for the combined effect of rare and common variants (p-value=0.02 for both rare and common variants, p-value=0.30 for only rare variants, p-value=0.03 for only common variants). **Conclusions:** Using WES data, the combined effect of rare and common variants for *GSDMB/ORMDL3* was associated with ICS response among asthmatic children in the PrecisionLink biobank with replication in the BIG Initiative. Future studies examining the combined effect of both rare and common variants may help elucidate mechanisms underlying ICS response variability. This proof-of-concept study demonstrates the potential power of biobanks of pediatric real-life populations in genomic investigations.

Session Title: Pharmacogenomics Poster Session I

PB2201 The RNA-editing ADAR proteins broadly regulate gene expression in hepatic cells

Authors:

J. Collins, F. Yu, D. Wang; Univ. of Florida, Gainesville, FL

The adenosine deaminase acting on RNA (ADAR and ADARB1) proteins are RNA-editing enzymes that broadly regulate gene expression through editing-dependent or editing-independent mechanisms. Using siRNA-mediated gene knockdown (KD) in HepaRG cells, we found broad changes in the expression of drug-metabolizing enzymes and related transcription factors in the KDs. We used whole transcriptome and miRNA sequencing to further understand the regulatory networks of ADAR and ADARB1 on gene expression. We detected global changes to gene expression after KD of either ADAR or ADARB1, resulting in differentially expressed (DE) transcripts for both ADAR (7028) and ADARB1 (3121), including many genes related to drug metabolism and disposition. Gene ontology (GO) analysis found that pathways related to viral defense and inflammation were enriched in the ADAR KD, in agreement with previous reports in other tissues and cell types. In contrast, after ADARB1 KD, cellular homeostasis and metabolism pathways became enriched, indicating that ADAR and ADARB1 have non-overlapping targets. We also identified numerous alterations to transcript splicing isoforms and sequence edits in both KDs. Among these, several key pharmacogenes were altered (e.g., CYP3A4 and CYP3A5), supporting previous work that the ADARs regulate drug metabolism. In addition, we observed broad changes in the expression of long noncoding RNAs (ADAR KD - 7854 DE, ADARB1 KD- 3208 DE) and miRNAs (ADAR KD - 143 DE, ADARB1 KD - 122 DE). Our results showed the broad regulatory roles of ADAR and ADARB1 on the expression of mRNA, lncRNA, and miRNA in the liver, highlighting the different targets of the ADARs. These results will help elucidate the pathways and cellular mechanisms leading to changes in the expression of the pharmacogenes. Since dysregulation of the ADARs has been directly implicated in the progression of many cancers, including hepatocellular carcinoma, the results will further our understanding of cancer progression and regulation of drug metabolism in the liver.

Session Title: Pharmacogenomics Poster Session II

PB2202 Treatment of MeCP2e1 deficient mice with microglia ablating PLX3397 correlated with sex disparate on body weight but not Rett like disease severity

Authors:

D. Yasui[1], H. M. Sophia[1], K. E. Neier[1], E. G. Thrall[1], R. Kawai[1], T. A. Tran[1], C. G. Torres[1], A. L. Humphries[1], A. M. Madany[1], P. Ashwood[2], J. Lasalle[1]; [1]UC Davis Sch. of Med., Davis, CA, [2]UC Davis Mind Inst., Davis, CA

Rett syndrome (RTT) is a complex neurodevelopmental disorder that occurs in 1 in 10,000 births. Approximately 95% of RTT patients are females heterozygous for X chromosome linked *MECP2* mutations and are thus mosaic for wild-type and mutant *MECP2* expression at a cellular level. *MECP2* encodes MeCP2e1 which is necessary for brain development and MeCP2e2 which is necessary for placental development. Analysis of a construct relevant, MeCP2e1 deficient mouse model suggests that microglia contribute to RTT like gait defects, as well as gut microbiome and metabolic changes. Thus, we hypothesized that pharmacologic ablation of microglia would ameliorate RTT like phenotypes in MeCP2e1 deficient mice. Therefore, PLX3397 a CSF1R/c-kit inhibitor dissolved in AIN-76A chow at 290 PPM was orally administered to *Mecp2e1*$^{-/+}$, *Mecp2e1*$^{-/y}$ and *Mecp2e1*$^{+/+}$ and *Mecp2e1*$^{+/y}$ mice, either acutely (2 weeks) or chronically (from 4-23 weeks) to ablate microglia in brain, compared to vehicle control. *Mecp2e1*$^{-/+}$, *Mecp2e1*$^{-/y}$, and *Mecp2e1*$^{+/+}$ and *Mecp2e1*$^{+/y}$ littermates were weighed and scored for neurophenotypes for 23 weeks post-weaning. Gait, fecal cytokines, fecal short chain fatty acids and gut microbial composition was assessed in all mice before, during and after disease onset. Surprisingly, neither acute or chronic PLX3397 treatment significantly reduced neurophenotype score either in *Mecp2e1*$^{-/+}$ females or *Mecp2e1*$^{-/y}$ male mice. Instead, chronic PLX3397 treatment significantly elevated disease severity in *Mecp2e1*$^{-/+}$ females compared to vehicle *Mecp2e1*$^{-/+}$ controls. However, chronic PLX3397 treatment significantly reduced weight gain over time in *Mecp2e1*$^{-/y}$ and *Mecp2e1*$^{+/y}$ males compared to vehicle male controls, but not *Mecp2e1*$^{-/+}$ females compared to untreated *Mecp2e1*$^{-/+}$ controls. Interestingly, the use of purified Envigo AIN-76A chow reduced disease severity overall in *Mecp2e1*$^{-/+}$ female and *Mecp2e1*$^{-/y}$ male mice compared to a previous study where *Mecp2e1*$^{-/+}$, *Mecp2e1*$^{-/y}$ and *Mecp2e1*$^{+/+}$ and *Mecp2e1*$^{+/y}$ mice were fed PicoLab mouse diet, possibly due to differences in lipid composition. Neither gait stride length nor organ weight were significantly altered by acute or chronic PLX3397 treatment in both *Mecp2e1*$^{-/+}$ females or *Mecp2e1*$^{-/y}$ males. As a caveat, PLX3397 treatment affected c-kit signaling as shown by coat color changes and may have produced significant liver pathology in certain mice. Although targeting of microglia with PLX3397 did not reduce disease severity in *Mecp2e1*$^{-/+}$ female or *Mecp2e1*$^{-/y}$ Rett model mice, the results of this study together with our prior investigations of lipid deficiencies suggest a potential dietary intervention for Rett syndrome.

Session Title: Pharmacogenomics Poster Session III

PB2203 Using genome-wide gene-drug interactions to test for genomic sensitivity to statins in the treatment of hypercholesterolemia

Authors:

B. Verhulst; Texas A&M Univ., Bryan, TX

Hypercholesterolemia is a major risk factor for cardiovascular disease, which is the leading cause of death world-wide. While prescription medications (chiefly statins) are undeniably beneficial, the magnitude of their benefit may depend on an individual's genotype. Our premise is that genomic variants interact with medications resulting in heterogeneous LDL cholesterol levels. To clarify the etiological factors the culminate in heterogeneous cholesterol levels, we use a genomic moderation approach that is more sensitive to gene-drug interactions than existing methods. Two factors have hindered research on gene-drug interplay: 1) existing methods struggle to identify robust and replicable interactions; and 2) genome-wide studies have low statistical power. To overcome these obstacles, 1) we developed genomic methods that are more sensitive to genetic interactions, and 2) we use an extremely large genetically-informed dataset. Our general hypothesis is that statins may be more effective for individuals with specific genotypes. After conducting gene-statin interaction analyses for LDL cholesterol, we collate the results with pharmaceutical databases to identify corresponding significant interactions and drug targets. Our results: 1) highlight specific genotypes that increase the effectiveness of statins; 2) provide a novel test of whether pharmacologically predicted gene targets correspond with observed differential genetic associations; and 3) identify genomic pathways for future drug refinement or repurposing.

Session Title: Pharmacogenomics Poster Session I

PB2204 Variants reaching Genomic wide significance in their association with therapeutic response to Atenolol or Topiramate in migraine prophylaxis are identified.

Authors:

F. Cutrer, L. Wang, C. J. Klein; Mayo Clinic, Rochester, MN

At present, the prophylactic treatment of migraine is given to patients without knowledge of the relevant mechanism of action and with no biologically based rationale for drug selection. To address this, we performed Whole Genome sequencing using the blood DNA samples from 370 patients treated for three months with monotherapy of either atenolol (n=179) or topiramate (n=191) for migraine prophylaxis. We assessed the treatment response using the percentage of headache days reduced. Association with treatment response reached genome-wide significance in 374 variants for atenolol treatment and 1,031 variants for topiramate using PLINK with a general linear model. Among the atenolol-associated variants, 77% were substitutions, 23% were insertions/deletions while in the topiramate-associated variants, 75% were substitutions and 25% were insertions/deletions. 22% of atenolol and 26% of topiramate treatment-associated SNPs were novel with no rs ID. No shared variants were detected between the atenolol and topiramate groups. None of the variants associated with atenolol or topiramate response were reported by prior migraine GWAS studies. Non-synonymous coding SNPs (chr1:161365641:C/T:1) in *CFAP126* for atenolol; (chr7:100057239: T/G:1) in *ZSCAN21* and (chr12:122976756: G/A:1) in *OGFOD2* for topiramate treatment response were identified (p<5E-8). Top pathway using SNPs associated with either atenolol or topiramate monotherapy treatment (p<0.006) include: **Atenolol:** Regulation of commissural axon pathfinding by SLIT and ROBO p=0.000891 Genes involved *ROBO2, SLIT3;* Citric acid cycle p=0.004416, *SDHC, SUCLG2* **Topiramate:** Non-integrin membrane-ECM interactions p=0.000632 Genes involved *COL4A2, COL4A6, ITGAV, PRKCA, TNC*; Synthesis of PIPs at Golgi membrane p=0.001146 *PIK3C2G, PIK3R4, TPTE*; ROBO receptors bind AKAP5 p=0.0047 *PRKCA, ROBO2*; Laminin interactions p=0.005146 *COL4A2, COL4A6, ITGAV*; Regulation of commissural axon pathfinding by SLIT and ROBO p=0.00583 *ROBO2, SLIT3* It is interesting to note that pathways involving *SLIT* and *ROBO2* are implicated in response to both Atenolol and Topiramate. Pathways involved Synthesis of PIPs at the Golgi membrane have been highly associated with Topiramate and with verapamil in prior similar investigations of migraine prophylaxis (Cutrer et al. 2022). We propose that response to prophylactic treatment is an element of phenotype linked to a discoverable underlying relevant mechanism of action in migraine headache suppression.

Session Title: Pharmacogenomics Poster Session II

PB2205 Variation in the response to synthetic glucocorticoids is dominantly explained by a single measure of efficacy.

Authors:

N. Giroux, G. Johnson, A. Barrera, T. Reddy; Duke Univ., Durham, NC

Synthetic glucocorticoids (GCs) induce the activity of the glucocorticoid receptor (GR) at thousands of sites across the genome to inhibit inflammation and suppress autoimmunity. The development of selective GR modulators (SGRMs) would see the dissociation of these therapeutic GC effects from concomitant adverse metabolic effects. We measured the gene expression and regulatory element activity responses to ten SGRMs, including full and partial GR agonists as well as antagonists, using mRNA-seq and STARR-seq, a genome-wide high-throughput reporter assay. Differential expression was induced by SGRMs at 8,072 genes, and differential activity was induced at 49,711 regulatory elements. We further reported the efficacies, defined as the number of differentially expressed genes or the number of regulatory elements with differential activity, measured for each SGRM compared to a vehicle control. We identified highly correlated linear relationships between responses to each SGRM and dexamethasone, a strong GR agonist used as a positive control for this study, contrary to claims that these synthetic GCs selectively modulate GR activity. Using these relationships, we estimated the change in gene expression and regulatory element activity with respect to dexamethasone, defined as the coefficient of a linear regression model and referred to as the relative response. In summary, we observed that the relative responses induced by agonists were consistently larger than those of non-agonists for both gene expression (agonists: 0.39-0.96; non-agonists: 0-0.14) and regulatory element activity (agonists: 0.37-0.87; non-agonists: 0.08-0.20). Furthermore, the variances in the relative responses were largely explained by the efficacies of the SGRMs alone, with 87% of the variance in relative gene expression responses and 93% of the variance in relative regulatory element activity responses accounted for. Though we found that two antagonistic SGRMs - CORT108297 and RU486 - produced a higher proportion of repressive effects compared to dexamethasone, overall, we demonstrated limited evidence of differential anti-inflammatory or metabolic gene set enrichment. Thus, we concluded that this functional assessment of selective modulation of GR activity instead identified conserved molecular effects which were tuned by the efficacy of the synthetic GC.

Session Title: Pharmacogenomics Poster Session III

PB2206 Whole-genome sequencing pharmacogenomic variants in a census-based Brazilian cohort of admixed elderly individuals

Authors:

M. Naslavsky[1,2,3], M. Scliar[4], G. Debortoli[5], B. Thiruvahindrapuram[6], S. W. Scherer[6,7], Y. A. O. Duarte[8], M. Zatz[9,10], G. Suarez-Kurtz[11], E. J. Parra[12], L. B. Nasciben[2,13,14]; [1]Dept. of Genetics and Evolutionary Biology, BioSci.s Inst., Univ. of São Paulo, Sao Paulo, Brazil, [2]Human Genome and Stem Cell Res. Ctr., Univ. of São Paulo, Sao Paulo, Brazil, [3]Hosp. Israelita Albert Einstein, São Paulo, Brazil, [4]Human Genome and Stem Cell Res. Ctr., Univ. of São Paulo,, São Paulo, Brazil, [5]Dept. of Anthropology, Univ. of Toronto at Mississauga, Mississauga, ON, Canada, [6]The Ctr. for Applied Genomics, The Hosp. for Sick Children, Toronto, ON, Canada, [7]Dept. of Molecular Genetics, Faculty of Med., Univ. of Toronto, Toronto, ON, Canada, [8]Med.-Surgical Nursing Dept., Sch. of Nursing, Univ. of São Paulo, São Paulo, Brazil, [9]Human Genome and Stem Cell Res. Ctr., Univ. of São Paulo, São Paulo, Brazil, [10]Dept. of Genetics and Evolutionary Biology, BioSci.s Inst., Univ. of São Paulo, São Paulo, Brazil, [11]Divisão de Pesquisa Clínica e Desenvolvimento Tecnológico, Inst. Natl. de Câncer, Rio De Janeiro, Brazil, [12]Dept. of Anthropology, Univ. of Toronto at Mississauga, Toronto, ON, Canada, [13]Sch. of Publ. Hlth., Univ. of São Paulo, São Paulo, Brazil, [14]John P. Hussman Inst. for Human Genomics, Univ. of Miami Miller Sch. of Med., Miami, FL

Pharmacogenomics (PGx) studies aim the identification of genetic variants that modulate response to drugs. The distribution of PGx variants varies across populations, and whole-genome sequencing (WGS) is a comprehensive approach to detect both common and rare variants. We studied the frequency of PGx markers in a WGS dataset from the “Health, Well-Being, and Aging Study” (SABE), a census-based admixed cohort from Sao Paulo, Brazil, which includes variants from WGS of 1,171 unrelated, elderly individuals. Stargazer tool was used to call star alleles and structural variants (SVs) across 38 pharmacogenes. Clinically relevant variants classified by PharmGKB and CPIC were investigated, and the predicted drug response phenotype was analyzed in combination with the medication record to assess individuals potentially at high-risk of gene-drug interaction. 352 unique star alleles or haplotypes were observed, of which 255 and 199 had a frequency < 0.05 and < 0.01, respectively. For star alleles with frequency > 5% (n=97), decreased, loss-of-function and unknown function accounted for 13.4%, 8.2% and 27.8% of alleles or haplotypes, respectively. Structural variants (SVs) were identified in 35 genes for at least one individual, and occurred with frequencies >5% for *CYP2D6*, *CYP2A6*, *GSTM1*, and *UGT2B17*. More than 98% of the individuals carried at least one high risk genotype-predicted phenotype in pharmacogenes with PharmGKB level of evidence 1A for drug interaction. The Electronic Health Record (EHR) Priority Result Notation and the cohort medication registry were combined to assess high-risk gene-drug interactions. Overall, 42.0% of the cohort used at least one PharmGKB evidence level 1A drug, and 18.9% of individuals who used PharmGKB evidence level 1A drugs had a genotype-predicted phenotype of high-risk gene-drug interaction. We described the applicability of next-generation sequencing (NGS) techniques for translating PGx variants into clinically relevant phenotypes on a large scale in the Brazilian population and explores the feasibility of systematic adoption of PGx testing in Brazil.

Session Title: Epigenetics Poster Session I

PB2207 † 3D features of the genome impact escape from X inactivation.

Authors:

X. Deng[1], H. Fang[1], A. R. Tronco[1], G. Bonora[2], G. N. Filippova[1], C. Disteche[1,3]; [1]Dept. of Lab. Med. and Pathology, Univ. of Washington, Seattle, WA, [2]Dept. of Genome Sci., Univ. of Washington, Seattle, WA, [3]Dept. of Med., Univ. of Washington, Seattle, WA

The number and escape levels of genes that escape X chromosome inactivation (XCI) in female somatic cells vary among tissues and cell types, potentially contributing to specific sex differences. Here we investigate the role of CTCF, a master chromatin conformation regulator, in regulating escape from XCI. CTCF binding profiles and epigenetic features were systematically examined at constitutive and facultative escape genes using mouse allelic systems to distinguish the inactive X (Xi) and active X (Xa) chromosomes. We found that escape genes are located inside domains flanked by convergent arrays of CTCF binding sites, consistent with the formation of loops. In addition, strong and divergent CTCF binding sites often located at the boundaries between escape genes and adjacent neighbors subject to XCI would help insulate domains. Facultative escapees show clear differences in CTCF binding dependent on their XCI status in specific cell types/tissues. Concordantly, deletion but not inversion of a CTCF binding site at the boundary between the facultative escape gene *Car5b* and its silent neighbor *Siah1b* resulted in loss of *Car5b* escape. Reduced CTCF binding and enrichment of a repressive mark over *Car5b* in cells with a boundary deletion indicated loss of looping and insulation. In mutant lines in which either the Xi-specific compact structure or its H3K27me3 enrichment was disrupted, escape genes showed an increase in gene expression and associated active marks, supporting the roles of the 3D Xi structure and heterochromatic marks in constraining levels of escape. Our findings indicate that escape from XCI is modulated both by looping and insulation of chromatin via convergent arrays of CTCF binding sites and by compaction and epigenetic features of the surrounding heterochromatin.

Session Title: Epigenetics Poster Session II

PB2208 A cistrome-wide association study identifies novel risk loci for renal cell carcinoma.

Authors:

Z. Zhang[1], D. Osei-Hwedieh,[1,2], J. Anderson[2], S. Abou Alaiwi[1,3], A. Nassar[4], J-H. Seo[1], E. Adib[1], T. El Zarif[1], T. Choueiri[1], M. Freedman[1], S. Baca[1]; [1]Dana-Farber Cancer Inst., Boston, MA, [2]Harvard Med. Sch., Boston, MA, [3]Brigham and Women's Hosp., Boston, MA, [4]Yale Cancer Ctr., New Haven, CT

Renal Cell Carcinoma (RCC), the most common type of kidney cancer, has been extensively studied through genome-wide association studies (GWAS) to identify genetic risk loci. Most RCC risk variants map to gene regulatory elements, but how they confer risk of RCC remains largely unknown. To discover RCC risk variants that confer risk through effects on regulatory element activity, we performed a cistrome-wide association study (CWAS). This analysis integrated summary statistics from a large-scale RCC GWAS (10,784 RCC cases and 20,406 controls) and novel genetic models of regulatory element activity. We leveraged ChIP-seq profiles for H3K27ac - a histone modification associated with active regulatory elements - from 64 RCC samples. We identified 13,534 regulatory elements that were influenced by genetic variation, as determined by correlation of SNP genotype with H3K27ac peak intensity or by allelic imbalance of heterozygous SNPs in ChIP-seq reads. We identified 12 regulatory elements that were significantly associated with RCC risk. Seven mapped to known RCC risk loci such as 2*p*21, 3*q*22.1, 11*q*13.3, 12*p*12.1, and 14*q*24.2, and provided mechanistic hypotheses to explain how genetic variation confers risk at these sites by modifying activity of specific regulatory elements. Five additional RCC-associated peaks were identified in regions that did not harbor any genome-wide significant SNPs within 2Mb. By utilizing chromatin conformational data and eQTL profiles, a regulatory element on 7*q*31.33 capturing a significant GWAS variant with limited functional annotation was linked to *POT1*, a protein involved in the regulation of telomerase-mediated telomere extension. In summary, our study identified known risk loci and nominated novel risk regions contributing to the development of RCC. We highlight the increased power of CWAS to nominate risk loci and investigate mechanisms of risk conferred by genetic variation in regulatory DNA. Mechanistic insights gained from these studies may advance the biological understanding, risk stratification, and therapeutic targeting of RCC.

Session Title: Epigenetics Poster Session III

PB2209 A classifier for lifetime cannabis use based on DNA methylation.

Authors:

L. Zhou[1], B. C. Quach[1], K. G. Lawrence[2], J. van Dongen[3], J. A. Marks[1], Z. Xu[2], L. J. Bierut[4], M. Ollikainen[5], J. A. Taylor[2], J. T. Bell[6], J. Kaprio[5], D. I. Boomsma[7], K. Xu[8], D. P. Sandler[2], D. B. Hancock[1], E. O. Johnson[1], F. Fang[1]; [1]RTI Intl., Research Triangle Park, NC, [2]NIH/NIEHS, Research Triangle Park, NC, [3]VU Amsterdam, Amsterdam, Netherlands, [4]Washington Univ Sch Med, St Louis, MO, [5]Univ Helsinki, Helsinki, Finland, [6]King S Coll. London, London, United Kingdom, [7]VU Univ. Amsterdam, Amsterdam, Netherlands, [8]Yale Sch. Med., New Haven, CT

Cannabis use is highly prevalent, and its effects on health vary with potential adverse outcomes and therapeutic benefits depending on the context. However, studying cannabis-related health outcomes can be difficult due to the lack of reliable biomarkers to accurately assess cannabis use patterns. One promising biomarker is DNA methylation (DNAm), as cannabis use may result in long-lasting changes to the methylome. In this study, we utilized peripheral blood DNAm data to develop a classifier predicting lifetime cannabis use (ever vs. never).

Analyzing data from the prospective Sister Study, we identified 108 significant CpGs ($p<1E-4$) associated with lifetime cannabis use among 2,073 participants of European ancestry (1,009 ever users). Employing LASSO regression with 10-fold cross-validation on centered and scaled DNAm beta values, we selected 77 CpGs to build a model predicting lifetime cannabis use. Testing this model on an independent sample of 517 Sister Study participants (243 ever-users), the classifier achieved an AUC (area under curve) of 0.67 (95% CI 0.62~0.71). We validated the classifier's performance in an independent sample of 1,195 participants (665 ever users) in the Gulf Long-Term Follow-Up Study (GuLF). The classifier was predictive of lifetime cannabis use in both European (AUC=0.62, 95% CI 0.58~0.66) and African (AUC=0.62, 95% CI 0.57~0.66) ancestry groups. In an independent sample of 2,109 participants (551 ever users) from the Netherlands Twin Register (NTR), the classifier achieved an AUC of 0.65 (95% CI 0.63~0.68).

To consider confounding by cigarette smoking, we further evaluated the 77 CpG-classifier in subsets of participants who never smoked cigarettes in each dataset. In the validation subset of 217 Sister Study participants, the classifier achieved an AUC of 0.66 (95% CI 0.59~0.72). In GuLF, the classifier performed better among participants of African ancestry (n=343, AUC=0.6, 95% CI 0.54~0.67) than those of European ancestry (n=360, AUC=0.54, 95% CI 0.47~0.61). In NTR (n=1,253), the classifier showed modest predictive ability, with an AUC of 0.58 (95% CI 0.54~0.62).

Our DNAm-based classifier predicts lifetime cannabis use, achieving AUC values around 0.65 in independent datasets. These findings support the potential for developing a peripheral blood-based DNAm biomarker to identify lifetime cannabis use across various population groups, enabling further research into the diverse health effects associated with cannabis use.

Session Title: Epigenetics Poster Session I

PB2210 A comparison of Whole-Genome Bisulfite Sequencing and Oxford Nanopore Technology for epigenetic studies.

Authors:

D. Tay, H. Ng, M. Lam, N. Sadhu, M. Loh, J. Chambers; Nanyang Technological Univ., Singapore, Singapore

Epigenetic studies provide insights into gene regulation and disease mechanisms. This knowledge is much coveted for personalized disease diagnosis, prevention, and therapy. Two widely used techniques for methylation detection are whole-genome bisulfite sequencing (WGBS) and long reads sequencing by Oxford Nanopore Technology (ONT). WGBS is based on short-read sequencing and requires bisulfite conversion which could degrade the DNA samples. On the contrary, long-read sequencing by ONT allows direct methylation detection, thereby eradicating the need for bisulfite conversion. Despite their difference in sequencing principles, a comprehensive assessment of the two approaches has not been performed to date. Such comparison is critical to enable informed experimental decision to be made prior to the inception of epigenetic studies. Thus, this study aims to assess i) the concordance of methylation levels of the 2 technologies in relation to the array, ii) number of CpGs detected with different number of aggregated ONT flow cells (FCs), and iii) relationship between the number of aggregated ONT FCs and WGBS. A total of 16 and 214 samples were sequenced using ONT and WGBS technologies. The WGBS data have a mean coverage of 32.3x (SD=2.0) while different numbers of ONT FCs (1 to 4) were used to sequence the samples. Methylation levels detected by WGBS and ONT were first assessed for concordance with methylation levels by Illumina methylation arrays. Performance was comparable ($p=0.0836$), with average Pearson correlation of 0.957 (SD=0.004) and 0.958 (SD=0.010) achieved for WGBS and ONT respectively. Average root mean squared error (RMSE) was lower in WGBS at 0.133 (SD=0.006), relative to 0.154 for ONT (SD=0.016). Next, the impact of the number of ONT FCs used on the methylation profile was evaluated. With 2 FCs, the average number of CpGs detected increased by 47.0 folds relative to the use of 1 FC. When 3 and 4 FCs were used, the increase in the average number of CpGs detected was at 1.54 and 1.09 folds (in relation to n-1 aggregated FCs) respectively. This suggests that there is little incremental value in using more than 2 FCs if the desired minimum coverage is at 30x. Lastly, comparison of WGBS and ONT demonstrates that with a single ONT FC, the number of CpGs detected at different minimum coverage (1x to 100x) is comparable to WGBS ($p=0.246$). However, with increased number of FCs, the CpGs' minimum coverage outperforms those from WGBS ($p<0.05$). This suggests that with at least 2 ONT FCs, we can achieve results that outperform WGBS, and that the use of ONT sequencing could potentially offer better insights into the epigenetic profile.

Session Title: Epigenetics Poster Session II

PB2211 A comprehensive integrated post-GWAS analysis of reproductive lifespan reveals enhancer-based NBR2 dysregulation in ovary

Authors:

S. Kim[1], J. Chang[2], Y. Suh[1]; [1]Columbia Univ. Irving Med. Ctr., New York, NY, [2]Univ. of Maryland, Baltimore, Baltimore, MD

The ovary exhibits the earliest aging among human body parts. The reproductive lifespan (RL), a period between the onset of menarche and menopause, greatly varies from woman to woman, ranging from 20 to 50 years. The length of the RL is partially genetically determined, but the underlying genetic mechanism is not yet understood. Using UK Biobank data, we calculated RL values from the age at menarche (AMC) and natural menopause (ANM) and performed genome-wide association tests using mixed model linear regression on the RL, AMC, and ANM of 50,698 women. We identified 7,364 variants that were significantly associated ($P < 5x10-8$) with 76 loci across all three traits. Through colocalization analysis, we observed that the majority of RL variants overlapped with ANM variants, indicating that the ANM variants are the primary genetic determinants of RL. To identify potential causal variants, we applied an integrative post-GWAS analysis with public epigenomic data and prioritized 55 high-probability causal variants in 22 loci, including 12 ovary eQTLs in the NBR2 locus (17q21.31). We measured and observed allele-specific changes in the long non-coding RNA NBR2, a known regulator of the AMPK-mTOR pathway. Decreased expression of NBR2 may contribute to a longer RL and later ANM, potentially by altering the activity of the AMPK-mTOR pathway. These findings provide new insights into the biological mechanisms underlying female RL and suggest potential targets for interventions to improve women's reproductive health.

Session Title: Epigenetics Poster Session III

PB2212 A computational toolkit to integrate multi-omics time-series data across species in brain development

Authors:

M. Yu[1], B. Borsari[1], E. S. Wattenberg[1], K. Xu[2], M. Frank[1], S. Liu[1], M. Gerstein[1]; [1]Yale Univ., New Haven, CT, [2]Yale Sch. Med., New Haven, CT

Steady-state molecular studies provide a static description of molecular processes in cell differentiation or organismal development. Longitudinal studies can instead offer a deeper understanding of epigenetic and transcriptional events and how these processes evolve over time. Here, we integrate multi-omics data (ATAC-seq, DNase-seq, ChIP-seq, RNA-seq) from different consortia to understand brain development at equivalent post-conception dates in human and mouse (7-20 post-conception weeks -pcw-, and 10.5-16.5 post-conception days, respectively).Genes with key roles in human brain development are associated with enhancers showing many different dynamic patterns, while genes performing non-brain functions are more often associated with just one type of enhancers (either up- or down-regulated). This suggests that proper expression of key human brain genes may require tight, multi-gate epigenetic control which is not seen in mice. To further understand differences in brain development between human and mouse, we trained a random-forest classifier that predicts the species-specificity of enhancer-based time-series data. We found both activation and repression of enhancers happen earlier in mice than humans. This is in line with previous studies comparing molecular processes in these species. The 10-12 pcw window is key to distinguishing enhancer activation patterns between the two species. At this time, 26% of mouse enhancers have already undergone major activation and 45% have undergone major repression, while human enhancers are mostly activated or repressed between 12-13 pcw.To further uncover the potential emergence of distinct neuronal lineages during development, we developed a multivariate HMM-based method to estimate brain cell composition at each time-point by integrating bulk and single-cell ATAC-seq data. Among all the deconvolved cell types, we observed that the most predominant excitatory neuron subgroup initially emerges at around 13 pcw and displays a marked enrichment at 16 pcw. Therefore, this subgroup might represent a distinct neuronal type that differentiates and matures during brain development. This observation is in line with a prior study, where this specific neuronal group has been shown to match to the newly formed migrating excitatory neurons from developing cortex of pcw 17-18. Taken together, our study reveals highly dynamic changes in epigenetic patterns and cell type compositions that underlie neurodevelopmental processes. Our future goal is to develop a computational toolkit to best harmonize and analyze time-series data generated across species, consortia and technologies.

Session Title: Epigenetics Poster Session I

PB2213 A DNA methylation atlas of normal human cell types

Authors:

N. Loyfer[1], J. Magenheim[1], A. Peretz[1], R. Shemer[1], B. Glaser[2], Y. Dor[1], T. Kaplan[1]; [1]Hebrew Univ. of Jerusalem, Jerusalem, Israel, [2]Hadassah Hebrew Univ Jerusale, Jerusalem, Israel

DNA methylation is a fundamental epigenetic mark that governs gene expression and chromatin organization, thus providing a window into cellular identity and developmental processes. Current datasets typically include only a fraction of methylation sites and are often based either on cell lines that underwent massive changes in culture or on tissues containing unspecified mixtures of cells. Here we describe a human methylome atlas, based on deep whole-genome bisulfite sequencing, allowing fragment-level analysis across thousands of unique markers for 39 cell types sorted from 205 healthy tissue samples. Replicates of the same cell type are more than 99.5% identical, demonstrating the robustness of cell identity programmes to environmental perturbation. Unsupervised clustering of the atlas recapitulates key elements of tissue ontogeny and identifies methylation patterns retained since embryonic development. Loci uniquely unmethylated in an individual cell type often reside in transcriptional enhancers and contain DNA binding sites for tissue-specific transcriptional regulators. Uniquely hypermethylated loci are rare and are enriched for CpG islands, Polycomb targets and CTCF binding sites, suggesting a new role in shaping cell-type-specific chromatin looping. The atlas provides an essential resource for study of gene regulation and disease-associated genetic variants, and a wealth of potential tissue-specific biomarkers for use in liquid biopsies.

Session Title: Epigenetics Poster Session II

PB2214 A novel gene regulatory element that modulates cardiomyocyte maturation.

Authors:

T. Anglen, K. Hagy, R. Perrelli, D. Tran, M. Ramaker, A. Landstrom, Y. Diao, S. Shah, C. Gersbach; Duke Univ., Durham, NC

Heart disease is the leading cause of death in the world and can be caused by genetic mutations. Heart disease results in pathogenic remodeling of the heart. A molecular marker of remodeling is a decreased expression of MYH6 relative to MYH7, the two cardiac isoforms of the sarcomeric myosin heavy-chain protein that is important for heart contractions. The silencing of the lncRNA MHRT is central to this switch from MYH6 to MYH7 in mice. Remarkably, transgenic expression of MHRT in mice attenuates cardiac hypertrophy. Dissecting how MYH6 and MYH7 levels are regulated in humans will lead to mechanistic understanding of development and disease progression and inform the development of therapeutic interventions. To this end, we have performed epigenome and transcriptome profiling and high-throughput CRISPR-based epigenetic screening to decode gene regulatory mechanisms in human iPSC-derived cardiomyocytes (iPSC-CMs). We identified a novel regulatory element involved in controlling MYH6 expression in iPSC-CMs in an MHRT-independent manner. We show that by repressing or activating this regulatory element with the CRISPR-based epigenetic editors we can modulate MYH6 expression. We then use the HiCAR chromatin conformation assay to show that repression of the regulatory element disrupts the regional 3D genomic architecture and decreases how frequently the element and the promoter of MYH6 interact. Finally, we observe that the repression of this element during iPSC to CM differentiation results in phenotypic changes of the resulting CMs. Following repression of this regulatory element, the iPSC-CMs have an increased tau of decay of calcium transients linked to an increase in CM hypertrophy and a decrease in calcium sparks, linked to cardiomyocyte maturity. The iPSC-CMs also have decreased expression of developmental factors NPPA and NPPB. A similar decrease in levels of NPPA and NPPB is observed during cardiac maturation after birth. These results implicate this genomic region as an important regulatory element during cardiac development and maturation as we observed similar phenotypic changes in the iPSC-CMs that occur during *in vivo* cardiac maturation when we manipulate the epigenetic state of this regulatory element in iPSC-CMs. In conclusion, we find that this regulatory element is involved in controlling MYH6 expression and is important during cardiomyocyte differentiation and maturation. Ongoing work is investigating causal roles that this element may have in disease progression and whether human genetic variation within this element contributes to heart disease phenotypes.

Session Title: Epigenetics Poster Session III

PB2215 A trans-ancestral meta-analysis of genome-wide association study identifies 146 susceptibility loci for rheumatoid arthritis

Authors:

A. Kwon[1], S-Y. Bang[2,3], H-S. Lee[2,3], K. Kim[1], S-C. Bae[2,3]; [1]Dept. of Biology, Kyung Hee Univ., Seoul, Korea, Republic of, [2]Hanyang Univ. Inst. for Rheumatology Res., Seoul, Korea, Republic of, [3]Dept. of Rheumatology, Hanyang Univ. Hosp. for Rheumatic Diseases, Seoul, Korea, Republic of

Rheumatoid arthritis (RA) is a complex autoimmune disorder characterized by inflammation in multiple joints. The development of RA is influenced by multiple factors, including genetic factors, with approximately 65% of heritability. While genome-wide association studies (GWASs) have successfully identified specific genetic variants associated with RA, a significant portion of the heritability remains unexplained. This study aimed to enhance our understanding of the genome-scale genetic architecture of RA by conducting a trans-ancestral meta-analysis of GWAS. To this end, we expanded our previous GWAS datasets (PMID 33310728) to include a total of 4,638 patients with RA and 74,949 healthy controls. After a general quality control procedure before and after whole-genome imputation, the genetic association between variants and RA was assessed by logistic regression for each cohort followed by an inverse-variance-weighted fixed-effects meta-analysis with recent large-scale GWAS association summary statistics (PMID: 36333501; 35,072 non-Korean patients with RA and 239,398 healthy controls). A SNP-based heritability was ~21%, which was improved over the previously established heritability of 17.6% in East Asian populations in our recent study (PMID 33310728). A total of 24 novel susceptibility loci were newly identified, exhibiting the genome-wide significance level (*P*-value < 5 x 10-8) and highly consistent effect sizes between datasets. Additionally, known RA association signals at 122 loci including the HLA region were replicated. In summary, this study provides an updated catalog of RA susceptibility loci that will facilitate a deeper understanding of the underlying biology involved in the development of RA.
Funding: 2017R1E1A1A01076388, 2021R1A6A1A03038899, 2022R1A2C2006073, NBK-2021-010

Session Title: Epigenetics Poster Session I

PB2216 African Americans with Sickle Cell Disease (SCD) demonstrate accelerated epigenetic aging compared to African Americans without SCD.

Authors:

A. Ashley-Koch[1], M. E. Garrett[1], B. Le[1], K. Bourassa[2], D. Hatch[3], Q. Yang[3], P. Tanabe[3], N. Shah[1], C. Oyedeji[1], N. Kimbrel[2], J. Beckham[2], M. J. Telen[1], M. Knisely[3], MIRECC Workgroup; [1]Duke Univ. Med. Ctr., Durham, NC, [2]Mid-Atlantic Mental Illness Res. (MIRECC), Ed. and Clinical Ctr., Veteran Admin., Durham, NC, [3]Duke Univ. Sch. of Nursing, Durham, NC

Sickle cell disease (SCD) is a chronic medical condition characterized by polymerization of deoxy-hemoglobin S (HbS), sickling of red blood cells, hemolysis, end-organ damage in multiple systems and reduced survival. Early mortality is associated with SCD-related end organ damage, especially to the heart, lung, kidneys and central nervous system, as well as with vaso-occlusive events. We hypothesized that early mortality is also associated with accelerated aging and that this could be quantified by measuring DNA methylation at multiple CpG loci across the genome. Several such epigenetic clocks have been developed. One of the most recent, DunedinPACE (Belsky et al., 2022), was developed from a longitudinal cohort and estimates the rate of an individual's biological aging compared to peers. Importantly, accelerated aging predicts mortality and poor health outcomes in many populations but has not been evaluated in SCD. We compared DunedinPACE epigenetic pace of aging scores in 134 African American individuals with SCD from the Outcome Modifying Genes in SCD (OMG-SCD, n=45; Elmariah et al., 2014) and SCD Implementation Consortium (SCDIC, n=89; DiMartino et al., 2018) cohorts to 1392 African American veterans without SCD from the MIRECC cohort (Calhoun et al., 2010). All subjects had global methylation data generated from the Infinium MethylationEPIC Beadchip (Illumina, San Diego, CA). Pace of aging was calculated with the R package "DunedinPACE" (Belsky et al., 2022). We found that the individuals with SCD displayed significantly accelerated epigenetic aging compared to the veterans without SCD, whereby individuals with SCD aged approximately 0.5 months more per year than the veterans ($p=5.31e^{-5}$). This was true, even though the individuals with SCD were significantly younger according to chronologic age than the individuals without SCD (mean age_{SCD}=34.4 yrs vs. mean age_{MIRECC}=39.4 yrs; $p<0.0001$), making the epigenetic aging discrepancy even more apparent. When we matched the SCD and MIRECC subjects on age, the effect size of this association was even larger (0.76 months more per year in SCD vs. veterans; $p=7.24e^{-3}$). These data support our hypothesis that individuals with SCD experience accelerated epigenetic aging as measured by global epigenetic variation. The assessment of epigenetic aging may prove useful to identify which SCD patients would most benefit from clinical interventions to reduce mortality.

Session Title: Epigenetics Poster Session II

PB2217 Alcohol sensitivity associated genome wide DNA methylation changes in alcohol use disorder patients and healthy controls

Authors:

S. Venkatesh[1], T. Gunawan[1], S. Bethany[1], M. Schwandt[2], P-H. Shen[3], C. Hodgkinson[3], M. De Biasi[4], D. Nancy[2], D. Goldman[3], V. Ramchandani[1]; [1]Human Psychopharmacology Lab., NIAAA/NIH, Bethesda, MD, [2]The Office of the Clinical Director, NIAAA/NIH, Bethesda, MD, [3]Lab. of Neurogenetics, NIAAA/NIH, Rockville, MD, [4]Univ. of Pennsylvania, Philadelphia, PA

Heavy alcohol consumption and alcohol use disorder (AUD) significantly alters genome-wide DNA methylation patterns. A lower level of response (LR) to alcohol predicts a greater risk for heavy drinking and alcohol-associated problems. In this study, we aim to identify AUD and alcohol sensitivity-associated DNA methylation changes in AUD patients and healthy controls. We analyzed genome-wide DNA methylation patterns in WBCs obtained from whole blood samples of individuals with current AUD (179) and non-AUD controls (95), using the Illumina MethylationEPIC BeadChip (850K). Differentially methylated positions (DMPs) were analyzed to identify the hypo- and hypermethylated regions between AUD and non-AUD groups after controlling for age, sex, race, principal components, and cell compositions. Given the impact of childhood trauma on long-term mental and physical health and associated epigenetic changes, we also controlled for Childhood Trauma Questionnaire scores. The LR to alcohol was measured with the Self-Report of the Effects of alcohol (SRE) questionnaire evaluating the number of standard drinks required to feel four different effects of alcohol. To identify the alcohol sensitivity-related DMPs we included SRE recent and total scores in the linear model. Between AUD and non-AUD groups, 4430 differentially methylated CpG sites were determined including 2719 hypermethylated and 1711 hypomethylated sites across exon, 3'UTR, 5'UTR, and gene body regions. After including alcohol SRE recent and total scores, the AUD group showed 909 differentially methylated CpG sites compared to the non-AUD group, including 631 hypermethylated and 278 hypomethylated sites across the 3'UTR, 5'UTR, and gene body regions. The dopamine transporter gene *SLC6A3*, which has been previously associated with heavy alcohol consumption and AUD, was among the significantly altered CpG-associated genes in both models. The 5'UTR CpG in the *SLC6A3* gene was significantly hypomethylated in the non-AUD group compared to the AUD group. Our study identified several novel AUD and alcohol sensitivity associated CpG sites/genes and provided evidence for the previously identified *SLC6A3* gene association with AUD and alcohol sensitivity. Future work will include examining the AUD and alcohol sensitivity related gene pathways and evaluating the alcohol sensitivity-association with epigenetic age acceleration.

Session Title: Epigenetics Poster Session III

PB2218 Allele-specific binding of EHMT2/G9a is required to maintain maternal imprintswithin Prader-Willi Syndrome candidate region

Authors:

S-E. Wang[1], M-A. Jang[2], J. Lim[1], Y. Cheng[1], Y. Kim[3], Y. Xiong[4], J. Jin[4], S. Wang[1], Y-H. Jiang[5]; [1]Yale Univ., New Haven, CT, [2]Soonhchunhyang Univ. Coll. of Med., Bucheon, Korea, Republic of, [3]St. Jude Children's Res. Hosp., Memphis, TN, [4]Icahn Sch. of Med. at Mount Sinai, New York, NY, [5]Yale Univ. Sch. of Med., New Haven, CT

Prader-Willi Syndrome (PWS) is a genomic imprinting disorder caused by the deficiency of paternally expressed genes in the human 15q11-q13 imprinting domain. The corresponding genes in the maternal 15q11-q13 region are transcriptionally repressed. The expression of paternally expressed and maternally silenced in the 6Mb imprinting domain of 15q11-q13 is coordinately regulated by a critical regulatory element designated as an imprinting center (IC), upstream of *SNRPN* gene. The PWS-IC is differentially modified by DNA methylation of 5mC and varies histone modifications in allele specific manner. Through a large scale small molecular screening using Snrpn-EGFP fusion protein, we previously reported that inhibitors of euchromatic histone lysine N-methyltransferase-2 (EHMT2/G9a) are capable of unsilencing the expression of PWS candidate genes, *SNORD116s* from the maternal allele both in PWS patient derived cells and in PWS mouse model without effects on the maternal specific DNA methylation. However, the exact molecular mechanism underlying the EHMT2-mediated imprinting maintenance has not been fully elucidated. In the present study, we report that *Ehmt2* deficiency in the embryonic (day18) and postnatal brain (p10) is capable of unsilencing the expression of *Snrpn-EGFP* gene from the maternal chromosome with reduced H3K9me2, but do not alter DNA methylation in PWS-IC. Unexpected, chromatin accessibility is not prerequisite for reactivation of maternal imprints. There is no chromatin state changes on PWS-IC, after treatment with EHMT2 or DNMT1 inhibitor. We first show allele specific 3D chromatin conformation in PWS-associated imprinted domains, contributing to imprinting maintenance. In addition, long noncoding RNA expressed from upstream region of PWS-IC interacts with EHMT2, regardless of its catalytic activity on H3K9. These results suggest that allele-specific recruitment of EHMT2 is necessary to maintain maternal imprint and offer a mechanism underlying the maternal reactivation of PWS paternally expressed genes via the inhibition of EHMT2 as an epigenetic therapy of Prader-Willi syndrome.

Session Title: Epigenetics Poster Session I

PB2219 Alternative promoter usage by NFkB tunes transcript diversity during the endothelial inflammatory response.

Authors:

A. Golebiewski, L. K. Stolze, C. E. Romanoski; Univ. of Arizona, Tucson, AZ

Endothelial dysfunction is a hallmark of complex diseases including coronary artery disease, the world's leading cause of death. Endothelial cells (ECs) line the innermost layer of blood vessels across tissues and are primarily responsible for maintaining blood-tissue barriers. When ECs are exposed to chronic inflammatory signals, they have aberrant expression of monocyte adhesion molecules and produce less nitric oxide, which contributes to leaky barrier function and invasion of immune cells into the subendothelial space. The EC inflammatory response is a key part of endothelial dysfunction; therefore, a better understanding of these mechanisms is necessary for a comprehensive understanding of complex disease risk. We and others have characterized the EC inflammatory response at the transcriptomic and epigenomic levels, but few have systematically evaluated the effects of inflammation on splicing. Alternative splicing is one way that cells quickly respond to their environment and is a previously uncharacterized mechanism by which ECs respond to inflammatory perturbations. To recapitulate inflammatory signaling in vitro, 53 primary human aortic EC cell lines were treated with 10 ng/mL interleukin 1 beta (IL1B), a pro-inflammatory cytokine, for 4 hours. We identified 1,056 genes whose RNAs are differentially spliced between IL1B and control treatments. 598 (57%) of the differentially spliced genes were also differentially expressed, out of a total of 3,437 differentially expressed genes. Of differentially spliced genes, 374 (35%) result in different protein-coding isoforms. For some genes, multiple splice variations were significant. Interestingly, 45% of all differentially spliced events involved alternative first exons. We found that binding of NFkB, a canonical inflammation-associated transcription factor, was enriched proximal to first exon start sites favored with IL1B treatment compared to first exon start sites favored with no treatment. We next used de novo motif enrichment analysis to identify motifs preferentially utilized at alternative promoters preferred under IL1B treatment. We identified that, in addition to the NFkB/RelA motif, the TCF12 motif is enriched at first exons preferred with IL1B treatment. The transcription factor TCF12 is an initiating factor in endothelial-to-mesenchymal transition, one hallmark of endothelial dysfunction. Taken together, this analysis provides novel insight into gene regulation during the EC inflammatory response and identifies a novel role for transcription factors as regulators of alternative promoter usage and protein diversity.

Session Title: Epigenetics Poster Session II

PB2220 Alzheimer's disease risk alleles are enriched in myeloid transcription factor binding sites.

Authors:

A. Chriss[1,2,3,4], D. Mattei[1,2,3,4], A. Ravi[1,2,3,4], M. Schertzer[5,6,7,8], D. Knowles[5,6,7,8], T. Raj[1,2,3,4]; [1]Dept. of Genetics and Genomic Sci., Icahn Sch. of Med. at Mount Sinai, New York, NY, [2]Nash Family Dept. of NeuroSci. & Friedman Brain Inst., Icahn Sch. of Med. at Mount Sinai, New York, NY, [3]Ronald M. Loeb Ctr. for Alzheimer's Disease, Icahn Sch. of Med. at Mount Sinai, New York, NY, [4]Icahn Inst. for Data Sci. and Genomic Technology, Icahn Sch. of Med. at Mount Sinai, New York, NY, [5]New York Genome Ctr., New York, NY, [6]Data Sci. Inst., Columbia Univ., New York, NY, [7]Dept. of Computer Sci., Columbia Univ., New York, NY, [8]Dept. of Systems Biology, Columbia Univ., New York, NY

Background: Genome-Wide Association Studies (GWAS) have revealed over 75 genetic loci associated with Alzheimer's Disease (AD) risk. Many of these GWAS hits map to noncoding regions of the genome, where it is challenging to determine the regulatory mechanisms underlying these genetic associations. Recent studies highlight the role of myeloid cells including microglia and monocytes in AD pathology, but little is known about the functional mechanisms of AD risk variants in these cell types.
Methods: To dissect potential functional mechanisms of these genetic associations with cell-type specificity, we generated ChIP-seq for four AD-implicated transcription factors - PU.1, CTCF, MEF2C, C/EBPβ - and one active histone mark (H3K27ac) in both primary human microglia isolated from postmortem brain tissue and peripheral monocytes from disease and control donors. To enhance the detection of variants that impact binding affinity, we employed a pooling strategy and performed ChIP-seq analysis on a cohort consisting of 14 microglia and 32 monocyte donors. We performed several key analyses including peak calling, motif enrichment, assessment of peak distribution, LD score regression, as well as integration with microglia eQTL data (n=550) and fine mapped variants from AD GWAS.
Results: Our results reveal a substantial enrichment of AD heritability within our myeloid-specific ChIP-seq peaks compared to other regulatory (e.g., enhancer or promoters) annotations of brain cell types. We observed heritability enrichment of AD GWAS in the peaks for PU.1 ($p = 2.40 \times 10^{-7}$), CTCF ($p = 2.13 \times 10^{-2}$), MEF2C ($p = 3.27 \times 10^{-3}$), and C/EBPβ ($p = 6.56 \times 10^{-4}$). Several of these peaks directly coincide with fine-mapped credible sets of variants from at least 18 AD GWAS loci. This includes the lead GWAS SNP located in the *RASGEF1C* locus, which is as an eQTL (*RASGEF1C*, $p= 1.52 \times 10^{-17}$) and is found within microglia ATAC-seq and H3K27ac peaks, and strongly disrupts PU.1 binding.
Conclusion: We are currently incorporating genotypes with ChIP-seq peaks to identify binding quantitative trait loci (bQTLs), linking genetic variants to changes in the binding affinity of these transcription factors. Using statistical colocalization, we will identify bQTLs that intersect fine-mapped AD risk variants, prioritizing functional variants that disrupt transcription factor binding and increase AD risk. Overall, our findings emphasize the significance of myeloid cells in AD and highlight the potential role of differential transcription factor binding in driving a significant portion of AD risk.

Session Title: Epigenetics Poster Session III

PB2221 Analysis of 14392 whole genomes reveals 5.2% of Qataris carry medically actionable variants.

Authors:

A. Elfatih[1], C. Saad[2], B. Mifsud[1], H. Mbarek[2], The Qatar genome research consortium; [1]Hamad Bin Khalifa Univ., Doha, Qatar, [2]Qatar Genome Program, Doha, Qatar

Arabic populations are underrepresented in large genome projects, therefore the frequency of clinically actionable variants in such highly consanguineous populations is largely unknown. Here, we investigated genetic variation in 14392 whole genomes from the Qatar Genome Programme (QGP) across the updated list of 78 actionable genes (v3.1) determined by the American College of Medical Genetics and Genomics (ACMG). We examined QGP variants that overlapped with the 78 ACMG genes with a minor allele frequency (MAF) < 5%. We categorized the variants that were not annotated in ClinVar as Benign (B) or Likely Benign (LB) into one of the following groups: 1) Validated and previously reported known Pathogenic (KP), 2) Expected pathogenic (EP), and 3) Rare variants of uncertain significance predicted to be pathogenic (pVUS). We identified 248 distinct variants in 50 ACMG genes that fulfilled our criteria to be included in one of the three groups. They affected 7.2% of the QGP cohort. Across the three categories, *TTN* variants were the most frequent, followed by variants in *RYR1* and *ATP7B*. The analysis revealed 104 KP variants in 425 genotype positive participants (GPP) identified based on ClinVar and HGMD disease entries. In addition, 82 PVS1 loss of function EP variants were carried by 244 GPP. Moreover, the study cohort was deeply phenotyped, featuring cardiovascular, cancer, and hypercholesterolaemia disease evidence, which allowed us to obtain significant disease-relevant phenotype association for 13 rare (MAF < 1%) variants that were either previously unreported, were reported as VUS or had a conflicting interpretation of pathogenicity in ClinVar. These variants also had in silico prediction support for pathogenicity (REVEL score > 0.7), therefore we could classify them as EP variants, which were carried by 78 GPP. Lastly, we described 49 additional variants in the pVUS list that had the same characteristics but lacked statistically significant disease association in the QGP cohort. The prevalence of reported secondary findings was 3.8%. A further 46 heterozygous variants in six genes with an autosomal recessive mode of inheritance (*MUTYH*, *ATP7B*, *BTD*, *GAA*, *RPE65*, and *TRDN*) were detected in 200 individuals, accounting for an additional 1.4%. In total, they affect 5.2% of the population. Due to the high consanguinity rate in the QGP cohort (28% in spouses and 60% in parents), actionable variants both in genes with dominant and recessive inheritance are important for developing better treatment options and preventive strategies in Qatar and the Arabic population of the Middle East.

Session Title: Epigenetics Poster Session I

PB2222 Analysis of competitive endogenous RNA (ceRNA) networks as epigenetic regulators of cardiogenesis.

Authors:

C. Bell[1], D. Bittel[2], J. E. O'Brien, Jr.[3], N. Kibiryeva[2]; [1]Kansas City Univ. - Coll. of Med., Kansas City, MO, [2]Kansas City Univ. - Coll. of BioSci.s, Kansas City, MO, [3]Ward Family Heart Ctr., Children's Mercy Hosp., Kansas City, MO

The epigenetic regulation of human embryonic heart development is not well understood. The regulation of heart specific genetic pathways is closely controlled in temporal and spatial patterning by transcription factors and post-transcriptional modifiers. Recently, functional non-coding RNAs (ncRNAs) have been discovered to play a post-transcriptional regulatory role during organ development. Competitive endogenous RNA (ceRNA) networks are large regulatory networks formed by both coding and noncoding RNA transcripts interactions. While there are several RNA:RNA interactions that are key to ceRNA network formation, one of the most important interactions is that between long non-coding RNA (lncRNA) and complementary microRNA (miRNA) transcripts. When these interactions form, the lncRNAs function as miRNA sponges, affecting downstream messenger RNA (mRNA) expression levels. These axes of interaction are frequently denoted as lncRNA/miRNA/protein coding gene axis of regulation. Evidence is beginning to accumulate, suggesting a link between dysregulation of epigenetic processes and congenital heart defects. But yet, little is known about the function and significance of ceRNA networks during fetal heart development. We compared expression patterns of lncRNA and mRNA during different gestational time points of human cardiogenesis. We identified several ceRNA networks that appear important for proper formation of the heart. Identifying these networks will help us understand complex epigenetic regulation of morphological changes in developing heart and potentially will lead to better understanding of mechanisms involved in congenital heart defects.

Session Title: Epigenetics Poster Session II

PB2223 ASD-associated alterations in chromatin accessibility and chromosome conformation pinpoint dysregulated transcription factor network and highlight genetic variants in distal regulatory regions

Authors:

J. Yin, J. Huang, J. Ou, L. Bicks, J. Haney, D. Geschwind; Univ. of California, Los Angeles, Los Angeles, CA

Despite phenotypic and genetic heterogeneity in autism spectrum disorder (ASD), recent studies have shown that molecular changes in ASD brains converge on common pathways. However, how genetic variants act via multiple epigenetic layers and contribute to ASD remains poorly understood. Here, we profiled bulk chromatin accessibility and cell-type specific 3D genome architecture in post-mortem brains of 48 ASD-affected and neurotypical subjects. We identify over 5,000 differentially accessible regions (DARs), showing coordinated changes with differential gene expression and Hi-C interactions. We demonstrate that idiopathic and syndromic (Dup15q) ASD subjects share common dysregulation of frequently interacting regions (FIREs). We further pinpoint transcription factors (TFs) underlying the DARs, and genetic variants that increase disease risk. Taken together, these findings elucidate how genetic risk factors alter the transcriptional regulatory landscape and contribute to ASD etiology. Our work provides a framework for employing multimodal profiling to decipher the molecular mechanism of complex human diseases.

Session Title: Epigenetics Poster Session III

PB2224 Assessing cell-type specificity of enhancer-promoter interactions in neurons.

Authors:

J. Koesterich, A. Kreimer; Rutgers Univ., Piscataway, NJ

Genome wide association studies (GWAS) have found that the majority of disease associated variants are located within the non-coding genome. It is still largely unknown how mutations in these non-coding sequences cause functional disruptions, even though these sequences constitute 98.5% of the human genome. One hypothesis is that they disrupt regulatory regions consisting of enhancer or promoter regions and disrupt the proper functioning of transcription. This hypothesis is challenging to explore since different cell types have different active enhancers, as well as different interactions between enhancers and nearby genes. To decipher the functional impact of non-coding mutations, we need to gain a better understanding of the mechanism by which they disrupt such regulatory regions.To this end, we collected epigenetic and gene expression data from multiple neuronal cell types, including Neural Progenitor Cells (NPCs), Excitatory cells, Inhibitory cells, and Motor Neurons. Using this data, we generated enhancer-promoter interaction pairs using the Activity By Contact (ABC) model and grouped the interaction pairs based on the pattern of their scores within each cell type. We have identified interaction pair groupings that are cell type specific and groupings that have similar interaction scores across closely related cells. For each of these groupings, we identify enrichments for cell type relevant biological processes. Next we applied bi-clustering techniques to the grouped interaction pairs network to observe and analyze the distribution of its different sub-structures. Namely, (i) single enhancer interacting with a single target gene, (ii) single enhancer interacting with multiple genes, (iii) multiple enhancers interacting with a single gene, (iv) multiple enhancers interacting with multiple genes. This analysis highlights different modes of gene regulation programs across cell types.We then predicted transcription factor (TF) binding motifs within enhancers to identify if there were preferences in TF usage and to quantify enhancers' possible involvement with transcription. Finally, we mapped neurological disease associated variants onto the predicted regulatory regions to identify enriched regions of overlap in the individual cell types. Overall, our results indicate that the enhancer-promoter interaction network is largely cell type specific and is enriched for cell type relevant biological annotations. Additionally, the clustering of this network can be leveraged to better understand the mechanism underlying disease associated variants, as well as a selection method for function assays in the studied cell type.

Session Title: Epigenetics Poster Session I

PB2225 Assessing HiFi Long Read Sequencing versus Whole Genome Bisulfite Sequencing and Methylation EPIC BeadChip Array: A Comparative Analysis Utilizing DNA Methylation Standards.

Authors:

X. Yang, R. Yancey, A. De Silva, K. Chang, J. Bhasin, K. Booher; Zymo Res. Corp., Irvine, CA

DNA methylation, one of the most abundant and extensively studied epigenetic modifications, plays crucial roles in various biological processes such as development, cancer, aging, and complex diseases. The Illumina array has been widely employed as a classical platform for high-throughput screening in large cohort studies like The Cancer Genome Atlas (TCGA). However, this type of array covers less than 3% of CpG sites in the human genome. The latest generation of DNA sequencing technologies, exemplified by the PacBio HiFi system, offers the unique capability of generating long sequence reads spanning up to 25 kilobases. Recent advancements by Pacific Biosciences (PacBio) have focused on enhancing per-base accuracy and the ability to detect DNA modifications. In this study, we evaluated the performance of PacBio HiFi sequencing using DNA methylation standards. The DNA standards comprised of human DNA enzymatically methylated at CpG sites and unmethylated human DNA derived from the HCT116 DKO cell line. 1 ug of DNA was prepared for sequencing on the PacBio Sequel IIe system using the SMRTbell prep kit 3.0. Samples were sequenced to about 8X coverage. DNA methylation data were generated by interpreting polymerase kinetics during HiFi read sequencing with a convolutional neural network developed by PacBio, and pb-CpG-tools were used to extract methylation values from BAM files. We then compared the results obtained from PacBio HiFi sequencing to those generated by the EPIC array and Whole Genome Bisulfite Sequencing (WGBS). We found that WGBS and PacBio HiFi native DNA methylation calling exhibited a high level of concordance with each other, outperforming the EPIC array in both concordance with the methylation standard and number of CpGs reported. Using a methylated standard sample, HiFi data reports approximately 85% of CpG sites have methylation ratios greater than 90%, with an average of 93% genome wide. Similarly, WGBS data presents approximately 85% of CpG sites have methylation ratios greater than 90%, with an average of 95% genome wide. In contrast, EPIC array only reports 40% of CpG sites had methylation ratios greater than 90%, and an average of 87% across the genome. These results demonstrate HiFi long read sequencing can detect DNA methylation signal accurately for regions close to 100% methylated. Our study provides insights into the performance of PacBio HiFi sequencing in detecting DNA methylation patterns and its potential as an alternative to the EPIC array. The findings from this study illustrate how DNA methylation standards can be used as a ground truth reference to evaluate DNA methylation calling models.

Session Title: Epigenetics Poster Session II

PB2226 Assessment of changes of mRNA isoform proportion after scaRNA knockdown.

Authors:

B. Earl[1], M. Filla[1], N. Kibiryeva[1], J. O'Brien[2], D. Bittel[1]; [1]Kansas City Univ. of Med. and BioSci.s, Kansas City, MO, [2]Children's Mercy Hosp., Kansas City, MO

Congenital heart defects (CHD) are the most common type of birth defect in the United States. Mendelian and chromosomal syndromes account for a small percentage of CHD cases, but nearly 80% of CHD cases are due to idiopathic genetic mechanisms that are poorly understood. Unfortunately, even with proof that alternate splicing of mRNA plays a role in the elusive genetic mechanisms that cause CHD, little data has been found to elucidate how these alternate splicing patterns lead to such lethal and debilitating heart defects. One such example of an impairing CHD is Tetralogy of Fallot (TOF). It is the most common form of cyanotic CHD that is often treated by one or several heart surgeries. This study will explore which isoforms created by alternative splicing after scaRNA1 knockdown lead to TOF. In our previous studies, it has been found that there was a significant decrease in pseudouridylation levels in the right ventricles of patients born with TOF when compared to children born with normal hearts. The amount of pseudouridylation in spliceosomal RNA U2 is influenced by scaRNA1 expression, and thus when scaRNA1 is knocked down, U2 spliceosomal activity is changed. This leads to alternative splicing of mRNA that is important in embryonic development. We propose to quantify mRNA isoform proportions after we knockdown scaRNA1 in cell culture samples to create a model for examining scaRNA1 function. By obtaining a numerical value for the isoforms, we can use this to guide our future studies that will follow how changes in mRNA isoforms contribute to the dysregulation of regulatory networks in patients with TOF.

Session Title: Epigenetics Poster Session III

PB2227 Associations between DNA methylation and cognitive function in early-stage breast cancer patients

Authors:

S. Liu[1], D. Liu[2], C. M. Bender[3], K. I. Erickson[4,5], S. M. Sereika[3,6], Y. P. Conley[3,1], J. R. Shaffer[1,7], D. E. Weeks[1,6]; [1]Dept. of Human Genetics, Sch. of Publ. Hlth., Univ. of Pittsburgh, Pittsburgh, PA, [2]Dept. of Genetics and Genomic Sci., Icahn Sch. of Med. at Mount Sinai, Pittsburgh, PA, [3]Sch. of Nursing, Univ. of Pittsburgh, Pittsburgh, PA, [4]Dept. of Psychology, Kenneth P. Dietrich Sch. of Arts and Sci., Univ. of Pittsburgh, Pittsburgh, PA, [5]AdventHlth.Res. Inst., NeuroSci. Inst., Orlando, FL, [6]Dept. of Biostatistics, Sch. of Publ. Hlth., Univ. of Pittsburgh, Pittsburgh, PA, [7]Dept. of Oral and Craniofacial Sci., Sch. of Dental Med., Univ. of Pittsburgh, Pittsburgh, PA

Nearly 33% of breast cancer patients experience decline in cognitive function before treatment, however, the biological mechanisms for this remain unclear. We aimed to use DNA methylation (DNAm), a dynamic gene regulator, to identify genes associated with cognitive function in postmenopausal women with early-stage hormone receptor-positive breast cancer.

Both objective and subjective measures of cognitive function were assessed. Seven objective cognitive function domains (executive function, mental flexibility, processing speed, attention, working memory, verbal memory, and learning and memory) were derived from a neuropsychological test battery. Subjective cognitive function was defined as the total score from the Patient's Assessment of Own Functioning Inventory. Genome-wide DNAm data were collected via the Infinium MethylationEPIC Beadchip using whole blood samples. After quality control procedures, 110 participants and DNAm measurements at 700,779 5'-cytosine-phosphate-guanine-3' sites (CpGs) were included for analysis. Epigenome-wide association studies (EWAS) for each cognitive domain phenotype were conducted while adjusting for age and verbal IQ scores, which we refer to as the cell type heterogeneity (CTH)-unadjusted EWAS. We further controlled for chip artifacts and CTH using surrogate variable analysis. Differentially methylated regions (DMR) were detected by the dmrff R package using EWAS summary statistics.

Participants were primarily diagnosed with stage 0 or stage I (73%) breast cancer and self-reported as White (89%) with an age of 62 years and a verbal IQ score of 112, on average. cg11875060 in a neurodegeneration-associated gene *SLC33A1* met the genome-wide significance threshold of 9E-08 in the CTH-unadjusted EWAS of learning and memory (*p*-value = 7.2E-08), although the signal was not observed when adjusting for CTH. In the CTH-adjusted EWAS, cg10331779 near *CTNND2* (*p*-value = 9.7E-09), which is involved in neuronal development and functioning in the Wnt signaling pathway, and cg25906741 in *MLIP* (*p*-value = 2.0E-08), were inversely associated with processing speed and subjective cognitive function, respectively. Three significant DMRs with Bonferroni-adjusted *p*-values<0.05 were detected for processing speed in/near the genes *SLC6A11*, *PRKG1/CSTF2T*, and *FAM3B*. Two DMRs were found for mental flexibility in/near the genes *PI4KB* and *SGCE/PEG10*.

In conclusion, we identified several genes associated with cognitive function domains in women with early-stage breast cancer. These findings will enhance the understanding of the biological mechanisms behind cognitive function decline in breast cancer patients.

Session Title: Epigenetics Poster Session I

PB2228 *BCL11B*-related disorder is associated with a DNA methylation episignature.

Authors:

S. Haghshenas[1], Q. Sabbagh[2,3], M. A. Levy[1], J. Kerkhof[1], H. McConkey[1,4], N. Karp[5,6], C. Prasad[5,6], V. M. Siu[5,6], T. B. Balci[5,6], D. Genevieve[2,3], B. Sadikovic[1,4]; [1]Verspeeten Clinical Genome Ctr., London Hlth.Sci. Ctr., London, ON, Canada, [2]Montpelier Univ., Inserm U1183, Montpellier, France, [3]Reference Ctr. for Rare Disease and Malformative Syndrome, Clinical Genetic Dept., Montpellier Univ. Hosp., Montpellier, France, [4]Dept. of Pathology and Lab. Med., Western Univ., London, ON, Canada, [5]Dept. of Paediatrics, Div. of Med. Genetics and Metabolism, Western Univ., London, ON, Canada, [6]Children's Hlth.Res. Inst., London, ON, Canada

Rare genetic conditions, including neurodevelopmental disorders, often present with overlapping clinical symptoms, complicating definitive diagnosis. Additionally, genetic testing often leaves many cases unresolved due to factors such as variants of uncertain clinical significance (VUSs). Recently, DNA methylation episignatures have emerged as reliable and stable biomarkers. They are currently being used worldwide to diagnose ambiguous cases in clinical settings and to reclassify VUSs. The *BCL11B*-related disorder (*BCL11B*-RD) is a disorder caused by variants in *BCL11B*. It is characterized by symptoms such as intellectual disability, distinct facial features, and immune impairment. In this study, two separate episignatures for *BCL11B*-RD were identified, and two classification models were developed. One model prioritized sensitivity, aiming to accurately identify samples with pathogenic or likely pathogenic variants in *BCL11B*. The other model prioritized specificity, aiming to minimize false positive results. Having two episignatures with distinct emphasis on sensitivity and specificity allows researchers to select the appropriate one based on the specific diagnostic needs and goals of a given scenario. These classifiers were applied to thousands of unresolved cases within the EpiSign™ Knowledge Database and identified one sample, which lacked any known variants. Subsequent review of patient records showed that a *de novo BCL11B* likely pathogenic variant had been detected on a multigene intellectual disability panel, providing support for the utility of these classifiers for screening of undiagnosed cases.

Session Title: Epigenetics Poster Session II

PB2229 Blood DNA Methylation in Infants Hospitalized for Bronchiolitis is Prospectively Associated with Recurrent Wheezing and Asthma During Childhood.

Authors:

Y. Li[1], Z. Zhu[2], C. A. Camargo, Jr.[2], J. A. Espinola[2], A. Yaskolka Meir[1], K. Hasegawa[2], L. Liang[1,3]; [1]Dept. of Epidemiology, Harvard T.H. Chan Sch. of Publ. Hlth., Boston, MA, [2]Dept. of Emergency Med., Massachusetts Gen. Hosp., Harvard Med. Sch., Boston, MA, [3]Dept. of Biostatistics, Harvard T.H. Chan Sch. of Publ. Hlth., Boston, MA

Background: Bronchiolitis is the most common lower respiratory tract infection among infants, and is associated with a substantially higher incidence of childhood asthma. DNA methylation may contribute to the elevated asthma risk among infants with bronchiolitis.
Objectives: Among infants hospitalized for bronchiolitis, we sought to identify DNA methylation differences in infant blood that are associated with subsequent outcomes of recurrent wheezing and asthma during childhood.
Methods: We analyzed data from a multi-center prospective cohort study of US infants hospitalized for bronchiolitis. Epigenome-wide DNA methylation profiling was performed using the Illumina EPIC array on blood samples collected at hospitalization from infants <12 months of age. Recurrent wheezing by age 36 months and asthma diagnosis at age 6 years were ascertained from follow-up interviews with parents. Linear regression models were used to assess the association between DNA methylation and wheezing/asthma at each CpG site, adjusting for potential confounders. The comb-p method was applied to identify differentially methylated regions (DMRs) related to wheezing and asthma based on the linear regression results. A state-of-the-art deconvolution method was employed to estimate the cell-type specific effects driving the DMRs.
Results: Among 506 infants hospitalized for bronchiolitis (median age: 3 months), 35% experienced recurrent wheezing by age 36 months, and 22% were diagnosed with asthma by age 6 years. After multiple testing corrections by the Šidák method, 21 DMRs were associated with recurrent wheezing (Šidák $p<0.05$). The top three DMRs (Šidák $p<5.0\times10^{-8}$), each containing 12 CpGs, were annotated to the promoter regions of *ANKRD3*, *GNMT*, and *ALOX12*, respectively. Notably, we identified a DMR (*chr2:54087008-54087343*, 9 CpGs, Šidák $p=7.2\times10^{-4}$) overlapping with the promoter region of *ASB3*, a gene involved in smooth muscle proliferation. This DMR was hypomethylated in recurrent wheezing cases compared to non-cases, with the hypomethylation driven by the effect in B cells. Additionally, 21 DMRs associated with asthma at age 6 years (Šidák $p<0.05$) were identified, including two (*chr17:6899085-6899758*, *ALOX12*; *chr6:42927940-42928056*, *GNMT*) overlapping with DMRs associated with recurrent wheezing by age 36 months.
Conclusion: Among infants hospitalized for bronchiolitis, blood DNA methylation at multiple novel genomic regions was associated with later outcomes of recurrent wheezing and asthma during childhood. These findings provide valuable insights into the underlying mechanism of asthma development in this high-risk clinical population.

Session Title: Epigenetics Poster Session III

PB2230 Both exercise intensity and training state influence the cell sources and magnitude of cell-free DNA accumulation in human plasma upon exercising.

Authors:

K. Rodrigues[1], Z. Weng[1], C. Tuggle[2], B. Peoples[3], G. Seay[3], C. Kelley[3], S. Yang[3], J. McAdam[2], K. Lavin[2], S. O'Bryan[3], D. Drummer[3], Z. Graham[2], M. Bamman[2], T. Broderick [2], S. Montgomery[1]; [1]Stanford Univ., Stanford, CA, [2]Florida Inst. for Human and Machine Cognition, Pensacola, FL, [3]Univ. of Alabama at Birmingham, Birmingham, AL

Exercise is critical for health and reducing risk for chronic inflammatory diseases. Exercise also drastically increases cell-free DNA (cfDNA) in blood plasma, but the origins of this DNA have not been comprehensively quantified in humans for all tissues and cell types. Neutrophils are a potential source, as they release chromatin in inflammatory contexts like exercise (NETosis). We hypothesized that enzymatic methylation sequencing (EMseq), paired with cell type proportion deconvolution, could precisely and comprehensively quantify the extent to which every cell type contributes to inducing cfDNA during exercise. To do this, we analyzed DNA methylation of cfDNA in 50 plasma samples over time from 10 healthy individuals in the Precision High-Intensity Training Through Epigenetics (PHITE) randomized dose-response trial. Participants exercised 3 d/wk for 12 weeks, followed by 4 weeks of detraining. To study the effects of exercise intensity, participants were randomized to high intensity tactical training (HITT) or traditional, moderate intensity combined endurance and resistance training (TRAD). Plasma was collected at five timepoints: before and immediately post exercise at weeks 0 and 12, then at rest during week 16. cfDNA was extracted, and EMseq was performed to measure methylation of cfDNA molecules. PCA showed post-exercise samples clustered distinctly from pre-exercise and resting samples. GREAT analysis found exercise-induced differentially methylated regions of cfDNA were enriched for immune-related gene ontology annotations. After implementing cell-type proportion deconvolution using DNA methylation atlases as reference and computing absolute abundances of cfDNA for each tissue and cell type, we observed exercise significantly increased cfDNA from neutrophils (>48-fold), dendritic cells (>5-fold), and macrophages (>2-fold). Surprisingly, muscle, bone, and fat cells did not contribute to the changes in cfDNA. Notably, we found that the amount of cfDNA released into circulation is proportional to exercise intensity (HITT > TRAD). Further, the extent of the increase in cfDNA after exercise diminished as an effect of training (wk 12 < wk 0) regardless of intensity, and specifically for dendritic cells and macrophages, but not neutrophils. These results suggest that exercise training desensitizes dendritic cells and macrophages to exercise-induced inflammatory activation. Combined, our study provides critical insight into which cell types contribute to exercise-induced accrual of circulating cfDNA and may guide understanding of how exercise could protect from or mitigate chronic inflammatory processes in humans.

Session Title: Epigenetics Poster Session I

PB2231 Cell-specific roles of 5-methylcytosine and 5-hydroxymethylcytosine in early AD onset and progression.

Authors:

Y. Li, H. Shen, B. Yao; Emory Univ., ATLANTA, GA

Alzheimer's disease (AD) is a progressive, irreversible neurodegenerative disorder representing the most common cause of dementia and affecting over 50 million people worldwide. Despite aberrant gene expression observed in multiple brain cell types during AD pathogenesis, molecular mechanisms behind these changes remain poorly understood. The most abundant DNA modification 5-methylcytosine (5mC) and its oxidized product 5-hydroxymethylcytosine (5hmC) has been extensively investigated as gene expression regulators and reported to be globally dysregulated in AD. However, accurate and comprehensive genome-wide brain cell-specific regulatory roles of 5mC and 5hmC at single-base resolution in early AD onset and progression remains elusive due to the lack of available early-stage human AD brain samples and the absence of high-resolution experimental techniques. In the present study, we employed a novel bisulfite treatment independent approach, TAPS (TET-assisted pyridine borane sequencing), to detect base-resolution DNA methylation signal by directly converting methylated cytosine to thymine and distinguished 5hmC from 5mC by additional β-glucosyltransferase labeling on 5hmC loci. Leveraging the power of TAPS combined with high-depth sequencing, we obtained accurate single-base resolution 5mC and 5hmC profiles during early onset and progression of AD using 3-month and 9-month 5xFAD, a well-established AD mouse model. Comparing with the relatively lower level of 5hmC in mouse embryonic stem cell, we detected comparable levels of 5hmC and 5mC on CpG loci in cortex of 9-month-old mouse, suggesting 5hmC accumulates in brain during development and aging. Global hyper-methylation and hypo-hydroxymethylation on CpG loci were simultaneously observed in 9-month 5xFAD mouse cortex, which is specifically preference in intragenic regions and potentially regulating expression of multiple neuronal genes such as *Grin2d* and *Bdnf*. Furthermore, integrated analysis of AD dysregulated DmR and DhmR with corresponding gene expression changes revealed specific roles of 5mC and 5hmC in gene transcriptional regulation. We then utilized immunopanning approach to successfully isolate neurons, astrocytes, microglia and oligodendrocyte progenitor cells in 3-month and 9-month 5xFAD mouse brains along with their age-matched littermate controls. We will apply TAPS to identify cell-specific DNA modification alterations during AD pathogenesis. Collectively, this study provides a comprehensive understanding of how aberrant epigenetic alteration promote early AD pathogenesis and progression in a cell type-specific manner.

Session Title: Epigenetics Poster Session II

PB2232 † Characterization of Native Allele Specific Methylation by Nanopore Adaptive Sampling Sequencing

Authors:

Y. Tian[1], L. Wang[2]; [1]Moffitt cancer center, Tampa, FL, [2]Moffitt Cancer Ctr., Tampa, FL

5-methylcytosine (5mC) is a crucial DNA modification occurring regularly across the human genome. To detect the modification at the single base resolution, bisulfite sequencing is commonly used to identify the 5mC status. However, this short read sequencing is often involved in the PCR amplification, which may cause 5mC detection bias by introducing duplicates. The short read also restricts co-methylation analysis between any two CpG sites. To address these issues and estimate the native 5mC methylation, we re-analyzed the Nanopore adaptive sampling sequencing run on human CpG islands. We applied the LD R^2 to calculate the co-methylation in nanopore data, and further identified 27875, 50481, 26542 and 51189 methylation haplotype blocks (MHB) in COLO829, COLO829BL, HCC1395 and HCC1395BL cell lines, respectively. Interestingly, while we found the majority of the co-methylation happens in a short range (≤200bp), a small portion (1~3%) occurs with long distance (≥1000bp), suggesting potential remoting regulatory mechanisms across the genome. To further characterize the epigenetic changes related to transcription factor binding, we profiled the 5mC percentage changes surrounding the JASPAR motif sites, and found that CTCF, KLF5, ZNF93 and ZBTB33 binding sites showed reduced methylation, while MEF2D, FOXE1, MAFK, and ZNF354A sites showed increased methylation in the above cell lines. To further investigate the allele specific 5mC in prostate genome, we designed a target region covering mQTL and GWAS germline variants and generated long reads with adaptive sampling run in BPH-1 and 22Rv1 cells. After phasing, we compare the methylation on both strands and identified 333 and 3031 significantly (q-value≤0.1) haplotype specific methylated regions in BPH1 and 22Rv1 cell line respectively. We also found that the haplotype specific methylation contributed largely to the MHB identified across the genome. By comparing to the ATAC-seq data in the BPH-1 and 22Rv1 cells, we found the haplotype specific methylation was highly correlated with chromatin accessibility in the human genome. Our work identified native methylome profiling with intact haplotype information and will provide innovative result to uncover the regulatory machinery of human prostate genome.

Session Title: Epigenetics Poster Session III

PB2233 Characterizing epigenetic aging in an adult sickle cell disease cohort.

Authors:

B. Lê[1], D. Hatch[1], Q. Yang[1], M. Garrett[1], N. Shah[1], P. Tanabe[1], A. E. Ashley-Koch[2], M. Knisely[1]; [1]Duke Univ., Durham, NC, [2]Duke Univ. Med. Ctr., Durham, NC

Sickle cell disease (SCD) affects approximately 100,000 predominantly African-American individuals in the United States. The pathophysiology of SCD contributes to significant cellular damage over time, increased disease complications, and ultimately premature death. While previous studies have sought to illuminate mechanisms of SCD pathophysiology through associations with genetic loci, the contribution of epigenetic factors to this pathophysiology is unexplored. DNA methylation (DNAm), a primary epigenetic mechanism for regulating gene expression in response to the environment, is an important driver of normal cellular aging. Several epigenetic clocks were developed to serve as a proxy for cellular aging, which use DNAm cytosine-phosphate-guanine (CpG) sites to calculate epigenetic age. Here, we calculated the epigenetic ages of an adult cohort of 90 people with SCD (mean age: 30.64 years; 60.64% female) across five published epigenetic clocks: Horvath, Hannum, PhenoAge, GrimAge, and DunedinPACE. We hypothesized that, in a chronic disease like SCD, patients would demonstrate epigenetic age acceleration. However, both epigenetic age acceleration and deceleration were observed in this cohort, dependent on the specific clock used. The effects of chronological age, gender, genetic ancestry, sickle cell genotype, and SCD disease severity on epigenetic age were also analyzed, to identify the presence of epigenetic acceleration/deceleration above and beyond those components. Chronologic age was significantly correlated with epigenetic age in four of the five clocks (Horvath, r=0.88; Hannum, r=0.89; PhenoAge, r=0.85; GrimAge, r=0.88). Sickle cell genotype was associated with epigenetic age in one clock (PhenoAge, p=0.02), but genetic ancestry, gender, and SCD disease severity were not associated with any clock. Finally, the CpG probes comprising the five clocks were compared to evaluate the level of overlap across the clocks, finding that each clock measures almost entirely distinct CpG sites. This likely explains the different observations of acceleration and deceleration observed within the same individual. We conclude that the specific epigenetic clock used to calculate epigenetic age influences associations with demographic and clinical phenotypes in adult SCD patients, and that further development of epigenetic clocks can improve predictive ability and utility in chronic diseases like SCD.

Session Title: Epigenetics Poster Session I

PB2234 Circulating cell-free methylated DNA reveals tissue-specific, cellular damage from radiation treatment.

Authors:

M. McNamara[1], N. Loyfer[2], A. J. Kiliti[1], M. O. Schmidt[1], S. Shabi-Porat[2], S. Jain[1], S. M. Roth[1], A. McDeed[1], N. Shahrour[1], E. Ballew[3], Y-T. Lin[1], H-H. Li[1], A. Deslattes Mays[4], S. Rudra[3], A. T. Riegel[1], K. Unger[3], T. Kaplan[2], A. Wellstein[1]; [1]Georgetown Univ., Washington, DC, [2]The Hebrew Univ. of Jerusalem, Jerusalem, Israel, [3]Medstar Georgetown Univ. Hosp., Washington, DC, [4]Sci. and Technology Consulting, LCC, Farmington, CT

Radiation therapy is an effective cancer treatment although damages to healthy tissues are common. Here we analyzed cell-free, methylated DNA released from dying cells into the circulation to evaluate radiation-induced cellular damages in different tissues. To map the circulating DNA fragments to human and mouse tissues, we established sequencing-based, cell-type specific reference DNA methylation atlases. We found that cell-type specific DNA blocks were mostly hypomethylated and located within signature genes of cellular identity. Cell-free DNA fragments were captured from serum samples by hybridization to CpG-rich DNA panels and mapped to the DNA methylation atlases. In a mouse model, thoracic radiation-induced tissue damages were reflected by dose-dependent increases in lung endothelial and cardiomyocyte methylated DNA in serum. The analysis of serum samples from breast cancer patients undergoing radiation treatment revealed distinct dose-dependent and tissue-specific epithelial and endothelial responses to radiation across multiple organs. Strikingly, patients treated for right-sided breast cancers also showed increased hepatocyte and liver endothelial DNA in the circulation indicating the impact on liver tissues. Thus, changes in cell-free methylated DNA can uncover cell-type specific effects of radiation and provide a readout of the biologically effective radiation dose received by healthy tissues.

Session Title: Epigenetics Poster Session II

PB2235 Community engagement in epigenomic and neurocognitive research on post-traumatic stress disorder in Rwandans exposed to the 1994 genocide against the Tutsi: lessons learned. 08/2022

Authors:

A. Rugamba; Univ. of Rwanda, Kigali, Rwanda

Epigenomic & neurocognitive studies have provided new perspectives on post-traumatic stress disorder & its intergenerational transmission. This article outlines the lessons learned from community engagement (CE) in such research on Rwandan genocide survivors. A strong trauma-related response was observed within the research project-targeted community (genocide survivors) during the explanation of the project. CE also revealed privacy concerns, as community members worried that any leakage of genetic/(epi)genomic data could affect not only themselves but also their close relatives. Adopting a culture of CE in the process of research implementation enables the prioritization of targeted community needs & interests. Furthermore, CE has stimulated the development of mental healthcare interventions, which married couples can apply to protect their offspring and thus truly break the cycle of inherited vulnerability. Plain language summary: Studies of how human genes are affected by the environment (epigenomic studies) have provided new perspectives on post-traumatic stress disorder & its intergenerational transmission. This article describes the lessons learned from community engagement (CE) in this type of research in a Rwandan genocide-exposed population. A strong trauma-related response was observed within the community while explaining the project. CE also revealed the participants' privacy concerns related to the leakage of genetic/(epi)genomic data that could also affect their close relatives. Adopting a culture of CE in the process of research implementation enables the prioritization of community needs and interests. CE has furthermore stimulated the development of preventive interventions for married couples to protect their offspring & thus truly break the cycle of inherited vulnerability.

Session Title: Epigenetics Poster Session III

PB2236 Comparative 3D genome analysis between neural retina and RPE reveals differential cis-regulatory interactions at retinal disease loci

Authors:

E. D'haene[1,2], V. López Soriano[1,2], P. M. Martínez-García[3], S. Kalayanamontri[3], A. Dueñas Rey[1,2], A. Sousa-Ortega[3], S. Naranjo[3], S. Van De Sompele[1,2], L. Vantomme[1,2], Q. Mahieu[1,2], S. Vergult[1,2], A. Bastos Neto[3], J. L. Gómez-Skarmeta[3], J. R. Martínez-Morales[3], M. Bauwens[1,2], J. J. Tena[3], E. De Baere[1,2]; [1]Ghent Univ., Ghent, Belgium, [2]Ctr. for Med. Genetics, Ghent Univ. Hosp., Ghent, Belgium, [3]Centro Andaluz de Biología del Desarrollo, Sevilla, Spain

Vision depends on the functional interplay between the photoreceptor cells of the neural retina and the supporting cells of the underlying retinal pigment epithelium (RPE). Many genes involved in inherited retinal diseases (IRD) display highly specific spatiotemporal expression within these interconnected retinal components through the local recruitment of *cis*-regulatory elements (CREs) in 3D nuclear space.
To understand the role of differential 3D genome topology in establishing tissue-specific expression patterns at IRD loci in the human neural retina and the RPE, we mapped genome-wide chromatin interactions by applying *in situ* Hi-C and H3K4me3 HiChIP to human adult post-mortem donor retinas. A comparative 3D genome analysis between neural retina and RPE/choroid using both Hi-C and HiChIP data revealed that almost 60% of known IRD genes were marked by differential 3D genome structure and/or *cis*-regulatory interactions. Next, we used UMI-4C to zoom in on regulatory chromatin interactions at the *ABCA4* locus, which is implicated in the most common autosomal recessive IRD. Upon integration with bulk and single-cell epigenomic datasets and *in vivo* enhancer assays in zebrafish, we revealed tissue-specific CREs interacting with *ABCA4*.
In summary, through extensive comparative 3D genome mapping, based on genome-wide (Hi-C), promoter-centric (HiChIP) and locus-specific (UMI-4C) assays of human neural retina and RPE, we have shown that gene regulation at key IRD loci is likely mediated by tissue-specific chromatin interactions. These findings do not only provide insight into tissue-specific regulatory landscapes at retinal disease loci, but also delineate the search space for non-coding genomic variation underlying unsolved IRD.

Session Title: Epigenetics Poster Session I

PB2237 Comprehensive methylation analysis of 309 children born SGA targeting imprinting disorders

Authors:

K. Yamazawa[1], M. Nakashima[1], T. Inoue[2,3], A. Nakamura[4,3], K. Matsubara[3], M. Kagami[3]; [1]Natl. Hosp. Organization Tokyo Med. Ctr., Tokyo, Japan, [2]Univ. of Tokyo, Tokyo, Japan, [3]Natl. Res. Inst. for Child Hlth.and Dev., Tokyo, Japan, [4]Hokkaido Univ. Graduate Sch. of Med., Sapporo, Japan

Background: Small-for-Gestational Age (SGA) is a heterogeneous condition caused by diverse factors, including maternal, placental, umbilical, and fetal elements. Various imprinting disorders (IDs) can be associated with fetal growth impairment, and thus, it is postulated that some cases of SGA may include IDs, though there is currently little evidence to substantiate this.
Objective: The aim of this study was to elucidate the role of IDs in SGA.
Method: A total of 309 SGA infants, having a birth weight below the 10th percentile for gestational age, were investigated. Cases with evident congenital anomaly syndromes were excluded. Buccal mucosa DNA was collected and subjected to bisulphite treatment. Subsequently, methylation analysis of ten differentially methylated regions (DMRs) associated with the onset of IDs was conducted using pyrosequencing.
Results: Of the 309 cases, methylation abnormalities in one or more DMRs suggestive of IDs were identified in 6 cases (1.9%):
Case 1: 2-year-old boy, hypermethylation in *PLAGL1*:alt-TSS-DMR, indicating a potential diagnosis of Silver-Russell syndrome (maternal uniparental disomy of chromosome 6).
Case 2: 13-year-old girl, hypomethylation in *MEG3*/*DLK1*:IG-DMR and *MEG3*:TSS-DMR, suggesting Temple syndrome.
Cases 3 and 4: 5-year-old boy and 10-month-old boy, respectively, both showing hypermethylation in *SNURF*:TSS-DMR, implying possible Prader-Willi syndrome.
Case 5: 1-year-old girl, with hypermethylation in *ZNF597*:3' DMR and hypomethylation in *ZNF597*:TSS-DMR, indicating a potential diagnosis of maternal uniparental disomy for chromosome 16.
Case 6: 5-year-old boy, with hypomethylation in *ZNF597*:TSS-DMR, suggesting a potential *ZNF597* methylation disorder.
In Cases 1 and 6, the above observations were confirmed with peripheral blood-derived DNA, establishing the diagnosis of IDs.
Discussion: This study reveals that a certain proportion of SGA cases, excluding those with apparent congenital anomaly syndromes, include IDs. Methylation anomalies and uniparental disomy, considered as epigenetic variants (epivariants), are usually assumed to occur *de novo*, a finding of significant importance for genetic counseling.
Conclusion: IDs are one of the causative factors in SGA, and methylation analysis is an effective tool for investigation.

Session Title: Epigenetics Poster Session II

PB2238 Context-specific single-cell transcriptomic analysis in hiPSC-derived neurons informs novel gene regulatory mechanism in schizophrenia

Authors:

J. Duan[1], H. Zhang[2], C. Li[3], S. Zhang[4], L. Liang[5], A. Kozlova[6], A. Sanders[4], Z. Pang[7], X. He[8]; [1]Northshore Univ Hlth.system/Univ of Chicago, Evanston, IL, [2]NorthShore Univ. Hlth.System, Evanston, IL, [3]Northshore Res. Inst., Evanston, IL, [4]North Shore Univ Hlth.System, Evanston, IL, [5]Univ. of Chicago, Chicago, IL, [6]The Res. Inst., NorthShore Univ. Hlth.System, Evanston, IL, [7]Child Hlth.Inst. of New Jersey, New Brunswick, NJ, [8]Univ Chicago, Chicago, IL

Despite a large number of risk genes identified by GWAS for neuropsychiatric disorders (NPD), their functional interpretation remains challenging, especially in the context of interaction between polygenic and environmental risk factors. Human induced pluripotent stem cell (hiPSC)-derived neurons have emerged as a promising cellular model for ascertaining the cell type-specific polygenic effects for neuropsychiatric disorders in a context-dependent manner, e.g., under neural activation that may mimic the physiological effects of environmental stimuli such as social experiences, stress, or drugs of abuse. To this end, we modelled neural activation by potassium chloride (KCl) stimulation in co-cultured excitatory/inhibitory neurons of ~100 hiPSC lines, followed by assaying activity-dependent single-cell multiomes (scRNA/ATAC-seq). For each main neural subtype (GABAergic inhibitory, NEFM+ or NEFM- excitatory neurons) clustered from scRNA-seq, we identified reproducible cell type-specific expression changes in thousands of genes after 1hr or 6hrs of KCl stimulation. We found that only the KCl-upregulated genes are enriched for synaptic gene ontology (GO) terms. Our MAGMA gene-set analysis further showed that mainly the KCl-upregulated genes were enriched for neuropsychiatric GWAS risk variants, with the strongest enrichment for schizophrenia (SZ). To further identify context-specific SZ-associated differentially expressed genes (DEGs), we performed single-neuron transcriptomic DEG analysis using MAST for 28 SZ cases and 50 controls (~700,000 cells). We found that 3-5%, 4-6%, and 15-23% of genes were SZ-associated DEGs in NEFM+ excitatory, NEFM- excitatory, and GABAergic neurons, respectively, and most of which were cell subtype-specific and stimulation-specific. Notably, Notably, DEGs in stimulated GABAergic neurons (up-regulated genes) and NEFM+ excitatory neurons (downregulated genes) were more enriched for synapse-related GO terms and for risk genes of SZ and other neuropsychiatric disorders than in unstimulated neurons. Interestingly, the upregulated genes in SZ cases in the stimulated (only at 6hr) NEFM+ excitatory neurons showed strong enrichment for GO terms related to cholesterol synthesis of which the top-ranking genes *ACAT2* and *HMGCR* were reported to be upregulated *in vitro* by atypical antipsychotic drug clozapine. Our study suggests that many neuropsychiatric genes may only elicit disease-relevant effects upon neuronal activation, providing novel insights into how polygenic risk factors interact with environmental stimuli for NPD.

Session Title: Epigenetics Poster Session III

PB2239 Cord blood epigenetic gestational age acceleration and childhood blood pressure trajectories from 3 to 15 years of age

Authors:

J. Hu[1], A. Yaskolka Meir[1], X. Wang[2], L. Liang[1]; [1]Harvard T.H. Chan Sch. of Publ. Hlth., Boston, MA, [2]Johns Hopkins Univ. bloomberg Sch. of Publ. Hlth., Baltimore, MD

Background: Epigenetic gestational age acceleration has been considered a biomarker for physiological development and maturity of fetus. Decreased epigenetic gestational age has been associated with perinatal environmental exposures and maternal adverse pregnancy outcomes. However, the impact of epigenetic gestational age acceleration at birth on childhood blood pressure (BP) trajectory remains unclear.
Methods: This study included 940 children (500 boys; 440 girls) from the Boston Birth Cohort, with data on cord blood DNA methylation (profiled using Illumina MethylationEPIC BeadChip) and BP at 3-15 years (6,570 observations; retrieved from electronic medical records). Systolic (SBP) and Diastolic (DBP) BP percentiles were calculated based on the 2017 American Academy of Pediatrics Clinical Practice Guideline. LOWESS (locally weighted scatterplot smoothing) curves were fitted to examine BP trajectories from 3 to 15 years. We identified 4 distinct trajectories for both SBP and DBP percentiles: high-steady, high-decrease, normal-increase, and normal-steady. Epigenetic gestational age was estimated using the *EPIC clock* model. Extrinsic age acceleration (EAA) was calculated as residuals of associations between epigenetic and chronological gestational ages (in weeks). Intrinsic age acceleration (IAA) was calculated using the same method with adjustment for cord blood cell compositions.
Results: We found inverse associations between EAA and IAA with repeated measures of SBP and DBP percentiles, after adjusting for maternal age, race, pre-pregnancy body mass index, education, smoking during pregnancy, gestational age at delivery, and birth weight. Significant associations were found between EAA and SBP percentiles in boys (β = -2.02; P = 0.02) but not in girls (β = -0.49; P = 0.58). Both EAA and IAA were inversely associated with SBP percentiles in girls who were born preterm (<37 weeks of gestation; β_{EAA} = -2.95; β_{IAA} = -3.00; $P < 0.05$). Compared to normal-steady SBP trajectory, significant inverse associations were observed between IAA and high-steady, high-decrease, and normal-increase SBP trajectories in boys (β = -0.21 to -0.31; $P < 0.03$), and significant positive associations were observed for high-decrease and normal-increase SBP trajectories in girls (β = 0.23 to 0.32; $P < 0.01$). Significant sex differences were also observed ($P_{sex\text{-}interaction} < 2\times10^{-16}$).
Conclusions: Epigenetic gestational age acceleration at birth was inversely associated with child BP, and such associations significantly differ by sex, indicating that reduced biological maturity at birth may be related to long-term cardiovascular risk in a sex-specific manner.

Session Title: Epigenetics Poster Session I

PB2240 CRISPR-targeting of scaRNA1 to investigate effects of mRNA splicing and heart development

Authors:

M. Brown[1], M. Filla[1], N. Kibiryeva[1], J. E. O'Brien, Jr[2], D. C. Bittel[1]; [1]Kansas City Univ., Kansas City, MO, [2]Ward Family Heart Ctr. Children's Mercy Hosp., Kansas City, MO

Congenital heart defects (CHDs) affect ~1% of babies. A clear genetic cause for CHDs has yet to be identified, but small Cajal body-specific RNAs (scaRNAs) appear to play a role in spliceosome function and regulation of heart development. scaRNAs are small non-coding RNAs that target the RNA subunits of the spliceosome for biochemical modification. Our research team identified 12 scaRNAs that were reduced in the right ventricle of babies with tetralogy of Fallot (TOF, a CHD). We subsequently showed that mRNA splicing was deregulated and furthermore we showed scaRNA played a direct role in the regulation of mRNA alternative splicing. We hypothesize that scaRNAs influence mRNA splicing and are important for heart development. To test this theory, I used CRISPR-Cas13d and a guide RNA (gRNA) that targets the scaRNA1 transcript for knockdown (KD) in quail QM7 cells to evaluate the efficiency of the vectors. These experiments will contribute to our understanding of a novel epigenetic regulatory mechanism that appears to be critical for vertebrate heart development.

Session Title: Epigenetics Poster Session II

PB2241 Deciphering Transcriptional Regulatory Elements during Inhibitory Interneuron Differentiation using Deep Learning

Authors:

R. Birnbaum[1], R. Alatawna[1], Y. Salomon[1], Y. Orenstein[1,2]; [1]Ben Gurion Univ., Beer Sheva, Israel, [2]Bar Ilan Univ., Ramat gan, Israel

During neurogenesis, the generation and differentiation of neuronal progenitors into inhibitory gamma-aminobutyric acid-containing interneurons is dependent on the combinatorial activity of transcription factors (TFs) and their corresponding regulatory elements (REs). However, the roles of neuronal TFs and their target REs in inhibitory interneuron progenitors are not fully elucidated. In this study, we developed a deep-learning-based framework to identify enriched TF motifs in gene regulatory elements (eMotif-RE), such as active, poised, repressed enhancers and putative silencers. Using histone modifications associated with gene regulatory elements datasets, such as ATAC-seq and H3K27ac/me3 ChIP-seq, from human interneuron progenitors, we distinguished with high accuracy between active enhancer sequences (open chromatin with H3K27ac) and non-active enhancer sequences (open chromatin without H3K27ac). Using our eMotif-RE framework, we analyzed potential regulatory elements and discovered enriched motifs of TFs such as ASCL1, SOX4, and SOX11 in the active enhancer set. We showed that these ASCL1 and SOX4/11 factors cooperativity function to control active enhancers of neuronal progenitors. Interestingly, we found that ZEB1 and CTCF motifs are enriched in the non-active set. We showed that ZEB1 and CTCF control the dynamic activity of regulatory elements. Using an *in vivo* enhancer assay, we showed that most of the tested putative REs from the non-active enhancer set have no enhancer activity. However, two of the eight REs (25%) showed function as poised enhancers in the neuronal system. Moreover, mutated REs for ZEB1 and CTCF motifs increased their *in vivo* activity as enhancers indicating a repressive effect of ZEB1 and CTCF on these REs that likely function as repressed enhancers or silencers. Overall, this study integrates a novel framework based on deep learning together with a functional assay that elucidated novel functions of TFs and their corresponding REs. Our approach can be applied to better understand gene regulation not only in inhibitory interneuron differentiation but in other tissue and cell types.

Session Title: Epigenetics Poster Session III

PB2242 † Deconstructing the cancer-associated Pol III transcriptome.

Authors:

K. Van Bortle, S. Lizarazo, R. K C, S. Zhou, R. Cheng, A. Jayanth, L. Mouli, D. Condroski, G. He; Univ. of Illinois Urbana-Champaign, Urbana, IL

RNA polymerase III (Pol III) transcription, which produces tRNA and other small RNA that promote protein synthesis and cell growth, is commonly upregulated in cancer. Despite evidence that specific Pol III-derived ncRNAs contribute to cancer initiation and metastasis, the mechanisms and functional significance of Pol III overactivity in cancer remain poorly defined. Additionally, due to limited genomic exploration of Pol III in humans, the full extent of Pol III transcription remains unclear, precluding a comprehensive understanding of how Pol III contributes to small RNA patterns observed in cancer. To address this, we developed a novel computational framework that integrates ChIP- and ATAC-seq to identify RNA polymerase occupancy and related gene activity signatures across the non-coding transcriptome. We apply this method to genomic data across hundreds of tissues and cancers, generating an unprecedented meta-map of Pol III occupancy and activity that recovers both canonical and non-canonical transcription patterns. Our genomic meta-map identifies a core Pol III transcriptome shared by all tissues while highlighting restricted activity patterns that emerge in human cancers. Cancer-associated Pol III signatures further segregate into a subrepertoire of ncRNAs that are related to the expression of a specific Pol III subunit, RPC7α, implicating dynamic Pol III composition as a mechanistic driver of Pol III overactivity and reorganization in cancer. By combining genomic and biochemical approaches, we further complement our survey of the cancer-associated Pol III transcriptome with evidence that Pol III overactivity increases the abundance of small RNA species retaining uncapped 5’ triphosphate signatures detected by RIG-I-like pattern recognition receptors, suggesting an intersection of Pol III upregulation and innate immune responses. These findings address important gaps in our understanding of Pol III transcription and further highlight the need for future exploration of Pol III as an important player in cancer.

Session Title: Epigenetics Poster Session I

PB2243 Deep learning predicts brain cell type-specific regulatory variants for DNA methylation and improves fine mapping for psychiatric disorders

Authors:

S. Han[1], J. Zhou[2], D. Weinberger[2]; [1]Lieber Inst. for Brain Dev., Ellicott, MD, [2]Lieber Inst. for Brain Dev., Baltimore, MD

DNA methylation (DNAm) is essential for brain development and function, and its aberrations are implicated in psychiatric disorders. Studies of DNAm quantitative trait loci (mQTLs) have uncovered numerous genetic variations associated with DNAm levels in the human brain, which may illuminate causal genetic variations within GWAS risk loci. However, most studies have relied on bulk brain samples that may not capture cell type-specific mQTLs, and face challenges in identifying functional variations impacting DNAm levels. Here we present INTERACT, a deep learning model that integrates convolutional neural network with transformer, to predict cell type-specific regulatory variants affecting DNAm levels in the human brain. We show that cell type-specific INTERACT models reveal transcription factors that potentially shape cell type-specific DNAm profiles. Furthermore, predicted DNAm regulatory variants reflect cellular context and show enrichment for heritability of brain-related traits in relevant cell types. Importantly, we demonstrate that incorporating the predicted cell type-specific effects as priors improves the fine mapping of risk loci for schizophrenia and depression, leading to the identification of potential novel risk genes within a cellular context. Our study highlights the potential of deep learning technologies in identifying cell type-specific regulatory variants, which could enhance our understanding of the genetic underpinnings of complex traits.

Session Title: Epigenetics Poster Session II

PB2244 Defining the role of MeCP2e1 in Rett Syndrome symptom progression

Authors:

V. Haghani[1], D. Yasui[1], O. Sharifi[1], I. Korf[2], J. Lasalle[1]; [1]Univ. of California, Davis, Sch. of Med., Davis, CA, [2]Univ. of California, Davis, Davis, CA

Rett Syndrome (RTT) is a progressive, X-linked dominant neurodevelopmental disorder that affects ~1:10,000 females. RTT patients suffer from seizures, language and motor regression, cognitive impairment, and metabolic abnormalities with variable disease onset after 12-18 months. Paternally inherited de novo mutations within the X-linked *MECP2* (methyl CpG-binding protein 2) gene account for roughly 95% of RTT cases. RTT patients are *MECP2* heterozygotes and thus mosaic for wild-type and mutant MeCP2 expression. MeCP2 is expressed as two isoforms, MeCP2e1 and MeCP2e2. Loss of MeCP2e1 contributes to RTT-relevant neurological phenotypes. Currently, the best characterized functions of MeCP2e1 are as a methyl CpG DNA binding protein, transcriptional modulator, and chromatin organizer. Few established links exist between cell type specific MeCP2 pathways and RTT symptoms due to the asynchronous onset of symptoms. One key phenotype in MeCP2-deficient mice is the loss of excitatory/inhibitory neuronal balance. Therefore, we aim to identify how loss of MeCP2e1 binding in Camk2a and Vip neurons correlates with gene expression and epigenetic patterns before (postnatal day 30, PND30), during (PND60), and after (PNP120 for males, PND150 for females) symptom progression in an MeCP2e1-deficient mouse model of RTT.

For this project, male mice homozygous for Camk2a-Cre and Sun1-GFP or Vip-Cre and Sun1-GFP transgenes were bred with Mecp2e1 heterozygous females to produce *Mecp2e1*$^{-/+}$, *Mecp2e1*$^{+/+}$ females and *Mecp2e1*$^{-/y}$ and *Mecp2e1*$^{+/y}$ males with Camk2a/Sun1GFP or Vip/Sun1GFP expression. Cortical nuclei were recovered using isolation of nuclei tagged in specific cell types (INTACT) at each time point. MeCP2 binding was assayed by CUT&RUN, analysis of gene expression by RNA-seq, and DNA methylation by WGBS in isolated nuclei. For RNA-seq validation, results were compared with single nuclei RNA-seq analysis of Camk2a and Vip neurons in a parallel study.

Initial findings suggest that MeCP2 acts as a distal transcriptional modulator with long-range regulatory activity, as most binding sites are approximately 50-500 kb from transcriptional start sites and associate with 1-2 genes. MeCP2 expression in cortex increases over time, thus significantly more MeCP2 bound regions are observed at later time points compared to the pre-disease stage. Gene ontology analysis revealed that genes with oxidoreductase functions emerge earliest, becoming significant at PND60 in male excitatory neurons. These findings are important for clinical translation as they identify specific gene targets in specific cell types and stage-specific changes in MeCP2e1 binding activity.

Session Title: Epigenetics Poster Session III

PB2245 Determining mechanisms of somatic and germ cell gene expression by KDM5C histone H3K4 demethylase.

Authors:

I. Venkatachalam, K. M. Bonefas, S. Iwase; Univ. of Michigan, Ann Arbor, MI

Several neurodevelopmental disorders (NDDs) are caused by mutations in chromatin regulators, and their mechanisms of altered development are unknown. Intellectual developmental disorder, X-linked syndromic, Claes-Jensen type (MRXSCJ) is one such NDD caused by mutations in KDM5C, a demethylase that removes the active transcription-associated histone 3 lysine 4 di- and tri-methylation marks (H3K4me2/3). In our *Kdm5c*-KO mouse model, we unexpectedly found that germline genes are derepressed in somatic cells including neurons. In addition to the germline genes, somatic cell genes such as neuronal genes were also dysregulated in *Kdm5c*-KO mice brains. Both germline genes and somatic-cell genes can be important contributing factors to behavioral and synaptic deficits found in *Kdm5c*-KO mice.

However, we do not know how KDM5C controls germline and somatic-cell genes. MRXSCJ patient mutations span several domains of KDM5C, including conserved domains with unknown function. Patients exhibit phenotypic variability in the severity of intellectual disability, epileptic seizures, short stature, and aggressive behavior. The molecular mechanisms that cause this range of phenotypes are unexplained. We hypothesize that KDM5C protein domains specify the regulation of different sets of genes that underlie the phenotypic variability among patients. Here, we determine which KDM5C protein segments specifically regulate germline and somatic-cell genes. To this end, we will perform a systematic deletion screen of KDM5C domains and deliver these proteins into *KDM5C*-KO human embryonic stem cells (hESCs) via lentiviral transduction and differentiate the hESCs into neuroprogenitor cells. KDM5C's gene silencing activity will be measured via qPCR of representative germline and somatic-cell genes and ChIP-seq of re-introduced KDM5C mutants and H3K4me2/3. Our research will provide insights into the functional roles of KDM5C domains, which could begin to explain the underlying biology and potential pathological mechanisms that arise from different mutations in patients with MRXSCJ. Characterizing the functional impacts of chromatin modifier domains can provide insight into the altered development course of different mutations and illuminate targeted therapeutic opportunities.

Session Title: Epigenetics Poster Session I

PB2246 Determining the molecular features of BAF-regulated chromatin loci.

Authors:

A. Gulka, D. Gorkin; Emory Univ., Atlanta, GA

Chromatin remodeling is essential for eukaryotic transcriptional regulation, and mutations in genes encoding components of the chromatin remodeling machinery are prevalent causes of cancer and neurodevelopmental disorders. In mammals, chromatin is remodeled at many cis-regulatory sequence elements (cREs) in an ATP-dependent manner by BRG-1/BRM-Associated Factors (BAF) complexes, generating an “accessible” chromatin state that permits binding by transcription factors (TFs) and other regulatory proteins. BAF complexes are deposited at tens of thousands of cREs, with little overlap between BAF-bound loci in different cell types. However, the mechanisms by which BAF complexes are localized to cREs remain unclear, as no subunits of this complex are known to recognize specific DNA sequences. Notably, inappropriate targeting of BAF complexes has been observed in several cancer contexts. Thus, in addition to furthering our understanding of fundamental aspects of transcriptional regulation, defining the mechanisms driving BAF localization may yield insight into how BAF mutations result in disease phenotypes. To identify factors that may play a role in BAF localization, we characterized patterns of TF binding and histone modification at genomic loci that depend on BAF complex activity for maintenance of their accessible state. Specifically, we measured genome-wide changes in chromatin accessibility in the immortalized human lymphoblastoid cell line GM12878 following treatment with small molecules that either inhibited or induced proteasome-dependent degradation of the BAF ATPase subunits. Importantly, the genomic distributions of over 200 TFs and histone modifications have been profiled by the ENCODE Consortium in this cell line. In agreement with previous reports, BAF perturbation resulted in broadly altered patterns of chromatin accessibility, with the most striking trend being loss of accessibility at transcription start site-distal elements. We found that these TSS-distal sites of lost accessibility were specifically enriched for binding by multiple “pioneer” TFs - including members of the Activator Protein 1 (AP-1) TF family, PU.1, and C/EBPB, all of which are known to interact in vitro with BAF complex subunits - as well as H3K4 monomethylation (H3K4me1). These findings provide a more nuanced understanding of the characteristics that distinguish BAF-dependent and -independent accessible chromatin, and implicate pioneer TFs and H3K4me1 in the localization and function of these complexes.

Session Title: Epigenetics Poster Session II

PB2247 Diet dependent epigenomic and molecular adaptations in the aging retina.

Authors:

A. Mondal[1], D. C. Brock[1], J. Gumerson[1], Z. Yang[2], K. V. Rajulpote[2], J. Kun[2], J. Tang[2], K. Jiang[1], J. Nellissery[1], A. Remaley[2], E. Chew[1], A. Swaroop[1]; [1]Natl. Eye Inst., NIH, Bethesda, MD, [2]Natl. Heart, Lung, and Blood Inst., NIH, Bethesda, MD

Purpose: Age-related macular degeneration (AMD) is a leading cause of blindness among adults and is often described as a "vascular-metabolic-inflammatory" disease with complex etiological origins. Genetics, aging and lifestyle factors such as diet and smoking are among significant contributors to AMD risk. Epigenome is an interface of cellular response to extrinsic factors like diet. Clinical trials have indicated lower risk of progression to late stages of AMD among patients adhering to Mediterranean diet (MED), while high-fat diet (HFD) can exacerbate AMD progression. AMD pathology is marked by large drusen that include apolipoproteins and cholesterol rich lipoproteins of potential dietary origin. Despite pathological links, retinal response to diet remains poorly. We designed this study to characterize aging-related retinal molecular adaptations to diets and examined both transcriptome and DNA methylation profiles.
Methods: Twelve-month-old male and female C57/BL6J mice were fed isocaloric MED or HFD for a period of six-months. In the end, retinal samples were harvested to generate transcriptomic and genome-wide methylation. Lipid and fatty acid abundances were quantified to identify diet-related retinal changes.
Results: PCA showed diet-related retinal DNA methylome clustering. Differential methylation between MED and HFD was observed around genes of MTOR signaling, complement cascade, and arachidonic acid metabolism. Transcriptomic analyses revealed significant alterations in metabolism and inflammation related pathways in the two diet groups. We found the pentose phosphate pathway to be enriched with MED related over-expressing genes such as *Taldo1*, *Pgd* and *Tkt*, some of which are targets of MTOR signaling. On the other hand, genes for multiple amino-acyl tRNA synthetases were found over-expressing in HFD retinas. In addition, fatty acid metabolism and PPAR signaling genes were found to be altering in the two diet groups. Interestingly, we also observed disparity in abundance of specific fatty acids, with the MED group showing enrichment of omega-3 and omega-6 fatty acids. All together our experiments show evidence of distinct diet-dependent molecular adaptations in the retina.
Conclusions: Our study shows notable diet-associated retinal alterations in metabolic and inflammation related genes and processes. Using an integrative approach, we have characterized diet-related molecular adaptations in the aging retina. Epigenomic shifts are known to be drivers of retinal aging and response to environment. Our results indicate a possibility of epigenome-metabolism crosstalk in the retina.

Session Title: Epigenetics Poster Session III

PB2248 Differential chromatin accessibility analysis defines regulators of human coronary artery disease.

Authors:

J. T. Stubbs[1,2], A. W. Turner[2], J. V. Mosquera[1,2], A. Gaëlle[2], C. J. Hodonsky[2], G. Cummings[2], J. Smith[1,2], C. Miller[1,2,3,4], N. C. Sheffield[1,2,3,4,5]; [1]Dept. of Biochemistry and Molecular Genetics, Univ. of Virginia, Charlottesville, VA, [2]Ctr. for Publ. Hlth.Genomics, Univ. of Virginia, Charlottesville, VA, [3]Dept. of BioMed. Engineering, Univ. of Virginia, Charlottesville, VA, [4]Dept. of Publ. Hlth.Sci., Univ. of Virginia, Charlottesville, VA, [5]Sch. of Data Sci., Univ. of Virginia, Charlottesville, VA

BACKGROUND: Genome-wide association studies (GWAS) have identified hundreds of loci associated with coronary artery disease (CAD); however, the regulatory mechanisms for candidate variants remain mostly unknown. Given that most disease-associated variants reside in non-coding, cis-regulatory elements, we evaluated the epigenomic landscape in coronary artery segments using the Assay for Transposase Accessible Chromatin followed by sequencing (ATAC-seq). METHODS: We analyzed bulk ATAC-seq of human coronary artery (HCA) tissue from 119 patients (heart transplant recipients or donors) and identified ~170,000 accessible regions. We stratified the data into three categories based on the heart and coronary segment classification: non-ischemic/non-lesion (Control; n=41), ischemic/non-lesion (I-N, n=51), and ischemic/lesion (I-L, n=27). Across these categories, we computed differentially accessible (DA) chromatin regions using DESeq2. We linked DA regions to the nearest coding genes, tested for pathway enrichment, compared categories to define the chromatin state of CAD, and prioritized GWAS candidates for follow-up. RESULTS: We found 12,226 control-specific and 1,023 I-L-specific regions in HCAs. Ischemic-lesioned CA tissue samples were enriched for immune-related GO terms, such as "immune system process" and "T cell activation." We found a novel association between ~300 CAD-associated SNPs and DA regions. We found endothelial/smooth-muscle-associated genes (e.g., *EDNRA*, *ADAMTS7*, and *COL4A2*) proximal to control-specific DA regions and CAD-associated genes (e.g., *FHL5*, *LTA*, and *C4B*) proximal to I-L-specific DA regions. Notably, we also identified two macrophage-associated genes (*C4B* and *FCHO1*) that were concordantly differentially expressed in bulk RNA-seq of lesion-containing coronary segments. SUMMARY: Consistent with global gene-set enrichment analyses, I-L HCAs are associated with inflammation, lipid, and immune epigenetic markers. We believe the role of gene regulatory networks in CAD will be better understood through continued illustration of chromatin accessibility in CAD disease specimens.

Session Title: Epigenetics Poster Session I

PB2249 Differential gene regulation associated with alcohol use disorder in human nucleus accumbens and dorsolateral prefrontal cortex.

Authors:

J. White[1], C. Willis[1], B. C. Quach[1], S. Han[2], R. Tao[2], J. Shin[2], A. Deep-Soboslay[2], B. T. Webb[1], R. D. Mayfield[3], E. O. Johnson[1], J. E. Kleinman[2], L. J. Bierut[4], D. B. Hancock[1]; [1]RTI Intl., Research Triangle Park, NC, [2]Lieber Inst. for Brain Dev., Baltimore, MD, [3]Univ. of Texas at Austin, Austin, TX, [4]Washington Univ. Sch. of Med., St. Louis, MO

Alcohol use disorder (AUD), characterized by compulsive alcohol seeking despite adverse social, occupational, or health consequences, is a leading cause of preventable death worldwide. Many genetic loci have been associated with AUD in increasingly large genome-wide association studies, and previous studies have identified differential gene expression and DNA methylation (DNAm) associated with alcohol consumption and AUD in peripheral blood as well as postmortem brain tissue. However, our understanding of the regulation of genes associated with AUD and their tissue specificity is limited. Here we integrated findings from genome-wide analyses of differential DNAm and differential gene expression by AUD status in the dorsolateral prefrontal cortex (DLPFC) and nucleus accumbens (NAc) from 120 decedents of European ancestry (59 cases with 2+ DSM-5 AUD symptoms, 61 controls with no AUD symptoms). Illumina HumanMethylation EPIC array data were analyzed using robust multivariable linear regression. Results were adjusted using *bacon* to minimize inflation. At false discovery rate (FDR) < 0.05, we identified 31 differentially methylated CpGs (DMCs) in DLPFC associated with AUD status. 52 DMCs were identified in NAc and none overlapped across the two regions. Bulk RNAseq data from the same samples yielded 20,666 genes in DLPFC and 20,324 genes in NAc for analysis. Differential expression was tested using covariate-adjusted negative binomial regression models. At FDR < 0.05, we identified 98 differentially expressed genes (DEGs) in DLPFC and 90 DEGs in NAc (12 genes significant in both). Within our significant DMCs and DEGs for each brain region, we identified three intersections, where a DMC occurred within 250kb and on the same strand as a DEG: *HMGB2* (cg19310307) in DLPFC and *HMGN1* (cg25077654) in NAc, both members of the *HMG* chromatin remodeling family, and *OGA* (cg10725316) in NAc, implicated in synaptic plasticity. We additionally tested for an interaction between DNAm and AUD status using multivariable linear regression, with expression as the outcome, for all CpGs located in gene promoter regions and found that the relationship between DNAm at cg01840401 and expression of the *TOMM7* gene differed by AUD status. *TOMM7* encodes a subunit of the translocase of the outer mitochondrial membrane and helps to stabilize the translocase complex on the outer membrane of depolarized mitochondria, which has been shown to occur more frequently in liver cells of mice exposed to alcohol. Overall, these results suggest complex gene regulatory mechanisms associated with AUD and implicate several avenues for better understanding this disorder.

Session Title: Epigenetics Poster Session II

PB2250 Dissecting the molecular mechanisms of BRPF1 in rare disease.

Authors:

M. Thaker, I. Angres, M. Singh, B. Russell, V. Arboleda; Univ. of California, Los Angeles (UCLA), Los Angeles, CA

Background: Pathogenic mutations in *BRPF1* cause a rare disease known as Intellectual Developmental Disorder with Dysmorphic Facies and Ptosis (IDDDFP). BRPF1 is an epigenetic reader that facilitates histone H3 acetylation in complex with the lysine acetyltransferases KAT6A/B. While BRPF1 is a critical epigenetic regulator, the molecular mechanisms by which BRPF1 functions in normal physiology and human disease are unknown. Our approach utilizes IDDDFP patient blood samples to study BRPF1 gene regulation (RNA-seq) as well as novel cell models to study BRPF1 protein functionality (immunoprecipitation, immunofluorescence). These cell models contain an endogenous reporter at the 5'-end of *BRPF1*, HiBiT, and have been subsequently edited to recapitulate the *BRPF1* mutations in our patient cohort. Together, these approaches will enable us to profile downstream gene expression, direct binding partners, and subcellular localization. **Methods:** To study the effects of disease-causing *BRPF1* mutations on the transcriptome, we performed RNA-seq on patient blood samples (n=6) and matched controls (n=10). Differential expression analysis between BRPF1-mutated samples and controls was performed using DESeq2. Genes with a p-adjusted<0.05 were considered significantly differentially expressed. To generate HiBiT-BRPF1 reporter lines, we performed HDR mediated CRISPR/Cas9 editing to knock-in the HiBiT tag at the 5'-end of the gene in an immortalized neural cell line. Homozygous knock-in HiBiT-BRPF1 clones were selected for a second round of editing to generate patient-specific *BRPF1* mutations. The HiBiT antibody was used for western blotting, immunoprecipitation, and immunofluorescence studies (Promega, #N7200). **Results:** Transcriptomic analysis showed that BRPF1 mutations significantly alter the transcription of 24 downstream genes in patient samples, and one gene that is highly expressed is NRG1 (log2FC=2.14, adj. p-val=3.99E-4). NRG1 has a key role in neurodevelopment through the inhibition of neural circuit mechanisms involved in synaptic plasticity. To facilitate functional studies, we knocked-in the reporter, HiBiT, at the 5'-end of *BRPF1* in a neural cell line. Western blotting, immunoprecipitation, and immunofluorescence assays confirmed localization of the protein. To study the effects of pathogenic mutations on BRPF1 functionality, we generated clonal, patient-specific *BRPF1* mutations in the HiBiT-BRPF1 lines to perform the studies above. **Conclusions:** In this project we are utilizing orthogonal approaches to investigate the normal and dysregulated functions of BRPF1 in rare disease using patient blood samples and novel cell models.

Session Title: Epigenetics Poster Session III

PB2251 Distinct methylation patterns in carriers and non-carriers of *APOL1* risk alleles in African American and Hispanic/Latino populations: results from the Population Architecture using Genomics and Epidemiology (PAGE) consortium.

Authors:

A. W. Scadden[1], X. Zhang[2], Y. Zheng[3], V. Buchanan[2], M. Graff[2], K. Ferrier[1], A. E. Justice[4], Y. Qu[3], L. Hou[3], C. Gignoux[1], S. Buyske[5], S. Rich[6], T. Matise[5], C. Kooperberg[7], C. Haiman[8], R. Loos[9], K. E. North[2], A. Mottl[2], M. A. Stanislawski[1], H. M. Highland[2]; [1]Univ. of Colorado Anschutz Med. Campus, Aurora, CO, [2]Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, [3]Northwestern Univ., Chicago, IL, [4]Geisinger Hlth., Danville, PA, [5]Rutgers Univ., Piscataway, NJ, [6]Univ. of Virginia, Scottsville, VA, [7]Fred Hutchinson Cancer Ctr., Seattle, WA, [8]Univ. of Southern California, Los Angeles, CA, [9]Icahn Sch. of Med. at Mount Sinai, New York City, NY

Chronic kidney disease (CKD) is disproportionately higher in individuals with African ancestry compared to individuals with primarily European ancestry, partly due to the higher prevalence of risk haplotypes (termed G1 and G2) of the apolipoprotein L1 (*APOL1*) gene on chromosome 22. While much is known about APOL1's role in immunity, the mechanisms surrounding *APOL1* risk alleles and their involvement in CKD are not fully understood. Here we compared DNA methylation patterns in African American and Hispanic/Latino carriers of two *APOL1* high-risk alleles versus non-carriers in six cohorts from the Population Architecture using Genomics and Epidemiology (PAGE) consortium (N = 3,926; 770 [19.6%] carriers, 3,156 [80.4%] non-carriers). Methylation of CpG sites in individuals with *APOL1* high-risk genotypes (G1/G1, G2/G2, and G1/G2) were compared to non-carriers (G0/G0) methylome-wide; heterozygotes (G1/G0 or G2/G0) were excluded. DNA methylation levels were measured using Illumina 450K or 850K BeadChip arrays and transformed to M-values for modeling. We estimated associations between the *APOL1* haplotypes and 790,422 CpG sites using linear regression models, stratified by study and controlling for age, age-squared, sex, smoking status, 10 genetic principal components (PCs), estimated blood cell proportions, and cohort-specific covariates. We then performed fixed-effects sample size-weighted meta-analysis and adjusted the p-values using the Benjamini-Hochberg method (FDR ≤ 0.05).
APOL1 risk haplotypes were significantly associated with methylation levels at 56 CpG sites. Of these 56 CpG sites, 48 were located on chromosome 22, many of which (N = 19) were located in the gene regions of *APOL1-4* and *MYH9*, which have also been previously reported to be associated with CKD. Among our findings on other chromosomes, some of the genes with *APOL1*-associated CpG sites are involved in cellular proliferation (e.g *ATAD2B* on chr2 and *DEPDC1B* on chr5) or cellular transport mechanisms (e.g. *SLC12A7* on chr5 and *SLC16A3* on chr17). Our results reveal differential methylation as a possible mechanistic link between *APOL1* risk haplotypes and CKD, which may help to open new avenues for CKD targeted interventions, such as early detection and personalized treatment approaches.

Session Title: Epigenetics Poster Session I

PB2252 DNA Methylation Association with Energy Homeostasis in a Hispanic Childhood Obesity Cohort

Authors:

C. Serrano[1], M. Kos[2], N. Butte[3], A. Comuzzie[4], S. Cole[5], M. Carless[5]; [1]Univ. of Texas at San Antonio, San Antonio, TX, [2]The Univ. of Texas Rio Grande Valley Sch. of Med., San Antonio, TX, [3]Baylor Coll. of Med., Houston, TX, [4]The Obesity Society, Rockville, MD, [5]Texas BioMed. Res. Inst, San Antonio, TX

Obesity is an ongoing issue characterized by an imbalance of energy intake, storage, and expenditure, which disproportionately affects Hispanics and African Americans. Neither genetics nor environment have been determined as solely causal, and therefore epigenetics is a potentially relevant mechanism contributing to obesity. We are examining the role of DNA methylation in obesity and energy metabolism in Hispanic children (age 4-19 years) that participated in the Viva la Familia study (VIVA). The VIVA study was designed to genetically map childhood obesity in a familial Hispanic population and included 24-hour room calorimetry, providing extensive phenotypic data on energy utilization and metabolism for our analyses. Using the Roche KAPA HyperPrep Targeted MethylSeq system and incorporating a custom probe pool that utilizes a hybrid candidate gene/genome-wide approach, we have profiled blood-based DNA methylation signatures for 2.6 million CpG sites in 916 Hispanic children using the NovaSeq 6000 system. Using SOLAR, we performed preliminary association analysis for BMI in 350 children from the VIVA cohort using the inverse-normalization of BMI and adjusting for covariates (sex, age, age^2, age x sex interaction, and age^2 x sex interaction). After adjusting for multiple comparisons using Benjamini and Hochberg's FDR method, we did not observe statistically significant associations. Among the most significantly associated genes was *KLB* ($p=2.3x10^{-7}$), which has been implicated in human obesity studies and was found to be underexpressed in children with low physical activity. Using DAVID, we performed functional annotation and enrichment testing of the 5,000 most significant CpG loci associated with BMI. The most significantly enriched pathways included those related to thermogenesis, insulin secretion, insulin resistance and adipocytokine signaling. We are currently analyzing the full data set of 916 samples, which we expect will allow us to identify DNA methylation loci indicative of obesity risk in a relevant, at-risk population.

Session Title: Epigenetics Poster Session II

PB2253 DNA methylation changes associate with kidney structural injury in an American Indian cohort with type 2 diabetes.

Authors:

S. E. Day[1], R. L. Hanson[2], H. C. Looker[2], C. Koroglu[2], K. Susztak[3], R. G. Nelson[2], C. Bogardus[2], L. J. Baier[2]; [1]Midwestern Univ., Glendale, AZ, [2]NIDDK, Phoenix, AZ, [3]Univ Pennsylvania, Philadelphia, PA

Diabetes is a leading cause of chronic kidney disease with characteristic structural lesions, but the molecular mechanisms underlying these lesions are not well understood. DNA methylation is an epigenetic mechanism that can regulate gene expression. The goal of this project was to identify DNA methylation patterns associated with kidney structural injury in American Indians with type 2 diabetes. Morphometric evaluation of kidney tissue by light and electron microscopy was performed to analyze 9 structural parameters in kidney biopsies from 76 participants (mean age=46 ± 10 years; 75% female). For all participants, DNA methylation was measured using the Infinium HumanMethylation450 array in whole blood collected from an exam within 2 years of the kidney biopsy (73% are within one year). The methylation data were analyzed for association with each of the 9 kidney parameters by analysis of variance (ANOVA) using Partek Genomics Suite. Among 196,642 DNA methylation probes that mapped to known enhancer/promoter regulatory regions, 86,110 had some evidence for association ($P < 0.05$) with at least one kidney structural parameter. Some probes associated with more than one parameter, and 5 probes associated with 7 of the 9 parameters. One probe (cg02213103) with 7 associated parameters is near *SOX2OT*, which alleviates glucose-induced kidney damage. These methylation patterns suggest associations encompassing many interconnected structural elements affected by diabetic kidney disease (DKD). However, since specific kidney structural lesions may also occur at different stages of disease progression, some valid associations may only occur with one measure. Probes of interest with only one association include cg10954392 near *BUB1* associated with glomerular basement membrane width ($r = -0.52$; $P = 3x10^{-6}$); *BUB1* is involved with acute kidney failure. Probe cg17285536 near *OGG1* associated with global glomerular sclerosis ($r = 0.50$; $P = 5x10^{-6}$); *OGG1* has a key role in kidney disease progression through a range of physiological processes. Probe cg13976219 near *INSR* associated with percent fenestrated endothelium ($r = 0.52$; $P = 2x10^{-6}$); *INSR* is critical for insulin signaling and is expressed in glomerular podocytes. Probe cg13614753 near *SEMA3B* associated with glomerular filtration surface density ($r = 0.59$; $P = 3x10^{-8}$); *SEMA3B* is an antigen in the glomerular basement membrane in membranous nephropathy. In summary, many epigenetic targets were identified that may potentially lead to better understanding of mechanisms responsible for morphometric changes involved in DKD. Our plans for this ongoing project include functional assessment of these DNA methylation sites.

Session Title: Epigenetics Poster Session III

PB2254 † DNA methylation correlates of chronological age in diverse human tissue types

Authors:

N. Jain[1], L. Tong[1], J. Li[1], K. Demanelis[2], M. Kibriya[1], F. Jasmine[1], L. Chen[1], B. Pierce[1]; [1]The Univ. of Chicago, Chicago, IL, [2]Univ. of Pittsburgh, Pittsburgh, PA

While the association of chronological age with DNA methylation in whole blood has been extensively studied, the tissue-specificity of age-related DNA methylation changes remains an active area of research. Studies investigating the association of age with DNA methylation in tissues such as brain, skin, immune, fat, and liver have identified tissue-specific CpGs, thus, motivating future studies to examine other human tissues. Here, we performed an epigenome-wide association study (EWAS), leveraging DNA methylation data (Illumina EPIC array, >800,000 CpG sites) from 9 tissues from The Genotype-Tissue Expression (GTEx) project: breast (n=38), testis (n=50), kidney (n=50), prostate (n=119), ovary (n=157), colon (n=223), lung (n=223), skeletal muscle (n=47), and whole blood (n=54). To examine how DNA methylation across the genome is associated with chronological age, we used a linear regression adjusting for sex, BMI, race/ethnicity, ischemic time, batch variables, and surrogate variables representing cell type composition. We identified age-associated differentially methylated CpGs (false discovery rate < 0.05) in 8 tissues (all except skeletal muscle). This included 132,841 unique hypermethylated and 68,559 hypomethylated CpGs, with 111,933 hypermethylated CpGs and 59,480 hypomethylated CpGs detected in a single tissue. Age-related CpGs in two or more tissues (e.g., *ELOVL2* (n=8) and *ZNF549* (n=7)) were found near gene regions known to be associated with aging. While we observed that a majority of age-related differentially methylated CpGs are tissue-specific, the patterns of enrichment among genomic features, such as chromatin state and CpG island status are similar across most tissues, suggesting common mechanisms underlying chronological aging. Consistent with previous findings, we observed that age-related hypermethylated CpGs are enriched in regions with polycomb repression marks and CpG islands, while hypomethylated CpGs preferentially occur in non-CpG islands and regions with active histone marks (e.g., enhancers). Additionally, we extended our analysis to assess the correlation between age-related gene expression and DNA methylation changes to identify age-associated expression quantitative trait methylation (eQTMs). Identifying age-related eQTMs provides insights into the functional effects of these DNA methylation changes, thus, improving our interpretation of the EWAS results. Overall, our findings will aid future work on identifying biomarkers of chronological aging and understanding the mechanisms of aging in different human tissues.

Session Title: Epigenetics Poster Session I

PB2255 DNA methylation in *ABCG1* mediates an effect of lipoproteins on atherosclerotic markers

Authors:

M. Nasr[1], A. Teumer[2]; [1]Univ. of Med. Greifswald, Greifswald, Germany, [2]Univ. Med. Greifswald, Greifswald, Germany

The association between atherosclerosis and cardiovascular disorders including myocardial infarction and stroke is widely established. Although it is widely accepted that atherosclerosis is associated with chronic low grade inflammatory processes in the vessel wall, the atherosclerotic process itself is not fully understood. Blood metabolome and DNA methylation changes may play a role in the development or downregulation of the underlying immune response related to atherosclerosis. The aim of this study is to identify markers of the plasma metabolome, i.e. lipoproteins associated with subclinical atherosclerosis and its risk factors, and to analyse if these associations are mediated by DNA methylation. Serum values of 112 different lipoprotein measurements assessed in up to n=4010 individuals of the population-based German SHIP-Trend cohort were grouped according to their densities and regressed with atherosclerotic and low grade inflammatory markers related to carotid intima media thickness (IMT), periodontitis, retinal microcirculation, and cytokines adjusting for potential confounders and technical factors. Next, an epigenome-wide association study (EWAS) on lipoprotein levels was conducted in a subset of 490 individuals typed from blood using the EPIC array. Significant CpG associations were selected by Bonferroni correction for the number of sites and lipoproteins. The regression results revealed that lipoproteins were significantly (false discovery rate < 0.05) associated with multiple analyzed traits. Among these results, IMT (n = 4010) and meta-analyzed low grade inflammatory markers (n = 209) were robustly associated with low density lipoproteins potentially mediated by DNA methylation. In detail, our EWAS detected one CpG site on chromosome 11 (cg05325763) and another CpG site on chromosome 21 (cg06500161) in *ABCG1* that were significantly associated with low density lipoproteins ($p < 4.73E-10$), DNA methylation beta of cg06500161 was associated with an increase in serum soluble tumor necrosis factor receptors 2 (STNF2) ($p = 0.024$), while (cg05325763) methylation beta was associated with a decrease in IL10 levels ($p = 0.044$). Our results provide a step towards understanding the inflammatory process of atherosclerosis. Next, we will focus on its impact on cardiovascular outcomes and on causal pathways of these associations.

Session Title: Epigenetics Poster Session II

PB2256 DNA methylome analysis of trauma exposure in the human postmortem cortex

Authors:

D. Nunez-RIos[1], T. Meyer[2], S. T. Nagamatsu[1], J. Martínez-Magaña[1], G. Fries[3], L. Stertz[2], C. Walss-Bass[2], J. Montalvo-Ortiz[1]; [1]Div. of Human Genetics, Dept. of Psychiatry, Yale Univ. Sch. of Med., New Haven, CT, [2]Dept. of Psychiatry and Behavioral Sci., Univ. of Texas Hlth.Sci. Ctr. at Houston, Houston, TX, [3]Dept. of Psychiatry and Behavioral Sci., Univ. of Texas Hlth.Sci. Ctr. at Houston, New Haven, CT

Background: Growing literature from our group and others have shown that epigenetic marks, which mediate the interplay between genetic and environmental factors, are associated with trauma and trauma-related disorders. However, these studies have been limited to peripheral tissue. Given the tissue and cell-type specificity of epigenetic marks, research evaluating human brain tissue is warranted. **Methods:** We assessed brain DNA methylation (5mC) patterns of trauma exposure in the prefrontal cortex (PFC) of 92 postmortem brain samples: 54 samples from the UTHealth Brain Collection (UTHBC) and 38 samples from the National Center of PTSD Brain Bank (NCPBB). Trauma exposure was defined as being exposed to at least one of the following: physical or sexual abuse, illness or accident, threat and loss of close relative, combat. In the UTHBC, 5mC was assessed using the Illumina EPIC BeadChip array and analyzed using meffil, minfi, and Enmix R packages for the epigenome-wide association analysis (EWAS). In the NCPBB, 5mC was assessed using reduced representation oxidative bisulfite sequencing and the EWAS was conducted using the methylKit R package. Genome-wide significance (GWS) was defined using Bonferroni correction. We also evaluated replication across the cohorts. **Results:** No GWS CpG sites were found in the UTHBC cohort. In the NCPBB, we identified 1141 GWS CpG sites associated with trauma exposure. Among these, cg05711542 was replicated in the UTHBC cohort (p value 0.02605). This CpG site maps to the exon 1 of *ZNF326*, a gene involved in alternative splicing. When evaluating CpG sites within 10 base pairs upstream and downstream of the identified GWS CpG site. Five additional CpG sites were replicated in the UTHBC cohort, including cg02009280 and cg18446159**NGFR*, cg13790603**THRB*, cg13031623 and cg06392442**STOX2*, and cg05250953**CDK6*. **Conclusion:** Our EWAS analysis in individuals with trauma exposure revealed genes with differential DNA methylation in the human postmortem cortical tissue replicated in an independent cohort. Our findings shed light on the epigenetic dysregulation in the brain associated with trauma exposure.

Session Title: Epigenetics Poster Session III

PB2257 Dynamic Alu CpG Methylation Changes of Dermal Fibroblasts in Normal Aging and Diastolic Hypertension

Authors:

S. Jiraboonsri, J. Meevassana; Chulalongkorn Univ., Bangkok, Thailand

Epigenetic alteration is an important element of the aging process. Changes in DNA methylation have been observed in association with chronological aging and age-related disease such as hypertension. Nonetheless, methylation of Alu element, the most abundant interspersed repetitive elements are largely still unexplored. Therefore, in this study, we aimed to determine the dynamic Alu methylation change during chronological aging and progression of hypertension, a recognized age-related disease using primary fibroblast as a model. Biopsies of dermal fibroblasts were performed on 39 patients with various ages and blood pressure statuses. The patients were grouped into three categories based on their systolic blood pressure (SBP) and diastolic blood pressure (DBP): Normal (SBP = 130 mmHg; DBP = 85 mmHg), High Normal (SBP = 130-139 mmHg; DBP = 85-89 mmHg) and Hypertensive (SBP ≥ 140 mmHg; DBP ≥ 90 mmHg) and three groups according to their age into three age groups: young (< 45 years old), middle (45-60 years old) and old (> 60 years old). The DNA was then extracted and subjected to Combined Bisulfide Restriction Analysis and Alu PCR to determine the level and pattern of Alu methylation. We found a highly significance difference of Alu methylation level between young, middle and old age groups using Two-way ANOVA with DBP status as covariate (31.2 ± 2.14%, 30.7 ± 1.79%, and 29.4 ± 2.03 respectively; $p < 0.01$). A statistically significant difference of Alu methylation level was also found between DBP groups with high normal the highest followed by the hypertensive and normal group (29.6 ± 1.85%, 31.1 ± 1.78%, and 30.2 ± 2.48% for normal, high normal and hypertensive respectively; $p < 0.05$). Regarding the pattern, the increase in proportion of $^{u}C^{u}C$ methylation pattern was observed with higher chronological age while the dynamic change of $^{m}C^{u}C$ and $^{u}C^{u}C$ proportion was found in the progression of diastolic hypertension. No methylation change was found in relation with SBP. Interaction between DBP and chronological aging was also observed. In conclusion, there is a dynamic Alu methylation change with possible methyl group transfer between Alu loci in the process of diastolic hypertension. Furthermore, Alu hypomethylation is associated with chronological aging which is in line with global hypomethylation found in aging.

Session Title: Epigenetics Poster Session I

PB2258 Effect of artificial light modern lifestyle and their association with circadian gene expression

Authors:

N. Fatima; King George's Med. Univ., Lucknow, India

The large no. of world population daily exposed to the blue light, from few minutes to several hours. Blue light/ screen light provide negative effect on circadian rhythm which causes sleep deprivation and develops many diseases. Total 12 Wistar rats were enrolled and divided into two groups. Control group and Blue Light (BL) treated group which consist of six rats in each group. BL model was developed by placing the rats in 12hr blue light and 12 hr in dark till three months. Half of the rats were sacrifices and remaining rats were shifted to Normal Light (NL) for three months to see the legacy effect. These rats were sacrificed and blood samples were collected. Body weight was measured monthly, with blood glucose, Insulin, melatonin, lipid profile were estimated and mRNA were expressed by RTPCR. The percentage gain of body weight of BL treated group was **27.9%** as compared of control group (**20.2%)**. Blood glucose levels were increased and circulatory level of insulin, melatonin, total cholesterol, TG, HDL and LDL were decreased in blue light treated rats. BL treated group were shifted on NL they showed increase in **2.3%** body weight, melatonin and total cholesterol and HDL were significantly increased (**p= 0.0258)** and **(p= 0.037)** respectively. Level of glucose were significantly decreased (**p>0.05**). Per1 and Bmal1 gene were up regulated in BL group and the expression of these gene shows the significantly down regulated in NL group **p= 0.0394 and p= 0.0403** respectively. We found that the exposure of artificial light increases the prevalence of obesity and metabolic disorders.

Session Title: Epigenetics Poster Session II

PB2259 Effects of Assisted Reproductive Technology on genome-wide DNA methylation, imprinting and gene expression in human placenta.

Authors:

P. Auvinen[1], J. Vehviläinen[1], I. Laukkanen[1], K. Rämö[1], H. Marjonen[1], E. Wallén[1], V. Söderström-Anttila[2,3], H. Kahila[3], C. Hyden-Granskog[3], T. Tuuri[3], A. Tiitinen[3], N. Kaminen-Ahola[1]; [1]Dept. of Med. and Clinical Genetics, Medicum, Univ. of Helsinki, Helsinki, Finland, [2]The Family Federation of Finland, Fertility Clinic, Helsinki, Finland, [3]Dept. of Obstetrics and Gynecology, Helsinki Univ. Hosp., Univ. of Helsinki, Helsinki, Finland

Background: Approximately one in six couples experience infertility worldwide and the usage of Assisted Reproductive Technology (ART) for its treatment has increased over the last decades. ART has been associated with increased risk for adverse perinatal and life-long outcomes but the mechanisms by which ART affects these risks are still largely unknown. ART has been suggested to affect fetal development through epigenetic modifications as the procedures take place during extensive epigenetic remodeling occurring in the periconception period and early development. Although ART-associated DNA methylation (DNAm) alterations have been found, especially in developmentally important and epigenetically regulated imprinted genes, it remains unclear which procedures contributes to these effects and what is the functional relevance in terms of gene expression. Moreover, it is still debated whether the adverse outcomes are a result of ART methods, underlying parental subfertility, or both.
Methods: To explore the effects of ART, we collected full-term placental samples from a total of 80 ART and 100 naturally conceived newborns. To discern the effects of different ART methods, we divided the ART samples into subgroups of newborns derived through *in vitro* fertilization (n = 50) and intracytoplasmic sperm injection (ICSI) (n = 30) and to study the effects of embryo freezing, into subgroups of newborns derived from fresh (FRESH) (n = 42) and frozen-thawed (FET) (n = 38) embryos. Newborns' anthropometrics was analyzed by using standard deviations (SDs) based on Finnish growth charts. We performed genome-wide DNA methylation (DNAm) and gene expression analyses of placentas by using microarrays (EPIC, Illumina) and mRNA sequencing, respectively.
Results: The birth measures (SDs) or placental weights (g) did not differ between ART and control newborns. However, FET placentas were significantly heavier compared to FRESH placentas and the head circumference of ICSI female newborns (n = 18) was smaller compared to control females (n = 48) (P = 0.022 and P = 0.015, respectively). Our preliminary results showed no alterations in placental genome-wide average DNAm or in repetitive elements' DNAm but there were 164 ART-associated differentially methylated CpGs with DNAm difference greater than 5% and 787 differentially methylated regions (DMRs). DMRs were associated with several imprinted genes as well as with genes that were enriched in pathways involved in hormonal regulation, nervous system development, and immune defense. Moreover, 71 differentially expressed genes associated with ART and were enriched in pathways involved in vascularization.

Session Title: Epigenetics Poster Session III

PB2260 Enhanced Single Cell DNA Methylation Analysis Using Combinatorial Sequencing.

Authors:

S. Khare[1], H. Sin[2], A. Woodfin[1], E. Pu[2], L. Lin[2], P. Dang[2], D. Skinner[1], A. Beck[1], M. Nakamoto[1], B. Walczak[2]; [1]Scale BioSci.s, San Diego, CA, [2]Scale BioSci.s, San Carlos, CA

The epigenetic landscape of the human brain undergoes a plethora of changes during natural development and in malignant transformation. In recent years, DNA methylation based epigenetic modifications have been widely studied using traditional techniques like bisulfite sequencing and enzymatic methyl sequencing (EM-seq). However, these methods analyze bulk cell populations and lack the granularity of single-cell analysis. Although advances in single-cell analysis have revolutionized our understanding of cellular heterogeneity and functional diversity within complex biological systems, they are still costly, low throughput, and laborious. To address these challenges, ScaleBio has pioneered combinatorial indexing technology, enabling a significant increase in cell throughput. This novel method utilizes the cell itself as a compartment to perform 2-3 rounds of sequential barcoding in a plate-based workflow, eliminating the need for complex and expensive instrumentation. This technology has been successfully adapted to assess DNA methylation at the single-cell level offering a robust, affordable, high-throughput protocol that enhances yield, diversity, and coverage. In this study we used ScaleBio's single-cell methylation kit to investigate DNA methylation patterns during development and oncogenesis using fetal and adult brain cells along with cancer cells with widespread DNA methylation changes, such as isocitrate dehydrogenase (IDH) mutant glioma cells at the single-cell level. The IDH gene family, comprising of *IDH1-3*, encodes enzymes involved in the tricarboxylic acid (TCA) cycle. Mutations in *IDH1/2* genes are frequently found in tumors and exhibit a distinct hypermethylation pattern. We achieved high cell recovery and robust cytosine coverage throughout our analysis of single cell methylomes isolated from brain tissue. Using this data we generated a ranked list of the top hypo- and hypermethylated genomic regions and identified cell type specific clusters seen in different developmental or pathological states by looking at Differentially Methylated Regions (DMR), and compared them to known bulk methylation profiles, uncovering unique single-cell methylation profiles that may be obscured by bulk or pseudo-bulk analysis. Taken together these data show that the ScaleBio single-cell methylation workflow offers increased sensitivity, specificity, and accuracy in identifying DNA methylation sites when compared to previously available techniques while offering a comprehensive view of methylomes and providing unprecedented insights into cellular heterogeneity and trajectories.

Session Title: Epigenetics Poster Session I

PB2261 Epigenetic characterization of pseudogenes across human tissues

Authors:

Y. Jiang, B. Borsari, M. B. Gerstein; Yale Univ., New Haven, CT

Pseudogenes have long been considered non-functional remnants of genome evolution. With the advent of long-read RNA sequencing, we now have a better understanding of their genome-wide annotation as well as their transcriptional activity across tissues and organs. Nonetheless, little is known about epigenetic features that distinguish pseudogenes from protein-coding genes. Here, we extensively characterize promoters of pseudogenes from an epigenetic perspective by integrating multi-omics data across 26 human tissues generated by the ENCODE EN-TEx project. We analyzed RAMPAGE data to identify Transcription Start Sites (TSSs) at base-pair resolution. We found that pseudogenes are systematically less supported by RAMPAGE peaks compared to protein-coding genes, independently of their degree of tissue-specificity, the accuracy of their annotation, or their level of gene expression. Thus, we propose that a subset of pseudogenes may undergo TSS choices different from what was previously annotated, with potential epigenetic mechanisms playing a role in these choices. Moreover, whenever present, RAMPAGE peaks at pseudogenes are usually upstream of the annotated TSSs, contrary to protein-coding genes, which have more often downstream peaks. When focusing on pseudogenes with RAMPAGE-validated TSSs, we found that they show persistently lower histone marking, especially H3K4me3, and H3K27ac, across tissues compared to protein-coding genes and long non-coding RNAs. These differences in epigenetic marking appear unrelated to differences in the level of expression of the different classes of genes. This suggests that patterns of histone modifications may instruct the genome to distinguish different biological types of genes, in particular pseudogenes from protein-coding and lincRNAs. In conclusion, our research lays the foundation for a more profound epigenetic comprehension of pseudogenes.

Session Title: Epigenetics Poster Session II

PB2262 Epigenetic research contribution to public health: experience of an academic health center in Brazil during the COVID-19 pandemic

Authors:

L. Kulikowski, Ana Ritto, Lucas Vieira, Gleyson Carvalho, BeatrizWolff, Marina Silva, Ester Sabino, Geraldo B Filh; Univ.e de Sao Paulo, São Paulo, Brazil

The COVID-19 pandemic, caused by the SARS-CoV-2 virus, has triggered an urgent global research effort to mitigate its impact. The Hospital das Clínicas and Faculdade de Medicina of the University of São Paulo (HC-FMUSP), the largest academic health system in Latin America, has installed an emergency institutional task force to support interdisciplinary research initiatives on COVID-19 and a Steering Committee was established to map, monitor and support research groups to utilize these institutional resources. Samples and correlated data from the hospitalization of more than 5,500 adults with a confirmed diagnosis of COVID-19 between October 2020 and April 2021, and from 800 adults reassessed between 6 and 11 months after hospitalization for COVID-19 were stored and compiled. There are more than 45,000 blood samples supporting different research initiatives that investigate pathophysiological aspects of COVID-19, disease sequelae to a wide range of biomarkers of inflammation, neurodegeneration, intestinal permeability, peptidomics and metabolomics, among others. In this context, the Cytogenomics Laboratory (HC-FMUSP) is studying Brazilian epigenomic data related to COVID-19. We analyzed 72 DNA samples using the Infinium ImmunoArray-24v2 BeadChip and identified patients who developed multiple sclerosis and asthma post-covid-19 and who have allelic compatibility for six specific SNPs associated with the respective diseases. Another study constructed the differential methylation profile of 30 hospitalized patients with severe and mild symptoms without comorbidities using the Infinium MethylationEPIC BeadChip 850K and the results revealed a clear difference in the methylation status between the groups of patients allowing recognition of two main signaling pathways affected by DMRs (the MAPK kinase and TNF) that contributed to the increase in symptoms. So far, we found epigenomic variants associated with the onset of post-Covid immune diseases with a significant impact on the Brazilian population, as well as an distinct epigenomic profile showing a significant difference in the global methylation status between groups with severe and less affected patients infected by SARS-CoV-2 that may influence the expression of genes that contribute to the progression of COVID-19. Innovative epigenomic research strategies, different from traditional clinical research methods, are essential to achieve breakthroughs that help improve public health.

Session Title: Epigenetics Poster Session III

PB2263 Epigenetic signatures of atopic asthma in nasal epithelium from African ancestry populations in the CAAPA consortium

Authors:

R. Natarajan[1], M. Boorgula[2], M. Campbell[2], B. Szczesny[3], K. Kammers[4], I. Ruczinski[1], S. Chavan[2], C. Arehart[5], C. Cox[2], R. Johnson[2], I. Konigsberg[6], E. Thompson[7], A. Morin[7], C. McKennan[8], C. Figueiredo[9], C. Rotimi[10], R. Landis[11], H. Watson[12], N. Hansel[3], I. Yang[2], C. Olopade[13], C. Ober[7], A. Liu[14], CAAPA Consortium, E. Kenny[15], K. Barnes[2], R. Mathias[3], M. Taub[1]; [1]Johns Hopkins Univ., Bloomberg Sch. of Publ. Hlth., Baltimore, MD, [2]Univ. of Colorado, Anschutz Med. Campus, Aurora, CO, [3]Johns Hopkins Univ., Baltimore, MD, [4]Johns Hopkins Univ. Sch. of Med., Baltimore, MD, [5]Univ. of Colorado, Boulder, CO, [6]Univ. of Colorado, Denver, CO, [7]Univ. of Chicago, Chicago, IL, [8]Univ. of Pittsburgh, Pittsburgh, PA, [9]Federal Univ. of Bahia, Salvador, Brazil, [10]NIH, Woodstock, MD, [11]Univ. of the West Indies, Cave Hill, Barbados, [12]Univ. of the West Indies, Queen Elizabeth Hosp., Bridgetown, Barbados, [13]Dept. of Med., Univ. of Chicago, Chicago, IL, [14]Dept. of Pediatrics, Chidren's Hosp. Colorado and Univ. of Colorado Denver, Anschutz Med. Campus, Aurora, CO, [15]Icahn Sch. of Med. at Mount Sinai, New York, NY

Asthma is a heterogeneous, complex disease, with variable mechanisms and clinical manifestations. Understanding endotypes of asthma that result in phenotypic heterogeneity has been shown to be important, and here we use atopy-defined subgroups. Epigenetic signatures of disease may elucidate genetic and environmental mechanisms that impact gene regulation and asthma outcomes. We use nasal epithelial cell DNA methylation (DNAm) collected from self-identified African-ancestry subjects across four sites (Baltimore, Chicago, Denver, and Washington DC) in the Consortium on Asthma among African-ancestry Populations in the Americas (CAAPA). We aim to understand the etiology of asthma by identifying differentially methylated CpGs (DMCs) associated with active and atopic asthma. DNAm was measured using the Illumina EPIC array and a new asthma and allergy (A&A) array that targets allergy-related regions. Sample/probe QC was performed with the minfi R package. Associations were adjusted for age, sex, plate, site, 2 genetic principal components estimated from MEGA chip genotypes, and latent factors estimated by the CorrConf R package to correct for unwanted variation. DMCs had q-value < 0.05. Cases were limited to those with active asthma (N=149, defined by severity index > 0) and compared to never-asthma controls (N=182). We then performed a subgroup DMC analysis, where atopic, active asthma cases (N=128) were compared to non-atopic, never-asthma controls (N=75). Atopy was defined by phadiatop-measured, allergen-specific IgE >0.35 PAU/L or total IgE >100IU. CpGs were mapped to genes based on proximity, an external promoter-capture Hi-C dataset, and the results from an eQTM analysis associating DNAm to nasal epithelium gene expression from RNA-seq in these samples. On the EPIC array, 193 DMCs were associated with active asthma agnostic to atopy; there were 428 DMCs in the atopy subset. The A&A array detected 0 DMCs agnostic to atopy and 26 DMCs considering atopy. The 11 most significant active asthma DMCs (q<0.001) mapped to genes associated with T cell development, chemokine binding, and responsiveness to glucocorticoid receptors, including *CDHR3* and *FKBP5*. *CDHR3* is an epithelial cell receptor for RV-C and was associated with asthma exacerbations/hospitalizations in the first 5 years of life. Our eQTM analysis revealed that DNAm at *FKBP5*, which is also associated with increased inhaled corticosteroid use, may mediate the association between *FKBP5* nasal epithelium gene expression and asthma. Our analysis highlights the importance of considering specific phenotypic subtypes when working with heterogeneous phenotypes in epigenetic association studies.

Session Title: Epigenetics Poster Session I

PB2264 Epigenome-wide association study of exposure to arsenic provides evidence that exposure to arsenic causally impacts DNA methylation.

Authors:

J. Li[1], N. Jain[1], L. I. Tamayo[1], L. Tong[1], K. Demanelis[2], F. Jasmine[1], M. Kibriya[1], L. S. Chen[1], H. Ahsan[1], B. L. Pierce[1]; [1]Univ. of Chicago, Chicago, IL, [2]Univ. of Pittsburgh, Pittsburgh, PA

Introduction: Inorganic arsenic (As) is a carcinogen, and >200 million people worldwide are exposed to As through drinking water. Long-term exposure to As increases the risk of chronic diseases including cancer and cardiovascular disease. One proposed As toxicity mechanism is dysregulation of the epigenome.
Methods: We performed an epigenome-wide association study (EWAS) of exposure to As using data on 1,186 adults from the Health Effects of Arsenic Longitudinal Study in Bangladesh. DNA methylation (DNAm) was measured at ~850,000 CpGs in leukocytes using the Illumina EPIC array. We evaluated the association between As measured in urine (and in drinking water) with DNAm using linear regression models adjusting for covariates and surrogate variables representing cell type composition. After identifying As-associated CpGs, we performed Mendelian Randomization (MR) to assess the causal effect of As on these CpGs, using genetic variants associated with arsenic metabolism/elimination efficiency as instruments. Next, we build an epigenetic biomarker of exposure to As using an elastic net regression consisting of all As-associated CpGs (FDR<0.05). We then assessed the relationship between this epigenetic biomarker of exposure to As and arsenical skin lesion status, the common sign of arsenic toxicity (93 cases, 186 controls).
Results: In total, we identified 135 CpGs associated with log2-transformed creatinine-adjusted urinary As using a Bonferroni significance threshold. Urinary As-associated CpGs were enriched in non-CpG islands and shores and depleted in islands. Of the 135 CpGs, 116 were also associated with log2-transformed As concentrations measured in drinking water at an FDR of 0.05. Using these 116 CpGs in the MR analysis, we discovered that the vast majority of these CpGs (80%) were directionally consistent (one-sided binomial p=1.8E-11). We built an epigenetic biomarker of exposure to As using all CpGs that were associated with urinary As at an FDR<0.05 (1,177 CpGs) and covariates, and observed moderate correlation between epigenetically predicted and measured As levels (R^2=0.49). In an independent validation set, we observed a strong association between our epigenetic biomarker of exposure to As and skin lesion status (OR=1.7, p=1.5E-3, AUC=0.56).
Conclusion: Our study provides evidence that exposure to As causally impacts DNAm at many CpGs in the epigenome. We introduce an epigenetic biomarker of arsenic exposure that can potentially be used in contexts where it is difficult to directly assess exposure to As. Furthermore, our study suggests that impacts of arsenic exposure on the epigenome correlate with human health.

Session Title: Epigenetics Poster Session II

PB2265 Epigenome-wide association study of incident type 2 diabetes in Black and White participants from the Atherosclerosis Risk in Communities (ARIC) Study.

Authors:

S. Venkataraghavan[1], J. Pankow[2], J. Coresh[1], E. Boerwinkle[3], M. Fornage[4], E. Selvin[1], D. Ray[1]; [1]Johns Hopkins Univ., Baltimore, MD, [2]Univ Minnesota, Minneapolis, MN, [3]Univ. of Texas Hlth.Sci. Ctr. at Houston, Houston, TX, [4]Univ. of Texas Hlth.Sci. Ctr. at Houst, Houston, TX

DNA methylation (DNAm) in several genes are associated with type 2 diabetes (T2D) in populations outside the US. However, the link between DNAm and incident T2D has not been previously examined in US populations. We conducted a prospective cohort study of 2,091 Black and 1,029 White participants with mean age of 58 years in the Atherosclerosis Risk in Communities (ARIC) using Cox regression model to examine the association between ~480K CpG sites, measured using the Illumina450Karray, and incident T2D. Participants were aligned at baseline using the visit of DNAm measurement (Visit 2: 1990-92 or Visit 3: 1993-95) with mean follow-up of 17 years. At a Bonferroni-corrected epigenome-wide significance threshold of 10^{-7}, we identified 7 novel T2D-associated CpG sites: *GPX6* (cg02793507: HR= 0.85, p=2.7 x 10^{-8}) and (cg00647063: HR= 1.20, p=2.5 x 10^{-8}), *C1orf151* (cg05380846: HR= 0.89, p=8.4 x 10^{-12}), *ZNF2* (cg01585592: HR= 0.88, p=1.6 x 10^{-9}), *JPH3* (cg16696007: HR= 0.87, p=7.8 x 10^{-9}), chr11p15 (cg13738793; HR= 1.11, p=7.7 x 10^{-8}) and chr17q25 (cg16865890: HR= 0.8, p=6.9 x 10^{-9}). The CpG sites at *GPX6*, *C1orf151*, *ZNF2*, and *JPH3* were identified in Black adults, chr17q25 was identified in White adults alone, and chr11p15 was identified in the meta-analysis of the 2 groups. After adjustment of BMI as a continuous covariate in the Cox regression model, there was minimal change in effect size of CpG sites at *JPH3* (HR=0.87) and *GPX6* (HR=0.85) which remained significant after adjustment. The CpG site at *JPH3* (HR=0.87) was also significant after adjustment for fasting glucose levels with minimal effect size change. In vivo studies have identified correlations between *JPH3* expression and GPx protein levels with insulin. Our analysis replicated known CpG site associations at significance level of 10^{-7} in one or more groups including cg19693031 at *TXNIP*, cg00574958 at *CPT1A*, cg16567056 at *PLBC2*, cg11024682 at *SREBF1*, cg08857797 at *VPS25*, and cg06500161 at *ABCG1*. We used differentially methylated region analysis to examine if groups of proximal CpG sites within gene-regions of promoter, gene-body, and 3' UTR were associated with incident T2D. We found promoter regions of *TP63* to be differentially methylated across all groups. A previous study showed TAp63, an isoform of *TP63*, knockout mice developed insulin resistance and glucose tolerance. Differences in DNAm levels identified across groups could reflect differences in statistical power, differences in environmental exposures and/or allele frequencies at variants controlling DNAm levels. Our study demonstrated that including participants from diverse populations can help discover yet-unidentified DNAm-T2D associations.

Session Title: Epigenetics Poster Session III

PB2266 Epigenome-wide association study of perceived discrimination in the Multi-Ethnic Study of Atherosclerosis (MESA)

Authors:

W. Zhao[1,2], L. Lin[2], K. Kelly[3], L. Opsasnick[2], B. Needham[2], Y. Liu[4], A. Diez Roux[5], S. Sen[6], J. A. Smith[2,1]; [1]Inst. for Social Res., Univ. of Michigan, Ann Arbor, MI, [2]Sch. of Publ. Hlth., Univ. of Michigan, Ann Arbor, MI, [3]Inst. for Behavioral Genetics, Univ. of Colorado, Boulder, CO, [4]Duke Univ. Med. Ctr., Durham, NC, [5]Drexel Dornsife Sch. of Publ. Hlth., Philadelphia, PA, [6]Michigan NeuroSci. Inst., Univ. of Michigan, Ann Arbor, MI

Background: Perceived discrimination, recognized as a chronic psychosocial stressor, has various negative consequences on health. DNA methylation (DNAm) may be a potential mechanism by which stressors get embedded into the human body at the molecular level and subsequently affect health outcomes. However, relatively little is known about the effect of perceived discrimination on DNAm. This study aimed to identify the DNAm sites across the genome that are associated with discrimination in a multi-ethnic study. **Methods:** We conducted epigenome-wide association analyses (EWAS) of three discrimination measures (everyday discrimination, race-related major discrimination, and non race-related major discrimination) in 1,151participants (565 self-reported non-Hispanic European American, 365 self-reported Non-Hispanic African American, and 221 self-reported Hispanic) from the Multi-Ethnic Study of Atherosclerosis (MESA). DNAm in purified monocytes was measured using Illumina HumanMethylation450 BeadChip. The primary analyses were adjusted for age, gender, smoking, study site, cell contamination, genetic principal components, and batch effects (random effect). Comb-p was subsequently applied to EWAS results to identify differentially methylated regions (DMR). We conducted both race/ethnicity-stratified analyses as well as trans-ethnic meta-analyses. For the identified CpGs/DMRs, we further evaluated their cis-effects on RNA transcripts or potential mediation effects on health outcomes related to hypertension and diabetes. **Result:** At false discovery rate of 10%, three CpGs, three DMRs (*ASPG*, *DUSP5*, *IL17C*), and one CpG were associated with everyday discrimination in African Americans, Hispanics, and meta-analysis respectively. One CpG was associated with race-related major discrimination in African Americans. For non race-related major discrimination, one CpG was associated in Hispanics whereas one CpG and one DMR (*GLI3*) were associated in meta-analysis. All associations remained substantially similar after adjusting for potential confounding/mediating factors (body mass index, alcohol consumption, educational attainment, and depression). Some of the identified CpGs have been previously reported to be associated with aging, inflammatory diseases, or tumor cells. One CpG, cg14656411, was associated with overexpression of *NDUFS5* and *AK1RIN1*. No mediation effects on health outcomes were identified. **Conclusion:** Our study demonstrated the potential influence of discrimination on DNA methylation and subsequent gene expression. However, the potential impact on health outcomes remains to be determined.

Session Title: Epigenetics Poster Session I

PB2267 Epigenomic profiling of the infrapatellar fat pad in osteoarthritis

Authors:

P. Kreitmaier[1,2,3], Y-C. Park[1], D. Swift[4], A. Gilly[1], J. Wilkinson[4], E. Zeggini[1,2]; [1]Helmholtz Zentrum Muenchen, Neuherberg, Germany, [2]Technical Univ. of Munich (TUM) and Klinikum Rechts der Isar, Munich, Germany, [3]Graduate Sch. of Experimental Med., TUM Sch. of Med., Technical Univ. of Munich, Munich, Germany, [4]Dept. of Oncology and Metabolism, The Univ. of Sheffield, Sheffield, United Kingdom

Osteoarthritis is a complex joint disease that affects more than 300 million people worldwide. Due to increasingly aging populations, the impact of osteoarthritis on public health systems is estimated to increase further. Current treatment methods are limited to pain management and joint replacement, underlining the need to develop novel, personalised treatment strategies. Thus, it is important to extend our knowledge about the genetic and genomic basis of osteoarthritis. Here, we investigate matched genotype and methylation profiles from infrapatellar fat pad as well as blood samples from 70 knee osteoarthritis patients undergoing joint replacement surgery. We perform an epigenome-wide association study to examine epigenetic differences between fat pad and blood methylation. We generate a genome-wide cis-methylation QTL (mQTL) map in fat pad. Furthermore, we combine this mQTL map with results from genome-wide association studies for osteoarthritis using colocalisation and two-sample Mendelian randomization to identify epigenetic mechanisms mediating the effects of genotype on osteoarthritis risk in fat pad. Comparing the DNA methylation profiles between fat pad and blood reveal 84,973 differentially methylated sites ($p < 6.4x10^{-08}$). We produce the first genome-wide methylation quantitative trait locus (mQTL) map of fat pad and identify 35,948 mQTL targeted methylation sites. Using colocalisation and Mendelian randomization analyses, we resolve eleven osteoarthritis GWAS signals and provide insights into the molecular mechanisms underpinning osteoarthritis aetiopathology. Together, we report widespread differences between fat pad and blood methylation, generate the first mQTL map in fat pad and reveal methylation sites with a putative causal role in osteoarthritis. Our findings provide the first view of the epigenetic landscape of infrapatellar fat pad primary tissue in osteoarthritis.

Session Title: Epigenetics Poster Session II

PB2268 Epigenomic signature of major congenital heart defects in newborns with Down syndrome.

Authors:

J. Mouat[1], S. Li[2], S. Myint[2], B. I. Laufer[1], P. J. Lupo[3], J. M. Schraw[3], J. P. Woodhouse[3], A. J. de Smith[2], J. M. Lasalle[1]; [1]Univ. of California Davis, Davis, CA, [2]Univ. of Southern California, Los Angeles, CA, [3]Baylor Coll. of Med., Houston, TX

Background: Congenital heart defects **(CHDs)** affect approximately half of individuals with Down syndrome **(DS)** but the molecular reasons for incomplete penetrance are unknown. Previous studies have largely focused on identifying genetic risk factors associated with CHDs in individuals with DS, but comprehensive studies of the contribution of epigenetic marks are lacking. We aimed to identify and characterize DNA methylation differences from newborn dried blood spots **(NDBS)** of DS individuals with major CHDs compared to DS individuals without CHDs.
Methods: We used the Illumina EPIC array and whole-genome bisulfite sequencing **(WGBS)** to quantitate DNA methylation for 86 NDBS samples from the California Biobank Program: 1) 45 DS-CHD (27 female, 18 male) and 2) 41 DS non-CHD (27 female, 14 male). We analyzed global CpG methylation and identified differentially methylated regions **(DMRs)** in DS-CHD vs DS non-CHD comparisons (both sex-combined and sex-stratified) corrected for sex, age of blood collection, and cell type proportions. CHD DMRs were analyzed for enrichment in CpG and genic contexts, chromatin states, and histone modifications by genomic coordinates and for gene ontology enrichment by gene mapping. DMRs were also tested in a replication dataset and compared to methylation levels in DS vs typical development **(TD)** WGBS NDBS samples.
Results: We found global CpG hypomethylation in DS-CHD males compared to DS non-CHD males, which was attributable to elevated levels of nucleated red blood cells and not seen in females. At a regional level, we identified 58, 341, and 3,938 CHD-associated DMRs in the Sex Combined, Females Only, and Males Only groups, respectively, and used machine learning algorithms to select 19 Males Only loci that could distinguish CHD from non-CHD. DMRs in all comparisons were enriched for gene exons, CpG islands, and bivalent chromatin and mapped to genes enriched for terms related to cardiac and immune functions. Lastly, a greater percentage of CHD-associated DMRs than background regions were differentially methylated in DS vs TD samples.
Conclusions: A sex-specific signature of DNA methylation was detected in NDBS of DS-CHD compared to DS non-CHD individuals. This supports the hypothesis that epigenetics can reflect the variability of phenotypes in DS, particularly CHDs.

Session Title: Epigenetics Poster Session III

PB2269 Evaluating Hi-C data using deep neural networks.

Authors:

M. Benton, A. Rodriguez Perez; Baylor Univ., Waco, TX

Background: The three-dimensional (3D) chromatin architecture of the human genome is vital for maintaining appropriate levels of gene expression. Disruptions to this structure can change the existing enhancer-promoter interactions and have been associated with a wide range of disease phenotypes, including cancers and severe developmental and neurological disorders. Thus, we must be able to evaluate the quality of existing Hi-C datasets, which map 3D chromatin organization and its interactions, and create accurate methods for comparing Hi-C matrices genome-wide. Current correlation-based metrics for comparing Hi-C interactions do not fully capture known structure in these data, while annotation-informed approaches rely on additional external datasets. Here, we present a novel data-driven approach to evaluate the reproducibility of Hi-C data using neural networks to predict whether two portions of Hi-C contact matrices correspond to biological replicates.
Methods: Neural networks are a strong function approximation model that have proven useful in a wide range of problems, including chromatin conformation prediction. Here, we employ Siamese networks, which have been used extensively for similarity problems. We consider several evaluation scenarios that test the ability of the models to identify biological replicates, evaluate how well the task generalizes across different cell types, and assess the differences in the quality of the results across the genome.
Results: Through our experiments, we show that models can learn to predict the desired output when trained and tested in multiple cell types simultaneously. Across 7 different cell types, we achieve accuracy ranging from 91.7-95.9% and find that the effectiveness of the models is largely stable across chromosomes. However, training the models with data from multiple cell types and testing on a novel one produces models with lower and more variable accuracy (46.0%-73.8%). This suggests that the model does not learn broader patterns when trained on cross-tissue data, and that patterns are not transferable across cell types.
Conclusions: Our preliminary results pave the way for new avenues of research in understanding spatial genome organization and dynamics using data-driven approaches. Using the power of neural networks and the vast amount of Hi-C data available, we can develop more robust reproducibility measurements and gain deeper insights into the complex mechanisms underlying chromatin organization.

Session Title: Epigenetics Poster Session I

PB2270 Evaluation of X chromosome inactivation escape heritability across human tissues and cell type.

Authors:

H. Markus[1,2], B. Jiang[3], L. Carrel[4], D. Liu[3,2]; [1]Penn State Coll. of Med., Hershey, PA, [2]Pennsylvania State Univ., University Park, PA, [3]Pennsylvania State Univ. Coll. of Med., Hershey, PA, [4]Penn State Univ Col Med, Hershey, PA

X-chromosome inactivation (XCI) silences one X in females to balance gene dosage with males. Yet up to 10% of genes consistently escape XCI and 15-30% of genes variably escape XCI in a subset of tissues or individuals. A few studies suggest possible genetic influences on variable XCI escape, yet its extent remains to be quantified. Creating a comprehensive landscape of XCI states across tissue/cell types requires improved methods to infer XCI escape from population scale RNASeq data. Two types of methods exist to identify escape genes. As XCI escape leads to increased expression level in females, male and female differential expression (DGE) allows identifying genes that escape XCI among many individuals but cannot determine whether the gene escapes XCI in a given sample. On the other hand, as XCI escape yields more balanced allelic expression, methods that compare allelic imbalance (AI) between escape and sample skewing (i.e., AI calculated from inactivated genes) through allele specific expression (ASE) can identify gene escape in a given individual. ASE-based methods are underpowered for samples with balanced skewing and needs heterozygous SNPs within transcribed regions. To improve existing approaches, we propose a novel statistical method that combines the strength of male-female DGE and AI to infer XCI states across all individuals irrespective of XCI skewing or heterozygous SNP in bulk RNA-seq data. In simulations and empirical estimations in the GEUVADIS dataset, our method has well controlled type 1 error and increased power to infer XCI states in moderate to balanced XCI skew samples (i.e., skewing ≤75:25) when compared to XCIR, an ASE-based method. We then assessed XCI states in female samples from the Genotype Tissue Expression (GTEx) Project for an average of 202 genes per tissue (range 19-356) across 42 tissues. Genes inferred as escape using our method were concordant with previous consensus XCI states. Together we infer XCI states of 22 genes not previously described and identify 187 previously silenced genes that escape or variably escape in at least one tissue. Lastly, we analyze XCI escape status as a trait and quantify its heritability of each variably escape gene. In total, we identified and replicated 108 genes with significant XCI escape heritability in at least one tissue (average heritability 0.19, range 0.015-1.13). Overall, our method enabled us to provide a comprehensive atlas of XCI escape genes across human tissues and quantify QTLs that influence XCI escape status. These results will increase our understanding of the mechanisms of XCI escape genes and sex-biased diseases.

Session Title: Epigenetics Poster Session II

PB2271 EWAS of Gulf War Illness in the Gulf War Era Cohort and Biorepository (GWECB)

Authors:

X. Qin[1,2], Y. Shi[3], S. Boyle[1], A. Stone[4], W. Carter[4], L. Griffin[4], S. Pyarajan[3], B. Despard[3], K. Markianos[3], J. Vahey[2,1], M. Johnson[1], A. Press[1], J. Upchurch[1], S. Petry[1], A. Thompson[1], T. Redding IV[1], E. Gifford[2,1], K. Sims[1], C. Williams[1], E. Hauser[2,1]; [1]Durham VA Hlth.Care System, Durham, NC, [2]Duke Univ., Durham, NC, [3]VA Boston Hlth.care System, Jamaica Plain, MA, [4]Central Arkansas Veterans Hlth.Care System, Little Rock, AR

Many Veterans who were deployed to the Persian Gulf during the Gulf War report a constellation of pervasive chronic symptoms such as alterations in mood and cognition, fatigue, and musculoskeletal problems that have become to be known as Gulf War Illness (GWI).The etiology of GWI is unknown, although there is evidence that some military exposures are risk factors for GWI. There are no biomarkers for GWI and treatments are very limited. A previous GWAS performed in this dataset suggested significant heritability of the CDC Severe GWI (CSGWI) case definition and genetic variants of modest effect were associated with CSGWI. The goal of this research is to explore the epigenetic landscape for CSGWI. The GWECB consists of DNA and survey data from 1275 individuals who served in the US Armed Forces in 1990-1991. Self-reported symptoms captured by the GWECB survey determined CSGWI case status. Methylation data were generated from the Illumina EPIC array with Illumina quality control (QC) performed at the Pharmacogenomics Analysis Laboratory. The analysis pipeline developed for the Million Veteran Program, incorporating Sesame, minfi, meffill, limma, SVA and implemented in GenISIS was used for additional sample, probe and batch QC, normalization and Housman cell type estimates. EWAS was conducted using linear regression of the log2 methylated/unmethylated ratio with CSGWI as the independent variable adjusted for covariates of age, sex, blood cell type estimates and 6 genetic PCs to adjust for ancestry. CpG sites were annotated and ranked by p-value from the CSGWI coefficient. The EWAS Atlas (https://ngdc.cncb.ac.cn/ewas/atlas) was searched for published associations of CpG sites and genes. After final QC, EWAS was completed for 685,395 probes in 1018 samples. Of these, 223 (21.9%) met criteria for CSGWI with 77% male and 35% self-reported non-European ancestry. Population stratification was reasonably well-controlled with QQ plot lambda of 1.07. Eight CpG probes had p-values in the suggestive ($p< 1 \times 10^{-5}$) range. The top annotated CpG results were in genes TMEM37 ($p=1.44 \times 10^{-6}$), SNX3 ($p=1.77 \times 10^{-6}$), SYT2($p=4.17 \times 10^{-6}$), LIPT1 ($p=4.28 \times 10^{-6}$), TLE3 ($p=5.49 \times 10^{-6}$), and MYL12A ($p=7.28 \times 10^{-6}$). In the EWAS Atlas, published EWAS associations with these genes included Gulf War Illness (SNX3, SYT2) and ME/CFS (TMEM37, SYT2, TLE3). We have completed the largest genome-wide epigenetic analysis of CSGWI to date which provides evidence for differential methylation between CSGWI cases and controls. The EWAS results indicate that larger sample sizes are needed for adequate statistical power but also that epigenetic analyses may be useful in identifying biological associations in GWI.

Session Title: Epigenetics Poster Session III

PB2272 Exploring the Relationship Between Epigenetic Age Acceleration and Carotid Plaque Presence in a Multi-Generational Family Study of Dominicans.

Authors:

S. Parmar[1], N. Dueker[2], H. Gardener[1], D. Cabral[3], L. Wang[3], T. Rundek[4], S. Blanton[3]; [1]Univ. of Miami Miller Sch. of Med. - - Miami, FL, MIAMI, FL, [2]Univ. of Miami Miller Sch. of Med., Miami, FL, [3]Univ. of Miami, Miami, FL, [4]Univ Miami, Miami, FL

Chronological age (CA) is a major risk factor for common diseases including hypertension, cancer, dementia, and stroke. Epigenetic clocks based on methylation patterns have been developed to estimate an individual's biological age, providing insight into lifestyles and exposures which may modulate these disease risks; this biological age provides a holistic assessment of an individual's health status. These clocks include the first-generation Horvath and Hannum, as well as second-generation GrimAge and PhenoAge. When biological age is regressed onto CA, an age acceleration (AA) measure is obtained; a positive value (pAA) indicates a person's biological age exceeds their CA with a negative value (nAA) indicating a biological age younger than their CA. Previous studies have found AA measures to be heritable, with estimates ranging from 21% to 64%. Furthermore, pAA has been associated with increased risk for cardiovascular disease (CVD) and carotid traits including carotid intima-media thickness and carotid plaque. However, these studies involved primarily European or African ancestry samples, and the generalizability of these clocks is unknown. Therefore, the present study aimed to estimate the heritability of various AA measures and to determine if AA is associated with carotid plaque presence in a sample of extended families from the Dominican Republic (DR). To accomplish this, DNA methylation in blood was assayed using the Illumina Infinium Human MethylationEPIC BeadChip in 61 extended DR families (n=792) for whom carotid plaque presence data was available. Heritability estimates for HannumAA, PhenoAA, GrimAA and IEAA (Horvath) were obtained using SOLAR, and t-tests were performed to determine if AA measures were associated with carotid plaque presence (y/n). In our sample, the mean CA was 46 y (range 18-94 y) and mean AA ranged from 0.03 ± 3.55y for HannumAA to 1.26 ± 5.94y for PhenoAA. All AA measures were heritable: PhenoAA (h^2=0.51, P=5.0x10^{-13}), GrimAA (h^2=0.51, P=5.8x10^{-16}), HannumAA (h^2=0.47, P=5.0x10^{-14}), and IEAA (h^2=0.47, P=1.8x10^{-16}). Additionally, we observed an increase in the mean PhenoAA in those with carotid plaque presence vs. those without (P=1.9x10^{-06}). No other AA measures were associated with carotid plaque presence. Mixed model analyses with covariate adjustments are underway to formally quantify the association between PhenoAA and carotid plaque presence. In conclusion, our study provides suggestive evidence for an association between PhenoAA and carotid plaque presence, highlighting the importance of PhenoAA in our Dominican sample and demonstrates the generalizability of these clocks across race/ethnic groups.

Session Title: Epigenetics Poster Session I

PB2273 † Extensive profiling of transcription factors in postmortem brains defines genomic occupancy in disease relevant cell types and links TF activities to neuropsychiatric disorders

Authors:

J. Loupe[1], A. G. anderson[1], L. Rizzardi[1], I. R. Nunez[1], B. Moyers[1], K. Trausch-Lowther[1], W. E. Bunney[2], P. Cartegena[2], A. Sequeira[2], S. J. Watson[3], H. Akil[3], G. M. Cooper[1], R. M. Myers[1]; [1]HudsonAlpha Inst. for Biotechnology, Huntsville, AL, [2]Univ. of California, Irvine, CA, [3]Univ. of Michigan, Ann Arbor, MI

Transcription factors (TFs) orchestrate gene expression programs crucial for cell physiology by binding to specific genomic loci and regulating nearby target genes. However, our knowledge of their function is limited, as the vast majority of the thousands of genomic occupancy maps available for human TFs are derived from a small set of immortalized cell lines that are biologically distinct from the cells and tissues relevant for many diseases. Particularly in the context of neurological diseases, few TFs have been studied in human brain tissues and cell-types, like neurons. To address this gap, we have generated a multi-omic resource highlighted by ChIP-seq for more than 100 TFs, (1,072 total experiments) using bulk tissues and NeuN/Olig2-sorted nuclei from multiple human post-mortem brain regions. Nearly a third of the TF-bound regions we have identified lack any annotation in the Registry of candidate cis-Regulatory Elements (cCREs) from the ENCODE Consortium and are enriched for neuronal pathways, highlighting the shortage of brain tissue in existing datasets. By uniformly profiling a large number of TFs, we are able to demonstrate unique occupancy profiles for each TF and how they change across brain regions and cell types. We identified regions of high TF occupancy and demonstrate improved measurements of TF activity upon their removal, including motif recognition and gene expression modeling. Further, we find that computational methods to predict TF binding demonstrate a bias for high occupancy sites, and miss many distal enhancer elements bound by subsets of TFs that are enriched for neuronal pathways. Neuronal TFs SATB2 and TBR1, for example, bind unique elements that regulate neuron-specific gene expression. Further, we find several TFs with far greater enrichment for GWAS loci associated with brain disorders compared to commonly assessed epigenetic measurements such as histone marks and ATAC-seq. The highest enrichment for neuropsychiatric disease variants was found in ChIP-seq performed on neuron-enriched NeuN+ sorted nuclei, including TBR1 and PKNOX1, while enrichment for neurodegenerative variants was found in TF binding sites identified in microglia-enriched $NeuN^-/Olig2^-$ nuclei. This dataset is one of the largest collections of TF binding maps in human tissues to date, and the only such resource for human brain. It will be a unique and powerful resource for future studies seeking to understand the role of TFs in epigenetic regulation in neurological disease.

Session Title: Epigenetics Poster Session II

PB2274 Functional characterization and network analysis of regulatory regions using genome-wide functional screens and chromatin interactions

Authors:

K-Y. Ko[1], K. Siklenka[2], A. Barrera[3], G. Yardimci[4], E. Aiden[5], T. Reddy[2]; [1]Duke, Durham, NC, [2]Duke Univ., Durham, NC, [3]Duke Univ., Durhan, NC, [4]Knight Cancer Inst., Portland, OR, [5]Baylor Coll. Med., Houston, TX

Objective

Establishing non-coding mechanisms for disease remains a major goal in human genetics. Identifying the genomic mechanisms that contribute to gene regulation is critical to achieving that goal. Two promising technologies for identifying those mechanisms are high-throughput reporter assays and CRISPR screens. The goal of this study is to test a hypothesis that high-throughput reporter assays detect regulatory element activity independent of chromatin context, while CRISPR screens identify regulatory elements that are active in the studied cell model.

Method

Phase 4 of the ENCODE project included a coordinated effort across eight labs to functionally characterize human gene regulatory elements using high-throughput reporter assay screens and high-throughput CRISPR perturbational screens. Together, ENCODE completed hundreds of such screens. Here, we compared the ability of the different functional characterization technologies used in ENCODE to identify genomic sites with transcription factor binding, histone modifications associated with active gene regulation, and chromatin interactions in the same cell models.

Results

Overall, we found substantial concordance across functional characterization assays, but also considerable assay-specific preferences in the characteristics of regulatory elements detected. While we identified enhancers such as GATA1 and MYC in multiple assays, we observed a mix of concordant and discordant signals in regions such as FADS and HBE1 loci. Our results suggested that variations in regulatory activities among the assays may be due to the differences of genomic features across the regions.

While functional annotations such as promoter/enhancer-like signatures were enriched in regulatory regions, some genomic features could partially explain the variations among the assays. TFs such as EP300 and AP1 were more enriched in ATAC-STARR-seq and TFs such as ELK1 and SP1 were more enriched in whole genome STARR-seq and MPRA. CTCF binding had lower enrichment scores than other TFs in STARR-seq and MPRA. However, several regions such as FADS loci were found not only had discordant signals across CRISPRi and reporter assays, but also contained CTCF and cohesin binding sites. Therefore, we hypothesized that chromatin organization may contribute to the differences among the assays.

To test the hypothesis, we constructed chromatin interaction networks and found CRISPR-based regulatory regions were more enriched at loop anchors than regions in other assays. Taken together, we proposed that CRISPR screens are more sensitive to chromatin spatial organization compared to reporter assays such as STARR-seq and MPRA.

Session Title: Epigenetics Poster Session III

PB2275 † Functional characterization of the mammalian genome.

Authors:

K. Siklenka[1], J. Zhang[2], D. Shigaki[3], K-Y. Ko[4], J. Oh[3], R. Venukuttan[1], A. Barrera[1], ENCODE Functional Characterization Center, M. Beer[5], E. Aiden[6], H. Yu[7], T. Reddy[1], G. Yardimci[8]; [1]Duke Univ., Durham, NC, [2]Cornell Univ., Ithica, NY, [3]Johns Hopkins, Baltimore, MD, [4]Duke, Durham, NC, [5]Johns Hopkins Univ, Baltimore, MD, [6]Baylor Coll. Med., Houston, TX, [7]Cornell Univ, Ithaca, NY, [8]Oregon Hlth.and Sci. Univ., Portland, OR

The ENCODE consortium has mapped millions of elements associated with regulatory potential across the human and mouse genomes. To assess the functionality of those elements in different genomic contexts, we systematically measured activity and essentiality of human and mouse regulatory elements across hundreds of complementary functional characterization experiments. The assays included genome-wide studies and coordinated high-resolution studies focused on select loci. Together, the studies comprise tens of billions of DNA fragments assayed in high-throughput reporter assays, millions of CRISPR-based perturbations, and thousands of in vivo whole-mouse transgenic reporter assays.

We first developed analysis methods to uniformly report similarities and differences between the regulatory element activity detected from these assays. Overall, 30% of tested open chromatin regions have detectable activity in high-throughput reporter assays, and 5% have a functional readout with CRISPR interference.

We observe that while a subset of high confidence regulatory regions are detectable across all assay types, there are considerable differences in the types of elements each assay observes. We reveal sequence specific differences between high-throughput reporter assay classes, with activity from MPRA explained by promoter-like motifs, while activity from STARR preferring distal. Moreover, we report similar patterns of enrichments of transcription factors and histone modifications at assay-specific active regions as measured by ChIP-seq.

Meanwhile, we observe 3D chromatin organization accounts for major differences in detection between reporter assays and CRISPR screens. CRISPR-based active-regions are enriched with chromosomal contacts, and constrained to the topological domain of the gene target, an observation not observed in reporter assays

Taken together, these results suggest a hierarchy of regulatory function in the human genome and provide a framework for integrative analysis and interpretation of those assays moving forward. In that model, DNA sequence features define a subset of open chromatin sites that have cell- and condition-specific regulatory activity in reporter assays; and the chromatin context further restricts regulatory function as detected with CRISPR-based perturbations.

Session Title: Epigenetics Poster Session I

PB2276 Further refinement of the differentially methylated distant lung-specific *FOXF1* enhancer in a neonate with alveolar capillary dysplasia

Authors:

P. Szafranski[1], R. Garimella[2], H. Mani[3], R. Hartman[4], G. deutsch[5], A. Silk[6], A. Benheim[7], P. Stankiewicz[1]; [1]Baylor Coll. Med., Houston, TX, [2]Dept. of Pediatrics, Inova LJ Murphy Children's Hosp., Falls Church, VA, [3]Dept. of Pathology, Inova Fairfax Hosp., Falls Church, VA, [4]Inova Dept. of Genetics, Inova Fairfax Med. Campus, Falls Church, VA, [5]Univ. of Washington Sch. of Med., Seattle, WA, [6]Neonatology, Fairfax Neonatology Associates, Inova Fair Oaks Hosp., Inova LJ Murphy, Children's Hosp., Fairfax, VA, [7]Div. of Pediatric Cardiology, Inova LJ Murphy Children's Hosp., Falls Church, VA

Heterozygous loss-of-function of the mesenchymal transcription factor *FOXF1* gene has been found in 80-90% of neonates with histologically-verified alveolar capillary dysplasia with misalignment of pulmonary veins (ACDMPV, MIM 265380), a lethal lung developmental disorder. Analyses of 34 ACDMPV-causative overlapping CNV deletions leaving *FOXF1* intact enabled us to define the distant ~ 60 kb lung-specific enhancer region (chr16:86,212,040-86,271,919, hg19) mapping ~ 286 kb upstream to *FOXF1* and harboring a LINE & *Alu* genomic instability hotspot. We previously reported that transcription of not only *FOXF1* but also of the neighboring and inversely oriented lncRNA gene *FENDRR* is strongly regulated by this enhancer. Intriguingly, *Fendrr*$^{-/-}$ mice developed lethal heart defects and hypoplastic lungs, similar to those found in *Foxf1*$^{+/-}$ mice. Recently, we proposed a bimodal structure and parental functional dimorphism of this *FOXF1* & *FENDRR* enhancer, with Unit 1 having a higher activity on the paternal chr16 and non-overlapping Unit 2 on the maternal chr16 (PMID: 36157490). Unit 1 harbors an evolutionarily ultra-conserved ~ 660 bp segment and encodes two antisense long non-coding RNAs, whereas Unit 2 features lung-specific histone H3 modifications characteristic for an active enhancer. Interestingly, two regions of differential methylation and allelic differences of the H3K27Ac profile were also identified in Unit 2. Here, we describe a novel unusual sized pathogenic CNV deletion involving a centromeric portion of the *FOXF1* enhancer on maternal chr16 that narrows Unit 2 to an essential ~ 9-kb segment. This region harbors the enhancer-promoter interaction site and the binding sites for TFs involved in chromatin modification and/or regulation of cell cycle and motility. Using a restrictase-based assay, we found that this segment is weakly methylated at ApT adenine, with twice the frequency of methylation on the maternal versus paternal chr16. This difference might contribute to the proposed higher activity of the enhancer Unit 2 on maternal compared to the paternal chr16. Our data provide further insight into structure and function of the lung-specific *FOXF1* & *FENDRR* enhancer.

Session Title: Epigenetics Poster Session II

PB2277 Genetics and Epigenetics of Alcohol-Associated Liver Disease: Analysis of Single Nucleotide Variants and DNA Methylation change.

Authors:

B. Shankarappa[1,2], J. Mahadevan[2], P. Murthy[2], M. Purushottam[2], B. Viswanath[2], H. Devarbhavi[1], A. Mysore[1]; [1]St Johns Med. Coll. Hosp., Bangalore, India, [2]Natl. Inst. of Mental Hlth.and NeuroSci.s, Bangalore, India

Introduction:To understand whether there is an underlying genetic signature for predisposition to Alcohol Use Disorder with cirrhosis (AUDC+ve) we studied single nucleotide polymorphism (SNPs) & DNA methylation change in genes involved in alcohol metabolism (ADH2, ADH3 and ALDH2), lipid metabolism (PNPLA3 and TM6SF2) and one carbon metabolism pathway (MTHFR and MTR). Alcohol has a toxic effect on all organs, and this may underlie disease syndromes such as cirrhosis, neuropathy, and neuropsychiatric manifestations. This study will help us to understand how genetic factors may intersect with methylation and increase risk of disease syndromes.

Methods:Study included men with AUDC+ve, N=136 and AUD without liver cirrhosis, N=107 based on ICD10 criteria, drawn from the clinical services of St.John's Medical College Hospital (Gastroenterology and Psychiatry) & NIMHANS. Fibroscan and/or sonographic findings were used to rule out fibrosis (Liver Stiffness Measurement, LSM <7kPa).PCR-RFLP was performed to study the SNPs. The SNP-based genetic risk score was calculated for all the SNPs studied. For DNA methylation, DNA was bisulfite converted followed by PCR for the LINE-1, ALDH2, MTHFR, PNPLA3 and TM6SF2 loci and Pyrosequencing (PyroMark Q24).Results:AUDC+ve group had consumed lesser quantities of alcohol for a shorter duration and had a later age of onset of drinking compared to AUDC-ve group. The AUDC+ve had higher levels of conjugated and direct bilirubin, AST/ALT ratio, ALP and GGT, also lower levels of total protein, albumin, Haemoglobin levels and MCV. We observed a global DNA hypomethylation (P<0.01) and ALDH2 hypomethylation (P 0.03) in AUDC+ve compared to AUDC-ve subjects. ALDH2 hypomethylation showed significant association with MELD score (cirrhosis severity). TM6SF2 loci showed lower methylation in AUD-C+ve group compared to AUD-C-ve group (P=0.02). Genetic Risk Score was significantly higher in AUDC+ve group compared to AUDC-ve group and control data (SAS population) (P = 0.03).

Conclusion and Discussion:Present study, we found DNA hypomethylation at LINE1 loci in AUDC+ve. Thus, a dip in global DNA methylation, following chronic, heavy alcohol use, may occur in patients who eventually progress towards HCC. ALDH2 hypomethylation might lead to mitochondrial damage, hepatic inflammation, and fibrosis due to the increased accumulation of acetaldehyde. TM6SF2 hypomethylation may affect it's transcription activity and leads to increased triglyceride accumulation in the liver.Altered methylation in blood DNA at candidate loci may serve as a biomarker for severe liver disease in the context of prolonged severe alcohol abuse.

Session Title: Epigenetics Poster Session III

PB2278 Genome-wide association analysis identified a novel locus for epigenetic age acceleration

Authors:

R. Noroozi[1,2], J. Rudnicka[3], A. Pisarek[3], B. Wysocka[4], A. Masny[4], M. Boroń[4], K. Migacz-Gruszka[5], P. Pruszkowska-Przybylska[6], M. Kobus[7], D. Lisman[8], A. Sitek[9], A. Ossowski[8], M. Spólnicka[10], W. Branicki[11], E. Pośpiech[3,7]; [1]Johns Hopkins Univ., Baltimore, MD, [2]Jagiellonian Univ., Krakow, Poland, [3]Malopolska Ctr. of Biotechnology, Jagiellonian Univ., Krakow, Poland, [4]Central Forensic Lab. of the Police, Warsaw, Poland, [5]Jagiellonian Univ. Med. Univ., Krakow, Poland, [6]Dept. of Anthropology, Univ. of Łódź, Łódź, Poland, [7]Inst. of Biological Sci., Faculty of Biology and Environmental Sci. Cardinal Stefan Wyszynski Univ., Warsaw, Poland, [8]Dept. of Forensic Genetics, Pomeranian Med. Univ., Szczecin, Poland, [9]Dept. of Anthropology, Univ., Łódź, Poland, [10]Ctr. for Forensic Sci. Univ., Warsaw, Poland, [11]Inst. of Zoology and BioMed. Res. of the Jagiellonian Univ., Krakow, Poland

The DNAm-based age estimators, also known as DNAm clocks can consider various aspects of the aging process such as health span, mortality risk, and the pace of aging. In addition, the epigenetic age acceleration (EAA) measure which is the discrepancy between DNAm age and chronological age, has emerged as a potential clinical biomarker in association with various diseases. Besides, previous GWAS analysis has emphasized the importance of genetic factors in epigenetic age acceleration and implies that genetic variations may contribute to inter-individual differences in epigenetic aging. Here we performed GWAS analysis on blood samples obtained from the Polish population to further investigate genetic variants associated with different EAA measures. The DNA methylation and genomic profile of samples were obtained using the Methylation EPIC array and the Global Screening Array (GSA). (Illumina, San Diego, CA, USA). DNAm ages were calculated for different DNAm clocks including Horvath2013, Hannum, Skin&Blood, PhenoAge, GrimAge, and FitAge. Respective epigenetic age acceleration (EAA) measures were calculated, as the residual of the DNAm age regressed on chronological age. Also, DNAm-based pace of aging (DunedinePACE) and Mortality Risk Score (MRS) were measured. GWAS analysis was done using PLINK 1.9. We found a novel genome-wide significant SNP for GrimAge Acceleration, which is an estimation of all-cause mortality risk, calculated based on the DNAm surrogate measures for plasma proteins, smoking pack years plus sex, and age. The detected SNP located on chromosome 18 is mapped to the *SOCS2* gene which encodes a member of the suppressor of cytokine signaling (SOCS) family and has been reported to act in aging and longevity, potentially through its negative regulatory impact on mitochondrial fatty acid oxidation, and plasma IGF1 levels. Decreased levels of plasma IGF1 have been linked to extended lifespan in mice and the absence of SOCS2 expression leads to reduced lifespan in mice with high-growth characteristics. Our study suggests that genetic variations can influence the DNAm-based mortality risk signal captured by the GrimAge clock and endorse the *SOCS2* gene as a target for further anti-aging investigation.

Session Title: Epigenetics Poster Session I

PB2279 Genome-wide Association and Mendelian Randomization Analysis among 120,671 Chinese Pregnancies Identifies Novel Genetic and Molecular Risk Factors of Gestational Diabetes and Glycemic Traits.

Authors:

Y. Gu, H. Zheng, Y. Liu, S. Bian, S. Liu; Sch. of Publ. Hlth.(Shenzhen), Sun Yat-sen Univ., Shenzhen, China

Gestational diabetes mellitus (GDM) and hyperglycemia are the most common disorders in pregnancy. However, the causes of GDM remain obscure. To investigate the genetic and molecular risk factors underlying the phenotypic spectrum of GDM and glycemic traits, we collected non-invasive prenatal sequencing (NIPT) data, five glycemic and 52 biochemical measurements, and electronic medical records of 120,761 pregnancies, including 12,024 patients diagnosed with GDM, during a four-year period at two hospitals in Shenzhen, China. We conducted genome-wide association studies (GWAS) between genotypes derived from the NIPT sequencing and the GDM diagnostic, as well as the five glycemic traits that reflect GDM severity. Mendelian randomization (MR) was performed to explore the potential causal effects of the 52 biochemical measurements on GDM and glycemic levels. Through meta-analysis, we identified nineteen genetic loci significantly associated with GDM susceptibility ($P < 5\times10^{-8}$). Notably, several genetic variants demonstrated a strong association with GDM but only a suggestive or insignificant association with type 2 Diabetes (T2D), which present in the *MTNR1B* and *FOXA2* et al gene loci, indicating different etiological factors between GDM and T2D. One hundred and five genetic loci including 66 novel ones, were also significantly associated with the five glycemic traits. We observed substantial genetic differences among the baseline glycemic level (FPG and OGTT0H), glycemic level after the oral glucose challenges (OGTT1H and OGTT2H), and glycated hemoglobin level (HbA1c). Genetically higher body mass index, lymphocyte percentage, and genetically lower hemoglobin F and absolute neutrophils were causally associated with increased fasting glucose concentration and an increased risk of GDM risk. As the most powerful genetic study on GDM nowadays, our study suggests that public health strategy may take into account a patient's genetic predisposition for a better understanding, prevention, and treatment of GDM. Future clinical trials may be warranted to investigate the modulation of BMI or the incidence of the inflammatory response through nutrition therapy or medicine.

Session Title: Epigenetics Poster Session III

PB2281 Genome-wide association study of an epigenetic signature of monozygotic twinning

Authors:

J. van Dongen[1], Twinning Genetics Consortium, Genetics of DNA methylation Consortium, E. A. Ehli[2], F. A. Hagenbeek[1], J. J. Hottenga[1], D. I. Boomsma[1]; [1]VU Amsterdam, Amsterdam, Netherlands, [2]Avera Inst. for Human Genetics, Sioux Falls, SD

Monozygotic (MZ) twins and higher-order multiples arise when a zygote or embryo splits during pre-implantation stages of development. The underlying mechanisms of this event have remained a mystery. We recently identified a DNA methylation signature of MZ twinning in blood samples from adult twins, which showed strong replication across cohorts and in buccal samples from children (van Dongen et al 2021 PMID: 34584077). Here, we report a genome-wide association study (GWAS) on the DNA methylation score for MZ twinning in two datasets from the Netherlands Twin Register (NTR): Illumina 450k array data from blood (N=2841, EU ancestry, mean age=37, 66% females, 56% MZ twins) and Illumina EPIC array data from buccal (N=1150, EU ancestry, mean age=10, 47% females, 84% MZ twins). Genotyping was done on the Affymetrix 6, Affymetrix Axiom or Illumina GSA array. Genotype data were imputed against the Human Reference Genome plus Genome of the Netherlands (GoNL) reference panel. GWAS analysis was performed in gcta using --fastGWA-mlm correcting for 10 genetic ancestry PCs and genotype platform. GWAS analysis of the blood 450k array data identified three genome-wide significant loci. The top SNP (rs76157694, MAF=0.25) maps to *TBX4*; an embryonic transcription factor primarily known for its role in hindlimb development. This finding is consistent with our previous results of MZ-differentially methylated positions being enriched within *TBX4* binding motifs (van Dongen et al 2021 PMID: 34584077). GWAS analysis of the buccal EPIC array data identified no genome-wide significant loci. Next steps will be to perform a GWAS meta-analysis with the Genetics of DNA methylation consortium (GoDMC).

Session Title: Epigenetics Poster Session I

PB2282 † Genome-wide chromatin conformation analysis in individuals of African versus European ancestry reveals ancestry-specific chromatin interactions. individuals of African versus European ancestry reveals ancestry-specific chromatin interactions.

Authors:

L. Wang[1], W. Xu[2], X. Liu[2], F. Rajabli[1], K. Celis[1], L. Lu[3], M. Gearing[4], D. Bennett[5], M. Flanagan[6], S. Weintraub[7], C. Geula[8], T. Schuck[9], D. Davis[1], W. Scott[1], D. Dykxhoorn[1], A. Griswold[1], M. Pericak-Vance[1], J. Young[1], F. Jin[3], J. Vance[1]; [1]Univ. of Miami, Miami, FL, [2]Case Western Reserve Univ, Cleveland, OH, [3]Case Western Reserve Univ., Cleveland, OH, [4]Emory Universit, Atlanta, GA, [5]Rush Univeristy Med. Ctr., Chicago, IL, [6]Northwestern Univ., Chicago, IL, [7]Northwestern Univeristy, Chicago, IL, [8]Northwestern Univ, Chicago, IL, [9]Univ of Pennsylvania, Phildelphia, PA

Introduction: Genetic risk for Alzheimer Disease (AD) varies across populations with different ancestry. Identifying the mechanisms underlying these differences can lead to therapeutic interventions. Three-dimensional (3D) genome architecture regulates gene transcription, which could confer differential disease risk. At the top layer of 3D genome structure, the genome can be partitioned into Compartment A and B representing transcriptionally active and silenced chromatin, respectively. At the fine-scale level, chromatin loops are formed, often representing promoter-enhancer interactions. Genome-wide Hi-C analyses have revealed dynamic 3D genome structure during development and disease progression. The variation of 3D genome structure on different ancestry backgrounds, however, has not been studied.
Methods: Hi-C libraries from brain dorsal lateral frontal cortex were constructed with a 4-cutter enzyme. On average, 700 million of paired-end reads per library were obtained in 8 *APOE*e4 homozygotes (4 AF and 4 EU) who are age and sex-matched. Local ancestry (LA) blocks were calculated using RFMix and a reference panel from 1000 Genomes for EU, AF, and Amerindian ancestry. The AF samples have 68~88% AF genome while each EU sample has >95% EU genome. DeepLoop was used to call chromatin loop at a 5-kb resolution. Ancestry-specific loops were identified using pairwise T-test ($P<0.05$, Fold Change>2)
Results: 6206 compartment bins were compared at a 500kb-resolution. Most of the A and B compartments were consistent among individuals with 2~3% of them switched between ancestry groups. Of the switched compartment bins, more (2x) transcriptionally active chromatin was observed over EU LA than AF LA blocks. Of the 1,095,151 chromatin loops called, 12,082 loops are EU-specific, and 2885 loops are AF-specific. Compared to the EU-specific loops (median size=496 kb), the AF-specific loops are smaller (median size=158 kb) and more likely represent promoter-enhancer interactions as they are enriched for co-localization with H3K4me3 and H3K27ac marks while the EU-specific loops are enriched for CTCF occupancy. DNA sequence variations within the CTCF binding motif contribute to some ancestry-specific loops (N=30).
Conclusion: Hi-C analysis demonstrates ancestry-specific chromatin interaction differences in brain frontal cortex. The compartment variation reflects large-scale genome activity, and the fine-scale loop level variation is likely governed by complex mechanism such as DNA sequence variations. Our data suggest novel epigenomic mechanisms for ancestry-specific genetic risk which could reflect GWAS differences between ancestries.

Session Title: Epigenetics Poster Session II

PB2283 Genomic dissection of enhancer context sensitivity and synergy

Authors:

M. Maurano, R. Ordoñez, G. Ellis, H. J. Ashe; NYU Sch. of Med., New York, NY

Noncoding disease and trait-associated genetic variation is frequently interpreted in the context of genomic regulatory elements such as DNase I hypersensitive sites (DHSs). Most genomic DHSs lie within a few kilobases of another DHS, but individual DHSs are typically analyzed independently without accounting for their surrounding context. Using our recently developed Big-IN approach for rewriting genomic loci, we have previously identified distance-dependent synergy among the DHSs at the *Sox2* Locus Control Region (LCR) in mouse embryonic stem cells. Here, we extend our synthetic regulatory genomics approach to identify widespread genomic examples of context-dependent enhancers which have no activity on their own but can double the activity of a nearby DHS. We show that this synergy between nearby DHSs decays as a characteristic function of distance and its influence extends up to several kilobases. Finally, we fine map this context dependency to the contribution of individual TF recognition sequences. Our approach implicates specific sequence and architectural features underpinning the influence of genomic context, and suggests that interpretation of noncoding variation must be done at a haplotype level rather than implicitly assuming the surrounding sequence matches the genomic reference.

Session Title: Epigenetics Poster Session III

PB2284 High-throughput DNA methylation BeadChip assay for population-scale EWAS studies.

Authors:

D. Goldberg[1,2], N. Renke[3], S. Gruber[3], W. Zhou[1,2]; [1]Univ. of Pennsylvania, Philadelphia, PA, [2]Children's Hosp. of Philadelphia, Philadelphia, PA, [3]Illumina, Inc., San Diego, CA

DNA methylation is a stable covalent chemical modification to DNA that shapes cell identity and impacts diverse processes from genome stability to gene expression. Infinium Methylation BeadChips contain probes for preselected CpG sites and can be used for highly accurate quantification of CpG methylation at selected loci. Such microarrays are a cost effective and computationally tractable alternative and have been successfully utilized for population scale EWAS studies. Despite the successes of previous BeadChip versions, some traits associate with only subtle differences in DNA methylation, necessitating ever larger sample sizes. Therefore, a highly consolidated BeadChip strategically designed for EWAS studies would be beneficial for the population epigenetics community. To that end, we designed a compact Infinium Methylation BeadChip containing ~300K probes, approximately 30% of the size of the newly released EPICv2 array. We incorporated known EWAS hits of high statistical significance while also adding new CpGs to facilitate discovery and scientific rigor in EWAS. Approximately 50% of probes were derived from mining EWAS databases and literature. The methylation of these CpGs are linked to 8511 biological features from broad categories including cardiovascular, metabolic, neurodegenerative/psychiatric, autoimmune, genetic, environmental exposure and infection-related traits and diseases that were studied using existing Infinium array platforms. The remaining 50% are novel probe designs obtained from integrative analysis of >300 public bulk and single-cell WGBS datasets. These probes target DNA methylation associated with cell type, gene expression, chromatin accessibility and mono-allelic expression. Additional cytosines were selected from recent regulatory region annotations throughout the human genome to further enhance discovery power. While the vast majority of EWAS studies have focused exclusively on 5mC, our identification and incorporation of tissue specific and variable regions of 5-hydroxymethylcytosine will enable deeper insights into the epigenetic landscape. Altogether, our newly developed EWAS array contains both previously identified markers and candidate CpG sites and will be a valuable, more scalable tool for population level EWAS studies.

Session Title: Epigenetics Poster Session I

PB2285 How the phenotypes and genotypes affect irritable bowel syndrome (IBS)

Authors:

S-Y. Tsai[1,2,3,4,5], S-C. Hsieh[6], C-F. Kuo[7], N-W. Hsu[3], S. Hsieh[3]; [1]Dept. of Hlth.Policy and Management, Johns Hopkins Bloomberg Sch. of Publ. Hlth., Johns Hopkins Univ., Baltimore, MD, [2]Dept. of Lab. Med., Mackay Mem. Hosp., Taipei City, Taiwan, [3]Dept. of Med., Mackay Med. Coll., New Taipei City, Taiwan, [4]Inst. of BioMed. Sci., Mackay Med. Coll., Taipei City, Taiwan, [5]Inst. of Long-Term Care, Mackay Med. Coll., Taipei City, Taiwan, [6]Dept. of Res., Mackay Mem. Hosp., Taipei City, Taiwan, [7]Div. of Infectious Diseases, Dept. of Internal Med., Mackay Mem. Hosp., Taipei City, Taiwan

Introduction: Patients with IBS demonstrate symptoms such as recurrent and persistent abdominal pain combined with change of stool passage over a period 3 months. We aimed to conduct the first study to investigate the impact of the phenotypes and genotypes on IBS to propose clinical implications. **Methods:** Genotyping Data and Quality Controls SNP genotyping was conducted by using custom Taiwan BioBank (TWB) chips. We implemented quality control procedures for subsequent analysis for SNP exclusion: failure to achieve Hardy-Weinberg equilibrium (with a p value$< 1 \times 10^{-5}$), MAF< 5%, and a genotyping call rate <95%. There were a total of 291,928 SNPs remaining. Patient selection and statistics methods We investigated the data, including questionnaires with the phenotypes, such as health status, environment factors, etc**.** and whole genome genotyping. After excluding the participants without genetic data, we used propensity score to match age and sex ratio. Sequentially, we analyzed 1316 participants filtering and merging using PLINK 1.9. Univariate and multiple logistic analysis were performed to identify risk factors for IBS using R (version 4.2.2) to analyze the relevant data including phenotypes and genotypes between IBS and relevant cormobidities. **Results:** The prevalence of IBS recorded in TWB was 2.56%, with 58.22 % in the male group and 41.78% in the female group. Univariate analysis revealed that all the included factors increasing the risk of IBS and then multivariate regression analysis through two models to develop a nomogram for estimating the risk of IBS through the following formula: The risk of IBS = -(0.69 x sleep quality=poor) -(0.48 x sleep quality= moderate)+(0.444 x Peptic ulcer disease)+ (1.059 x gastroesophageal reflux disease)+ (0.869 x depression)+(0.751 x migraine) +(0.364 x Back and waist pain) -(0.772 x rs12116694=C pair to C) +(0.538 x rs9287949=C pair to C) +(0.756 x rs544951885=pair without A) +(0.45 x rs544951885= pair with one A) **Conclusions:** Depression is our significant comorbidity and correlated with a SNP, rs193922209, in the *FBN1* gene. Our findings have uncovered novel genetic contributions to IBS, given its multifactorial nature and the likely interactions of multi-omics, represent a promising approach to integrating phenotypes and genome-wide analysis and highlight the importance of genetic predisposition to IBS. To explore disease mechanisms, and the associations between genetic variation and syndromes, and intermediate phenotypes for identifying potential therapeutic targets is a warrant. Healthcare providers should be aware of those comorbidities and variants with their clinical implications.

Session Title: Epigenetics Poster Session II

PB2286 † HTRAnet predicts regulatory activity for sequences in diverse reporter assays and enables systematic comparative analysis of high throughput reporter assays

Authors:

Z. Chen[1], K. Siklenka[2], R. Castro[3], R. Tewhey[3], T. Reddy[2], A. Kundaje[1]; [1]Stanford Univ., Palo Alto, CA, [2]Duke Univ., Durham, NC, [3]Jackson Lab., Bar Harbor, ME

High-throughput reporter assays (HTRAs) such as episomal and lenti-virus massively parallel reporter assays (MPRAs) and self-transcribing active regulatory region sequencing (STARR-seq) can directly measure regulatory potential of libraries of candidate regulatory elements. Despite apparent similarities, different HTRAs often exhibit poor concordance of measured activity for matched sequences. Systematic comparison of these assays is challenging due to distinct input libraries and different biases. Further, the causal sequence syntax driving similarities and differences between these assays is not well understood. Deep learning models have been effective at modeling diverse regulatory profiling experiments such as DNase-seq and ATAC-seq. The models can also be interrogated to decipher predictive sequence motifs and syntax rules. Although previous studies have proposed deep learning models for reporter assays, they are often optimized for each specific assay and do not attempt to generalize across them. Further, reporter assays have not been compared through the lens of cis-regulatory syntax learned by predictive models.

In this study, we introduce HTRAnet (High Throughout Reporter Assay Network), a convolution neural network architecture to learn accurate sequence models of regulatory activity from diverse reporter assays. We trained HTRAnet on MPRA, lenti-MPRA, whole-genome-STARR-seq and ATAC STARR-seq experiments in the K562 cell-line. For all pairs of assays, HTRAnet predicted activity was substantially more concordant across matched sequences relative to measured activity, suggesting effective denoising of the signal. We interrogated the models with sequence attribution and motif discovery methods to extract and compare predictive sequence features. The ATAC-STARR-seq model revealed a strong sequence composition bias, which motivated the development of an automated bias correction approach. We further identified systematic differences of predictive sequence motifs and syntax between assays. e.g. SP1 motifs in promoters were preferentially identified in MPRA, whereas GFI1B and SNAI2 repressors were preferentially detected by STARR-seq. We further identified substantial similarity between sequence syntax from HTRAnet models and syntax derived from analogous BPNet models trained on ATAC-seq and DNase-seq data. HTRANet is a unified deep learning network that predicts regulatory activity from sequences accurately and facilitates a more comprehensive understanding of different reporter assays.

Session Title: Epigenetics Poster Session III

PB2287 Identification of exposure- and genetically driven epigenetic associations with lifetime cannabis use.

Authors:

S. Giamberardino[1], L. Zhou[1], J. Marks[1], K. A. Lawrence[2], Z. Xu[2], J. Taylor[2], D. P. Sandler[2], D. B. Hancock[1], E. O. Johnson[1,3], F. Fang[1]; [1]GenOmics and Translational Res. Ctr., RTI Intl., Research Triangle Park, NC, [2]Epidemiology Branch, Natl. Inst. of Environmental Hlth.Sci., NIH, Research Triangle Park, NC, [3]Fellow Program, RTI Intl., Research Triangle Park, NC

Cannabis use is prevalent among the U.S. population, yet reliable, long-term biomarkers for quantifying its use over an individual's lifespan are limited. This knowledge gap hinders our ability to accurately assess cannabis exposure and its potential effects on health outcomes. DNA methylation (DNAm) may serve as a viable and sensitive biomarker of cannabis use. We previously identified DNAm associations with lifetime cannabis use, which may reflect exposure-based (i.e., DNAm change resulting from exposure to cannabis) or predisposing risk (i.e., DNAm influenced by nearby genetic variation). Differentiating between these two interpretations will help inform their utility as biomarkers. We conducted a *cis*-methylation quantitative trait loci (meQTL) meta-analysis of two cohorts, GuLF (Gulf Long-term Follow-up Study; n=843) and ALSPAC (Avon Longitudinal Study of Parents and Children; n=904). We tested cannabis-implicated CpGs (p < 0.001 from our previous study) for association with 1000 Genomes-imputed genetic variants. *Cis*-meQTLs were identified using Matrix eQTL (adjusting for age, sex, surrogate variables, cell type proportions, cigarette smoking status, and study-specific covariates) and defined as occurring within a 1Mbp window around each CpG. Results were combined via inverse variance-weighted meta-analysis (using the software METAL). A significant *cis*-meQTL was defined based on meta-analysis p-value < 1×10^{-8}. Location-based enrichment for CpG and genic regions was tested using the LOLA R package. Of the 1,297 cannabis-implicated CpGs, 331 (~25%) had at least one significant *cis*-meQTL. CpGs with *cis*-meQTLs were enriched for intergenic and CpG shore regions and depleted in CpG islands and 5'UTRs, as compared to the total set of cannabis-implicated CpGs. Incorporating publicly available *cis*-meQTL results from the Genetics of DNA Methylation Consortium (GoDMC) study comprised of 32,851 individuals from 36 cohorts independent from ours, we found that 166 of our CpGs with *cis*-meQTLs were also present in GoDMC, and 165 of them have significant *cis*-meQTLs reported in GoDMC. In our *cis*-meQTL analyses, we identified a subset of cannabis-DNAm associations that are potentially driven by nearby genetic variation, suggesting the remaining CpGs may be primarily influenced by cannabis exposure. Future work will include incorporating additional cohorts of different ancestral backgrounds into the meta-analysis to increase the generalizability of our findings. Furthermore, identified *cis*-meQTL variants will be incorporated into epigenome-wide association studies of cannabis use to minimize genetic confounding.

Session Title: Epigenetics Poster Session I

PB2288 Identification of multi-sample consensus open chromatin regions using ROCCO

Authors:

N. Hamilton[1], T. Furey[2]; [1]UNC Chapel Hill, Chapel Hill, NC, [2]Univ North Carolina, Chapel Hill, Chapel Hill, NC

Analysis of open chromatin regions across multiple samples from two or more distinct conditions can determine altered gene regulatory patterns associated with biological phenotypes and complex traits. The ATAC-seq assay allows for tractable genome-wide open chromatin profiling of large numbers of samples. However, stable, broadly applicable genomic annotations of open chromatin regions are not available. Thus, most studies first identify open regions as peaks using methods designed for individual samples. These results are then combined heuristically to obtain a consensus peak set. But reconciling these sample-specific results post hoc from a large cohort with many samples is challenging, and informative spatial features specific to open chromatin signals are not leveraged effectively.

To address these issues with popular existing methodology, we propose a novel method, *ROCCO*: **R**obust Detection of **O**pen **C**hromatin Regions via **C**onvex **O**ptimization. The proposed method determines consensus open chromatin regions across multiple samples simultaneously by employing robust summary statistics to formulate a constrained optimization problem accounting for both enrichment and spatial features of open chromatin data. We show this formulation admits attractive theoretical and conceptual properties as well as superior empirical performance compared to benchmarks using publicly available data from 56 lymphoblastoid ATAC-seq samples.

Session Title: Epigenetics Poster Session II

PB2289 Identification of pathological circRNA landscape signatures and pathways at a critical window for severe progression of Alzheimer's disease

Authors:

F. Wang[1], Y. Li[1], H. Shen[1], P. Martinez-Feduchi[1], J. Kuehner[1], Y. Feng[2], B. Yao[1]; [1]Dept. of Human Genetics, Emory Univ. Sch. of Med., Atlanta, GA, [2]Dept. of Pharmacology and Chemical Biology, Emory Univ. Sch. of Med., Atlanta, GA

Circular RNAs (circRNAs) are covalently closed, single-stranded transcript, conserved and enriched in the brain, which are changed in age and longevity of a wide spectrum of eukaryote from nematodes to mammals. With the in-depth studies of circRNA functions in age-related neurodegenerative disorders, such as it has been showed a novel window into mechanisms of Alzheimer's disease (AD). circRNAs could contribute to AD by sponging microRNA (miRNA) and RNA-binding proteins (RBPs) to regulate the messenger RNA (mRNA) targets. Our recent study established a robust platform to identify circRNA landscape in human neurons and oligodendrocytes. Using this platform, we discovered alterations of cortical circRNA landscapes in the 5xFAD mouse model of AD between 5- and 7-month of age, a critical window for the transition of starting pathology to full manifestation. We revealed that circGigyf2, which was reported that it locates in neuron body and the whole dendritic arbor and is high expressed in brain, is the most significantly dysregulated circRNA in the 5xAD mouse cortex associated with pathologic advancement, which is highly conserved and progressively downregulated along with increased severity of dementia in human AD brains. Moreover, the alteration in circGigyf2 is probable positively associated with the age of onset of AD in the 5xFAD mouse model. Mechanistically, we identified AD-affected splicing factors that regulate the biogenesis of circGigyf2, which can sponge specific miRNAs that were hyperactivated along with the disease-stage-specific deficiency of circGigyf2. Additionally, we found that circGigyf2 interferes with key polyadenylation factors to modulate global and loci-specific mRNA polyadenylation and subsequent protein level. Together, our results unveiled pathological circRNA landscape changes during a critical window for severe AD progression and identified novel molecular mechanisms underlying dysregulation of conserved circRNA pathways that contribute to AD pathogenesis.

Session Title: Epigenetics Poster Session III

PB2290 Identification of protective genome-wide DNA methylation profiles for asymptomatic Alzheimer's Disease in postmortem brains.

Authors:

X. Han[1], D. Li[1], D. Goldstein[1], N. Sahelijo[1], R. Au[1,2,3,4], T. D. Stein[3,5,6], L. A. Farrer[1,3,7,8], J. Mez[3], G. R. Jun[1,7,8]; [1]Dept. of Med., Boston Univ. Sch. of Med., Boston, MA, [2]Dept. of Anatomy & Neurobiology, Boston Univ. Sch. of Med., Boston, MA, [3]Dept. of Neurology, Boston Univ. Sch. of Med., Boston, MA, [4]Dept.s of Epidemiology, Boston Univ. Sch. of Publ. Hlth., Boston, MA, [5]Dept. of Pathology & Lab. Med., Boston Univ. Sch. of Med., Boston, MA, [6]Dept. of VA Med. Ctr., Bedford, MA, [7]Dept. of Ophthalmology, Boston Univ. Sch. of Med., Boston, MA, [8]Dept.s of Biostatistics, Boston Univ. Sch. of Publ. Hlth., Boston, MA

Introduction: The Alzheimer's disease (AD) hallmark pathological abnormalities of neuritic amyloid-β plaques and neurofibrillary tangles are often present in postmortem brains from persons who were cognitively normal prior to death, i.e. asymptomatic AD (AsymAD). We hypothesized that DNA methylation (DNAm) patterns in postmortem brain tissue differ between AsymAD and symptomatic AD (SymAD).
Methods: We obtained RNA sequencing and DNAm data derived from brain tissue in participants of the Religious Orders Study and Rush Memory and Aging Project (ROSMAP). Genome-wide differential methylation between 185 AsymAD and 254 SymAD individuals was analyzed using a linear regression model with covariates for age at death, sex, and batch effect. Next, we evaluated differential expression of genes within 50 kb of significantly differentially methylated CpG sites using a regression model with covariates for age at death, sex, RNA integrity number, postmortem interval, and batch. For significantly differentially expressed genes, we tested for association tests of DNAm at their corresponding CpG site and expression level using linear models adjusted by the same covariates in the differential expression analysis. To identify shared methylated CpG sites accounting for both protection on AsymAD and changes in gene expression, we investigated colocalization of DNAm with AsymAD and gene expression. The posterior probability (PP) threshold that both traits were linked to a common site was set at 0.9.
Results: We identified 18 differentially methylated CpG sites within 50 kb of 53 genes between AsymAD and SymAD with false discovery rate adjusted P<0.05. Top-ranked findings include CpG site cg19832721(logFC=-0.020, $P=5.4x10^{-8}$) in the *MAPT-KANSL1* region and cg20510285 (logFC=0.012, $P=8.7x10^{-7}$) in *NEDD9*. *MAPT*, *MAPT-AS1*, *KANSL1*, *KANAL-AS1*, and *NEDD9* were differentially expressed between AsymAD and SymAD (top-ranked finding with *KANSL1*: log2FC=0.086, $P=4.6x10^{-5}$). Methylation of cg19832721 site was associated with *MAPT* (β=1.76, P=0.007), *MAPT-AS1* (β=2.02, P=0.005),and *KANSL1-AS1* (β=-4.81, P=0.008, Colocalization analysis revealed that methylation of cg20510285 was a common epigenetic variation (PP=0.92) for clinical symptoms of AD or expression of its nearby gene *NEDD9* (β=4.95, $P=9.6x10^{-4}$).
Conclusion: DNAm profiles in the *MAPT-KANSL1* region may explain resilience to clinical manifestations of AD by modulating expression of genes in this region. Our study provides insights about mechanisms that could be targeted to prevent onset of AD clinical symptoms and the preservation of cognitive health even in presence of AD-related neuropathological changes.

Session Title: Epigenetics Poster Session I

PB2291 Identification of RNA biomarkers and potential regulatory mechanisms and therapeutic targets for type 1 diabetic using a competing endogenous RNA regulatory network analysis

Authors:

M. Malik[1], M. Dashti[1], A. Channanath[1], S. John[2], R. Iqbal[1], M. Alhusayan[1], F. Al Mulla[1], T. Thanaraj[1]; [1]Dasman Diabetes Inst., Kuwait, Kuwait, [2]Dasman Diabetes Inst., 15462, Dasman, Kuwait

Type 1 diabetes mellitus (T1D) is a growing autoimmune disease characterized by a genetic predisposition. There is a rising interest in understanding the involvement of microRNAs (miRNAs), circular RNA (circRNA), and long non-coding RNA (lncRNA) in the pathogenesis of this disease. However, the specific roles of these molecules in T1D remain unclear. Hence, the objective of this study was to identify distinct miRNA, circRNA, and lncRNA signatures and explore their contributions to T1D through the analysis of competing endogenous RNA (ceRNA) networks. Accumulating evidence suggests that mRNA, miRNAs, circRNA, and lncRNA play crucial roles in the processes associated with the pathogenesis of T1D. To identify differentially expressed lncRNAs, mRNAs, and miRNAs, expression profiles from GSE133217, GSE133225, and GSE55100 were obtained from the Gene Expression Omnibus. These profiles were derived from peripheral blood mononuclear cells of both T1D patients and healthy controls. Subsequently, a ceRNA regulatory network was constructed, followed by functional and pathway enrichment analysis. Additionally, a T1DM-related ceRNA regulatory network was established using the Human microRNA Disease Database for further pathway enrichment analysis. Simultaneously, T1D-related pathways were extracted from the Comparative Toxicogenomics Database (CTD). In this study, potential blood biomarkers related to T1D are identified using large-scale gene expression data. This is achieved through the computational construction of miRNA, circRNA, and lncRNA-target gene interactomes, as well as transcription factor networks. To enhance our understanding of T1D diagnosis, prognosis, and therapeutic interventions, further validation of these computational results is warranted through in vitro and in vivo studies in the future.

Session Title: Epigenetics Poster Session II

PB2292 Identification of small volatile inhibitors targeting methylation mediated dysregulated ribonucleoproteins in lung adenocarcinoma

Authors:

A. Mukherjee, M. Kuppusamy Selvam; Manipal Inst. of Technology, Manipal, India

Lung carcinoma is one of the most prevalent and life-threatening cancers globally, with tobacco smoking being the most significant cause of lung cancer deaths. Alterations in DNA methylation at CpG sites are associated with smoking-induced lung cancer. Enhancers play a vital role in gene expression, and aberrantly methylated enhancers can lead to malignant traits. The study involved linking enhancer status with the expression of target genes. The network-based approach aided in determining the dysregulated ribonucleoproteins TWISTNB, TRUB1, and RSL1D1, playing an important role as central regulators of dysregulated Ribosome biogenesis and RNA processing. Small molecules targeting epigenetic regulations have emerged as attractive therapeutics to overcome resistance to anticancer drugs. The study focused on volatile compounds with low molecular weight, high vapour pressure, and lipophilic. The Principle Component Analysis (PCA) was employed, and 1042 compounds were clustered based on structural similarity using the k-means algorithm to discover novel hits from large chemical libraries with similar pharmacokinetic and structural properties. The study investigated a technique for predicting the altered methylation in enhancer regions in modulating target genes by combining epigenomic and transcriptomic data. The PCA demonstrated the vigour of the novel method in screening potential leads by reducing high dimensional volatile chemical space, and the pharmacokinetics and toxicity profiling eliminated unfavourable leads, averting the side effects. Three potential leads were identified, #309 and #949, from the chemical class of benzenoids, and #347, from organic acids and derivatives, possessing higher binding affinity with the dysregulated ribonucleoprotein targets. The MD simulation of the complexes revealed their stability and rigidity and intermolecular Hydrogen bonds contributing to higher binding energies of the complexes.

Session Title: Epigenetics Poster Session III

PB2293 † *In vitro* model of gastrulation: Effects of alcohol on methylome, transcriptome and metabolome.

Authors:

E. Wallén[1], K. Rämö[1], J. Vehviläinen[1], J. Sokka[2], H. Marjonen-Lindblad[1], H. Skottman[3], T. Otonkoski[2,4], R. Trokovic[2], P. Auvinen[1], O. Kärkkäinen[5], N. Kaminen-Ahola[1]; [1]Dept. of Med. and Clinical Genetics, Medicum, Univ. of Helsinki, Helsinki, Finland, [2]Res. Programs Unit, Stem cells and Metabolism and Biomedicum Stem Cell Ctr., Faculty of Med., Univ. of Helsinki, Helsinki, Finland, [3]Faculty of Med. and Hlth.Technology, Tampere Univ., Tampere, Finland, [4]Children's Hosp., Helsinki Univ. Central Hosp., Univ. of Helsinki, Helsinki, Finland, [5]Sch. of Pharmacy, Faculty of Hlth.Sci., Univ. of Eastern Finland, Kuopio, Finland

Prenatal alcohol exposure (PAE) can have detrimental effects on early embryonic development, including alterations of gene regulation and DNA methylation (DNAm) as well as a wide range of physiological and neurological defects associated with Fetal Alcohol Spectrum Disorders (FASD). The underlying molecular mechanisms are poorly understood, but epigenetic changes can play a major role as early pregnancy is a sensitive period for environment-induced epigenetic changes. Gastrulation is one of the most critical events in embryonic development, involving the formation of the three germ layers. As human gastrulation is difficult to study due to technical and ethical limitations, we investigated the effects of alcohol exposure in human embryonic stem cells (hESCs) and in differentiated endodermal, mesodermal, and ectodermal cells as a model of gastrulation. hESCs (H1) were differentiated into the embryonic germ layers by using the STEMdiff™ Trilineage Differentiation Kit. For the alcohol-exposed wells, the media were supplemented with alcohol at final concentrations of 20 mM (moderate exposure) or 70 mM (severe exposure). Genome-wide DNAm was measured using the Illumina Infinium MethylationEPIC Array, gene expression was analysed with 3'mRNA-sequencing and metabolite profiles were characterized using a liquid chromatography-mass spectrometry-based (LC-MS) non-targeted metabolomics method. Our results show that especially severe alcohol exposure alters DNA methylation and gene expression of genes important for early development in hESCs and all germ layers. Metabolomic data indicate extensive perturbations in hESCs and all germ layers, highlighting the broad cellular impact of alcohol exposure. We observed the most prominent alterations in the ectodermal cells, including DNAm and gene expression alterations in the developmentally important genes *MAP2*, *NEUROD4*, *RAX* and *SIX3* and disruptions in the energy metabolism, such as changes in the levels of citrulline, citric acid and succinic acid. Our final aim is to combine these data sets with multi-omics methods to create a versatile picture of the possible mechanisms underlying the alcohol exposure. This study demonstrates that each of the germ layers exhibit a unique response to alcohol and provides new insights into the effects of alcohol in the beginning of embryonic development.

Session Title: Epigenetics Poster Session I

PB2294 Insights into pathogenetic basis of KMT2D related hearing loss from mice and humans

Authors:

A. Kalinousky[1], T. Luperchio[2], K. Schrode[1], L. Zhang[1], J. Harris[3], A. Lauer[1], H. Bjornsson[4]; [1]Johns Hopkins Univ., Baltimore, MD, [2]Johns Hopkins Sch. of Med., Ellicott City, MD, [3]Kennedy Krieger Inst., Baltimore, MD, [4]Univ. of Iceland, Reykjavik, Iceland

Background: Kabuki syndrome (KS) is a Mendelian disorder of the epigenetic machinery caused, most commonly, by loss of function variants in *KMT2D*, resulting in KS type 1. Clinical criteria for diagnosis of KS includes intellectual disability, characteristic facial features, postnatal growth deficiency, and skeletal anomalies. In addition to these phenotypes, another penetrant phenotype is progressive hearing loss. Current clinical dogma states that hearing loss in KS is the result of either repetitive ear infections in early life in context of immune dysfunction or related to the abnormal craniofacial structure. There are currently no clinical recommendations to prevent or mitigate the progressive hearing loss and the mechanistic basis is currently unknown. **Methods:** Hearing loss in a mouse model of KS1 was determined using auditory brainstem response testing (ABR). ABR stimuli consisted of brief clicks that excite auditory neurons to assess overall integrity of auditory function. Immunofluorescent staining was performed to quantify inner and outer hair cells (OHC) along the cochlea. Lastly, we gathered information about hearing loss from 21 children and adults with a molecularly confirmed diagnosis of KS1. **Results:** 15 participants reported hearing loss, with the average age of hearing loss occurring at 7 years. Three participants have hearing loss from birth, and 5 individuals reported having either middle and/or inner ear abnormalities. 53% of those with hearing loss experienced a severe number of infections (16-50). Interestingly, 40% of those without hearing loss experienced either a severe or a moderate number of ear infections (6-15). When performing ABRs using KS mice, unexpectedly, we observe significant abnormalities of hearing from the onset of hearing in KS mice at all frequencies in comparison to their wildtype (WT) littermates. Examination of the sensory epithelium by immunofluorescence staining in a small cohort showed a subtle increase in percent missing OHCs in adolescent KS mice compared to WT littermates at higher frequencies. **Conclusions:** Our findings suggest that KMT2D-dysfunction leads to early and progressive hearing loss, with outside factors, such as infection, aiding to this hearing loss. 40% of participants with hearing loss reported having sensorineural hearing loss, which results from damage to the hair cells within the inner ear, the vestibulocochlear nerve, or the brain's central processing centers. Staining of hair cell markers showed subtle damage in KS mice, but further quantification would need to be done to make any strong conclusions on etiology of the hearing loss in KS mice.

Session Title: Epigenetics Poster Session II

PB2295 Integrating 3D genome maps with vertebrate synteny blocks to identify genomic regulatory blocks for conserved traits

Authors:

A. Zimmerman[1,2], J. Palermo[3], A. Chesi[1], F. Doldur-Balli[1], S. Sonti[2], M. Pahl[2], E. Brown[3], J. Pippin[2], A. Wells[2,1], D. Mazzotti[4], P. Gehrman[1], A. Keene[3], A. Pack[1], S. Grant[2,1]; [1]Univ. of Pennsylvania, Philadelphia, PA, [2]Children's Hosp. of Philadelphia, Philadelphia, PA, [3]Texas A&M, College Station, TX, [4]Univ. of Kansas Med. Ctr., Kansas City, KS

The last decade has seen advances in genome sequencing along with large-scale genome-wide association studies (GWAS) for complex traits; however, functional interpretation of these data is complicated by modest effect sizes, causal gene prediction, and model organism interrogation. A major limitation in GWAS efforts is identification of a causal effector gene connected to variants residing predominantly in non-coding regions. Epigenomic information such as chromatin conformation and gene expression can help link GWAS variants to a putative effector gene(s). Ultimately, functional characterization in cultured cells or model organisms is necessary to determine the role of a variant and/or gene on a phenotype, yet this step is rarely performed. Functional interpretation of GWAS results using model organisms can be obscured by poor genetic conservation between human and the model or misidentification of the true effector gene. We propose a pipeline for the identification of true effector genes by integrating high-resolution epigenomic data with model organism synteny maps to implicate not only effector genes, but causal regulatory interactions that have been conserved across evolution for a highly conserved trait: sleep. Sleep is one of the most conserved behaviors in the animal kingdom, and many genes that regulate sleep are highly conserved. For this reason, we took human GWAS sentinel variants for sleep traits, such as insomnia and sleep duration, and performed our validated variant-to-gene mapping approach integrating these variants with ATAC-seq and high-resolution promoter-focused capture C across a panel of cell-types. This approach identified putative causal effector genes, whose promoters harbor an accessible chromatin contact by ATAC-seq/ promoter capture C and are highly expressed within a trait-relevant cell type, such as neural progenitor cells. We then performed a screen of candidate causal genes using neuron-specific (nSyb) RNA interference in *Drosophila* to narrow our list to high-confidence effector genes producing robust sleep phenotypes. Finally, we leveraged genomic synteny maps in a diurnal vertebrate, zebrafish, to identify genes which are conserved within a syntenic block harboring the GWAS-implicated gene (nearest to the GWAS variant) and the effector gene implicated through our variant-to-gene approach. Interestingly, the chromatin contacts from human data overlap with syntenic blocks in zebrafish, whereby genes conserved at both the gene and synteny level produced the most consistent phenotypes across species. These data suggest GWAS findings represent critical genomic regulatory blocks conserved across species.

Session Title: Epigenetics Poster Session III

PB2296 Investigating the potential of single-cell DNA methylation data to detect allele-specific methylation and imprinting

Authors:

N. Johnson, D. Cutler, K. Conneely; Emory Univ., Atlanta, GA

Allele-specific methylation (ASM) is an epigenetic modification whereby one parental allele becomes methylated and the other unmethylated at a specific locus. ASM is most often driven by the presence of nearby heterozygous variants that influence methylation, but also occurs somatically in the context of genomic imprinting. In this study, we investigate ASM using publicly available single-cell reduced representation bisulfite sequencing (scRRBS) data on 608 B cells sampled from six healthy individuals. We developed a likelihood-based criterion to test whether a CpG exhibited ASM, based on the distributions of methylated and unmethylated reads both within and across cells. Applying our likelihood ratio test, 65,998 CpG sites exhibited ASM according to a Bonferroni criterion ($p<8.4\times10^{-9}$). We also called ASM at the sample level: across the 6 samples, 93,290 CpG sites exhibited ASM in at least 1 sample, while 4,613 showed ASM in ≥2 samples, and 298 CpG sites showed ASM in ≥3 samples. To evaluate the accuracy of our method, we called heterozygous variants from the scRRBS data, which enabled variant-based calls of ASM within each cell for CpGs within 50bp of a heterozygous variant. Comparing sample-level ASM calls to the variant-based measures of ASM, we observed a positive predictive value of 68-100% across samples. We observed high concordance of ASM across samples and an overrepresentation of ASM in previously reported imprinted genes (OR: 3.7, 95% CI [2.4, 5.8]; $p=8.9\times10^{-11}$) and genes with imprinting binding motifs (OR: 5.3, 95% CI [5.0, 5.5]; $p<10^{-15}$). Our study demonstrates that single-cell bisulfite sequencing is a potentially powerful tool to investigate ASM, especially as studies expand to increase the number of samples and cells sequenced.

Session Title: Epigenetics Poster Session I

PB2297 Large-scale cross-ancestry genome-wide meta-analysis of serum uric acid

Authors:

C. Cho[1], B. Kim[1], D. Kim[1], M. Hwang[2], I. Shim[3], M. Song[1], Y. Lee[1], R. Do[4], W-Y. Park[5], W. Myung[6], B-J. Kim[7], Y. Kim[7], H-H. Won[1]; [1]Sungkyunkwan Univ., Seoul, Korea, Republic of, [2]NIH, Cheongju-si, Korea, Republic of, [3]Sungkyunkwan Univ., Seongnam-si, Gyeonggi-do, Korea, Republic of, [4]Icahn Sch. of Med. at Mount Sinai, New York, NY, [5]Samsung Med Ctr, Seoul, Korea, Republic of, [6]Seoul Natl. Univ. Bundang Hosp., Seongnam, Korea, Republic of, [7]Korea NIH, Cheongju-si, Korea, Republic of

Background Serum uric acid (SUA) is a biological waste produced by breaking down purine. A high SUA level (hyperuricemia) is associated with increased risk for gout and several diseases. We aimed to identify novel variants, genes, and pathways associated with SUA through a large-scale cross-ancestry genome-wide association study (GWAS) meta-analysis.
Methods We performed a large-scale cross-ancestry meta-analysis of 1,029,323 individuals of multiple ancestries (Europeans = 677,373, East Asians = 219,768, others = 132,182) to identify novel loci and ancestry-specific meta-analysis (European and East Asian) to examine the shared and distinct genetic architecture of each ancestry. We performed a transcriptome-wide association study (TWAS), colocalization analysis, and functional enrichment analysis to identify SUA-associated genes and pathways.
Results The cross-ancestry meta-analysis additionally identified 21 loci that were previously unreported. The genetic correlation of SUA between two ancestries (European and East Asian) was estimated to be high (ρge = 0.942, s.e. = 0.079). In total, 173 genes in glomerular (GLOM) and tubulointerstitial (TUBE) tissues from the Nephrotic Syndrome Rare Disease Clinical Research Network III (NEPTUNE) were colocalized with our GWAS results (posterior probability of colocalization > 0.8). As a result of TWAS, while 178 genes were commonly significant in the three meta-analyses, 183 of 945, 83 of 780, and 75 of 334 significant genes were significant only in the cross-ancestry, European, and East Asian, respectively. The tissue enrichment analysis identified various significantly enriched tissues in the urogenital, digestive, and endocrine systems. Moreover, we identified additional results of the KEGG pathway, such as "systemic lupus erythematosus" related to the high-level functions of biological systems through gene-set enrichment analysis.
Conclusion In summary, we investigated variants, genes, tissues, and pathways associated with SUA through the large cross-ancestry and ancestry-specific meta-analysis for SUA and various subsequent analyses. Our study adds insights into the genetic architecture by aggregating abundant genomic data.

Session Title: Epigenetics Poster Session II

PB2298 Leveraging multi-omics data to interpret a genome-wide meta-analysis of DNA methylation and PTSD in 23 military and civilian cohorts.

Authors:

M. Logue[1], S. Katrinli[2], A. Wani[3], A. Maihofer[4], A. Zannas[5], N. Nugent[6], X. Zhao[7], C. Nievergelt[8], M. Uddin[9], A. Smith[10], PGC PTSD Epigenetics Workgroup; [1]Natl. Ctr. for PTSD at VA Boston, Boston, MA, [2]Emory Hlth.care, Atlanta, GA, [3]Coll. of Publ. Hlth., Univ. of South Florida, Tampa, FL, [4]UCSD, La Jolla, CA, [5]Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, [6]Brown Univ., Providence, RI, [7]Boston Univ. Sch. of Publ. Hlth., Boston, MA, [8]Univ California San Diego, La Jolla, CA, [9]Univ. of South Florida, Tampa, FL, [10]Emory Univ, Atlanta, GA

Background: Epigenetic factors, including DNA methylation (DNAm), are impacted by traumatic stress, and may help to distinguish between individuals with and without posttraumatic stress disorder (PTSD). Here, we present the results of the largest epigenome-wide association study (EWAS) of PTSD to date. **Methods:** This study includes 5,077 participants (2,156 PTSD cases and 2,921 trauma-exposed controls) from 23 cohorts participating in the Psychiatric Genomics Consortium (PGC) PTSD Epigenetics Workgroup. DNAm was assayed from blood with the Illumina HumanMethylation450 BeadChip, which measures DNAm at approximately 450K methylation sites across the genome, or the Illumina MethylationEPIC BeadChip (850K sites). For each cohort, a common consortium-developed quality control pipeline was applied, and DNAm of each CpG site was regressed on PTSD, sex (if applicable), age, blood-cell proportions, and principal components for ancestry. Meta-analyses were performed using an inverse variance-weighted approach. TOAST was used to test for cell-type-specific associations in the subset of 11 cohorts with EPIC BeadChip data. An epigenome-wide significance threshold of $p<9e^{-8}$ was applied across all analyses. **Results:** We identified 11 CpG sites associated with PTSD in the overall model, as well as 14 associated with analyses of specific strata (sex, ancestry, and military vs civilian cohorts), including sites implicated in our previously published analyses. Many of these CpGs exhibit blood-brain correlation in methylation levels, cross-tissue associations with PTSD in multiple brain regions of postmortem samples, and associations with genetic and gene-expression variation. For example, methylation of cg04987734 (*CDC42BPB*) was higher in PTSD cases than controls in blood (Z=5.53; $p= 3.26e^{-8}$) and the dorsolateral prefrontal cortex (T=2.99; $p= 3.90e^{-3}$); blood-based eCpGs (R=0.2, p=0.023) and gene-based tests (Z=3.20; p=6.94E-04) also showed significant associations for this locus. None of the CpGs from the bulk-tissue analysis could be attributed to individual immune cell types. However, additional associated CpGs were identified in the cell-type specific analyses: 88 in B cells, 1 in CD4+ T cells, 2 in CD8+ T cells, and 5 in NK cells. **Conclusions:** This study identifies novel PTSD-associated CpGs from bulk tissues and individual lymphocyte subtypes. Leveraging data from postmortem brain samples, in vivo models, GWAS, and genome-wide expression data can elucidate the biological processes associated with PTSD by identifying methylation in particular pathways and cell types involved in PTSD pathophysiology.

Session Title: Epigenetics Poster Session III

PB2299 Locus-specific neuroepigenome editing for X-linked intellectual disability

Authors:

J. Halmai, J. Waldo, C. Gonzalez, K. Braczyk, D. Cameron, M. Singh, K. Fink; UC Davis, Sacramento, CA

A disproportionate number of genes causative for intellectual disability are found on the X chromosome (XLID). A potential therapeutic approach for XLIDs that affect females is to reactivate the silenced wild-type allele in cells that express the mutant allele. We have previously demonstrated that targeted X chromosome reactivation (XCR) using CRISPR/dCas9 is possible in human and mouse cells. We demonstrated that DNA CpG de-methylation editing of an X-linked gene promoter results in reactivation of the target gene. To further characterize the rule sets underlying targeted XCR, we aim to establish genomic and epigenomic features that will predict targeting of X-linked genes using epigenomic perturbations on a X chromosome-wide scale. This screen was first applied to human ENCODE tier 1 cell lines, and future studies will address this approach *in vivo* in the mouse genome. In addition, these experiments will interrogate the temporal order required for epigenome editors to change chromatin marks and removal of promoter DNA hypermethylation in favor of targeted XCR. In order to do so, chemical induced proximity ligation of CRISPR/dCas9 editors was used following a washout period to allow sequential epi-editing of chromatin loci. Finally, to investigate the usage of a hypercompact Cas variant, we have established a toolbox of epi editor fusion proteins tested in a MeCP2 reporter cell line and *in vivo*. If successful, locus-specific neuroepigenome editing holds great promise for those suffering from XLID and other neurological disorders with an X-linked etiology.

Session Title: Epigenetics Poster Session I

PB2300 Longitudinal DNA methylation highlights genes for COPD progression.

Authors:

O. Rafique[1], Y. Huang[1], K. H. Shutta[1,2], R. Min Hyung[1], K. Kim[1], P. Kachroo[1], C. Lopes-Ramos[1], K. A. Pratte[3], R. Bowler[4], E. K. Silverman[1,5], V. J. Carey[1], D. L. DeMeo[1,5]; [1]Channing Div. of Network Med., Brigham and Women's Hosp., Harvard Med. Sch., Boston, MA, [2]Harvard T.H. Chan Sch. of Publ. Hlth., Dept. of Biostatistics, Boston, MA, [3]Div. of Biostatistics and Bioinformatics, Dept. of Med., Natl. Jewish Hlth., Denver, CO, [4]Div. of Pulmonary and Critical Care Med., Dept. of Med., Natl. Jewish Hlth., Denver, CO, [5]Div. of Pulmonary and Critical Care Med., Brigham and Women's Hosp., Harvard Med. Sch., Boston, MA

To investigate the impact of time varying nature of DNA methylation (DNAm) on the progression of chronic obstructive pulmonary disease (COPD), we conducted a longitudinal epigenome-wide association study (EWAS). The EWAS findings were integrated with proteomic data to reveal the interplay between methylation and proteomics in COPD. Methods: Infinium MethylationEPIC BeadChip data was obtained from peripheral blood leukocyte DNA from 4,493 current and former smokers from the Genetic Epidemiology of COPD (COPDGene) study consisting of two time points of DNAm and FEV1 data (baseline and 5-year follow-up visits) of non-Hispanic white and African American subjects with a history of smoking, a median age of 62.2 years at enrollment and 51 percent male subjects. SomaScan blood proteomic data was available from the 5-year follow-up visit. Linear mixed-effects regression (LMER) was used to find the associations between longitudinal DNAm and longitudinal forced expiratory volume in 1 second (FEV1); adjusted for sex, age, race, cell counts, pack-years of smoking, current smoking status, height and time since baseline visit. Causal mediation analysis to identify proteins potentially mediating the relationship between DNAm and FEV1 was conducted using a nonparametric bootstrap procedure from the R package Mediation. Functional overlap analysis on the genome-wide significant CpGs was performed using the eFORGE tool. Results: From the longitudinal FEV1 and longitudinal DNAm data, LMER identified 66 CpGs having genome-wide significant associations with FEV1 ($p \leq 5 \times 10\text{-}8$). Intraclass correlation (ICC) of the 66 CpGs calculated using DNAm M-values from both time points revealed that the majority (40) have ICC between 0.5 and 0.8 and 22 CpGs have ICC greater than 0.8, suggesting that FEV1-associated CpGs are stable over 5 years. Chromosome 17 demonstrated three significant associations over a 440 base pair region annotated to the SOCS3 gene (cg13343932 with $p = 2.12 \times 10\text{-}18$, cg11047325 with $p = 8.77 \times 10\text{-}16$ and cg18181703 with $p = 1.07 \times 10\text{-}13$). An X chromosome CpG near the zinc-finger repressor, BCOR, was also associated with longitudinal FEV1 (cg18422972 with $p = 2.14 \times 10\text{-}8$). Functional Overlap analysis of the 66 CpGs revealed significant enrichment in histone methylation (H3K4me1 and H3K4me3) in blood and lung tissue. Mediation analysis identified ITPRIPL1, NOTCH1 and IGFBP5 as three candidate proteins partially mediating the signal between DNAm and FEV1. Conclusions: EWAS of longitudinal DNAm data revealed that the CpG sites that are stable over time are associated with longitudinal lung function (FEV1) and are linked with epigenetic and inflammatory signals.

Session Title: Epigenetics Poster Session II

PB2301 Long-read Khoe-San genomes reveal structurally divergent loci overlapping genic regions

Authors:

D. Soto[1], W. Palmer[2], J. Prado-Martinez[3], P. Norman[4], M. Dennis[5], B. Henn[6]; [1]Univ. of California Davis, Davis, CA, [2]Gencove, Long Island City, NY, [3]Inst. Pasteur, Paris, France, [4]Univ. of Colorado, Aurora, CO, [5]Univ. of California, Davis, Davis, CA, [6]UC Davis, Davis, CA

Long-read sequencing of hundreds of individuals is increasing our understanding of the genomic variation landscape of modern humans, including complex structural variation. While West African populations have been included in long-read sequencing efforts, most regions of Africa remain underrepresented. Considering that genetic diversity within African populations can be even greater than between continental groups, the lack of long-read sequencing representation suggests that ample genetic diversity is missing in these datasets. In particular, the Khoe-San indigenous peoples of southern Africa contain many of the most divergent haplotypes in extant human lineages. We performed 10X Genomics long-range sequencing and de novo assembly of three individuals of Khoe-San ancestry, including one ǂKhomani San and two Nama individuals, obtaining diploid pseudo haplotypes (N50: 54, 88,and 155 kbp, respectively). Collectively, the Khoe-San genomes harbor 146 kbp of short non-reference unique insertions (NUI), or non-repeat sequences (<10% repeats), not represented in the human reference assemblies GRCh38, which is of mostly West African ancestry, and T2T-CHM13, of European ancestry. NUIs range in size from a lower bound of 50 bp to 1.5 kbp, and overlap 18 genes, three of which are predicted to disrupt coding sequence. Compared to previously published assemblies from other ancestries (n=8), we found that 63 kbp (43%) of NUIs are Khoe-San-specific. We also identified 13,161 contigs not mapping to either reference, mostly lacking repetitive sequences, potentially representing the most divergent loci and totalling ~12 Mbp. Further, focusing on structural variants (>50 bp), we identified 34,164 deletions, 1,925 mobile element insertions, and 341 inversions, out of which 19,396 (57%), 790 (41%), and 230 (67%) were only found in the Khoe-San assemblies, respectively. Notably, of the Khoe-San specific variants, 28 deletions and 19 inversions ablated or affected splicing of protein-coding genes. To establish their population frequencies, we genotyped these events in a short-read sequencing cohort of Khoe-San ancestry (n=92). From this cohort, we also identified genes with divergent copy-number in Khoe-San populations, some of which have not been annotated as segmental duplications in the T2T-CHM13 reference of European ancestry. Finally, we leveraged long haplotypes to investigate divergence of structurally complex loci, including the highly divergent HLA locus, and assess the improvement of mappability on reference bias in individuals of Khoe-San ancestry.

Session Title: Epigenetics Poster Session III

PB2302 † Massive-scale meta-analysis of genetic regulation of RNA editing in the human brain identifies new risk genes and mechanisms for neurological disease

Authors:

J. Humphrey, W. H. Cuddleston, B. Z. Muller, K. BP, T. Raj; Dept.s of NeuroSci., and Genetics and Genomic Sci., Ronald M. Loeb Ctr. for Alzheimer's Disease, Icahn Sch. of Med. at Mount Sinai, New York, NY

Genome-wide association studies (GWAS) have identified genetic loci associated with neurodegenerative and neuropsychiatric diseases, but it remains challenging to describe the precise causal variants and link them to genes and underlying molecular mechanisms. Quantitative trait locus (QTL) mapping associates genotypes with molecular phenotypes, and determining whether GWAS risk loci and QTLs share a genetic basis can elucidate molecular pathways which potentiate disease risk. Efforts to identify genetic variants which regulate gene expression (eQTLs) and splicing (sQTLs) outpace that of all other gene regulatory phenotypes, but there is a critical need to consider other molecular traits.

Adenosine to inosine (A-to-I) RNA editing is among the most abundant of post-transcriptional modifications, whose role in brain homeostasis and disease has been extensively reported. Although a large body of work has focused on regulation of A-to-I editing by trans-acting factors, regulation of A-to-I editing by cis-acting genetic variants (edQTL) is largely unresolved in the brain, as nearly all studies which report edQTLs make use of the relatively small GTEx resource (<300 unique donors).

To test the hypothesis that edQTLs explain neurological disease heritability, we are mapping edQTLs in the largest collection to date of paired transcriptomic and genetic data from human brain tissues, the BigBrain project, which unites 15 cohorts of 13,061 samples sourced from 7 unique brain regions, representing a 30-fold enhancement of sample size from previous edQTL studies.

We have mapped edQTLs in 11 cohorts (2,707 samples, 1,574 unique European donors), using random effects meta-analysis to identify 26,564 significant cis-edQTLs in 3,261 genes, comprising ~16.3% of all tested A-to-I editing sites (FDR < 0.05). Using GWAS summary statistics from several neurological and psychiatric diseases we performed genetic colocalization. By including eQTL and sQTL summary statistics mapped in the same BigBrain cohorts, we have identified several examples of shared associations between editing, expression and splicing, such as *ICA1L* in Alzheimer’s disease and *CTSB* in Parkinson’s disease. However, several GWAS loci colocalized solely with RNA editing, such as the Alzheimer’s disease locus *APH1B*.

Comprehensively defining genetic regulation of RNA editing in the brain is informing novel relevance for non-coding A-to-I events, and will support our understanding of how RNA editing drives complex brain disease phenotypes. We will discuss our further plans for testing for mediation between editing, splicing and expression QTLs, and searching for trans-genetic effects.

Session Title: Epigenetics Poster Session I

PB2303 MethPhaser: Methylation-based haplotype phasing of human genomes

Authors:

Y. Fu[1], S. Aganezov[2], M. Mahmoud[3], J. Beaulaurier[2], S. Juul[2], T. Treangen[1], F. Sedlazeck[3]; [1]Rice Univ., Houston, TX, [2] Oxford Nanopore Technologies Inc, New York, NY, [3]Baylor Coll. of Med., Houston, TX

The assignment of variants across haplotypes, a process called phasing, is crucial for predicting the consequences, interaction, and inheritance of mutations and is a critical step in improving our understanding of phenotype and disease. This is exemplified in the case of the genes HLA or TPMT, where the functional inactivation of the gene depends on whether certain variants are in cis- or trans-relationships to each other. While there are three main phasing methods, only one (read-based phasing) can comprehensively provide information also about de novo single nucleotide variants (SNVs) and their origin. But this is often limited by the read size and length of homozygous regions in the genome.

To address this, we developed MethPhaser, the first method that uses haplotype-specific methylation signals from Oxford Nanopore Technologies (ONT) to extend SNV-based phasing. MethPhaser operates on a set of already-phased SNVs to extend or merge individual phased regions together, often by extending the phase blocks into homozygous regions that contain haplotype-specific methylation signatures.

Benchmarking using trio-based phasing data showed that MethPhaser is able to extend the genome-wide phasing by 1.6 to 2.5 fold, depending on the overall read length and ONT chemistry used, while only marginally increasing the phasing error rate from 0.03% to 0.05%. We observed that MethPhaser can successfully improve phasing across TPMT and HLA regions that are medically relevant. Here it is further able to even include SV in phasing, which is often challenging. To demonstrate the versatility of MethPhaser, we evaluated its performance on various human populations (HG01109, HG02080, and HG03098), as well as across blood samples from a cohort of patients with cardiovascular disease. In each case, MethPhaser is able to improve phaseblock N50 with methylation information. MethPhaser represents a novel approach that uses easily accessible nanopore methylation data to improve phasing and thus the interpretation of variant interactions across many medically important genes. In the future, we plan to investigate the utility of MethPhaser across cancer samples to obtain insights into oncogene regions. MethPhaser is open-source and available at https://github.com/treangenlab/methphaser.

Session Title: Epigenetics Poster Session II

PB2304 Methylome analysis of whole blood in ancestrally diverse Alzheimer's disease cohorts as a biomarker of disease

Authors:

T. Gu[1], L. Gomez[1], M. Mews[2], P. L. Whitehead[1], K. L. Hamilton-Nelson[1], M. Contreras[1], J. Sanchez[1], S. Tejada[1], L. D. Adams[1], P. R. Mena[1], T. Starks[3], C. Silva[4], M. R. Cornejo-Olivas[5], M. Illanes-Manrique[5], M. L. Cuccaro[1], J. M. Vance[1], W. S. Bush[2], G. S. Byrd[3], B. E. Feliciano-Astacio[4], J. L. Haines[2], G. W. Beecham[1], M. A. Pericak-Vance[1], A. J. Griswold[1]; [1]Univ. of Miami, Miami, FL, [2]Case Western Reserve Univ., Cleveland, OH, [3]Wake Forest Univ., Winston-Salem, NC, [4]Univ. Central Del Caribe, Bayamón, PR, [5]Inst. Natl. de Ciencias Neurologicas, Lima, Peru

Background: Alzheimer's disease (AD) is a neurodegenerative condition that is the most common form of dementia in the elderly. The disease process starts before the onset of symptoms and before the diagnosis can be made, making development of preventative therapies challenging. More than 30 genetic loci have been associated with AD; however, these loci can vary drastically between individuals of different genetic ancestral backgrounds. Thus, investigation into differences in risk markers enabling early intervention across diverse populations is critical. One such marker could be epigenetic features such as DNA methylation. **Methods**: As part of an ongoing study of the genetics and epigenetics of AD in diverse populations, we performed a methylome analysis in a cohort of 698 individuals. DNA from peripheral whole blood was analyzed using the Illumina MethylationEPIC 2.0 targeting more than 935,000 CpG sites. After filtering for quality and principal component outliers, 626 samples were included in the analysis. These consisted of both AD and cognitively intact (CI) individuals of European (68 AD, 67 CI), African (98 AD, 106 CI), or Hispanic (Puerto Rican - 85 AD, 76 CI; Peruvian - 41 AD, 41 CI; or Cuban - 22 AD, 22 CI) backgrounds. For quality control of methylation probes and downstream analysis, we utilized the SeSAMe R package. Probes with detection P-value > 0.05 in more than 5% of the samples were removed, as well as probes with low mapping quality or having a common SNP with MAF > 5%. We performed differential methylation analysis between AD and CI in the overall dataset and within each ancestral population. We used a linear model with covariate variables sex, age of exam, batch effect, global ancestry and estimated immune cell type proportions. **Results**: After quality control filtering, a total of 878,853 CpG sites were tested for differences between AD status. Overall, we identified 563 CpG sites with nominally significant differences (p-values <= 0.001) between AD and CI. Within each ancestral group, the number of differentially methylated sites differed: European - 442 sites, African - 217 sites, Hispanic - 475. Notably, however, these markers within the ancestral group did not overlap. **Conclusion**: Identifying the genes and biological pathways involved in AD is critical in the effort to develop effective therapies. Our analysis suggests the possibility of ancestry specific whole blood DNA methylation patterns as signatures of AD pathogenesis. Convergence of these methylation sites with existing genomic and transcriptomic data may reveal distinct genes but similar underlying pathological processes contributing to AD across individuals of diverse ancestries.

Session Title: Epigenetics Poster Session III

PB2305 Multigenerational impacts on DNA methylation signatures in autism spectrum disorder.

Authors:

G. E. Kuodza[1], Y. J. Ludena[1], J. S. Mouat[1], S. M. Hakam[2], R. Kawai[1], T. N. Sullivan[1], C. R. Torvick[1], D. H. Bennett[1], I. Hertz-Picciotto[1], J. M. Lasalle[1]; [1]Univ. of California Davis, Davis, CA, [2]Univ. of California, Berkeley, Berkeley, CA

The prevalence of Autism Spectrum Disorder (ASD) has been on the rise in recent years, and the underlying causes are not fully understood. The etiology of ASD is complex, as several hundred genes have been implicated as well as multiple environmental exposures. Epigenetic mechanisms, including DNA methylation, have emerged as a novel way to capture the complex gene by environmental interface of multivariate ASD etiologies. Past epidemiological investigations have suggested that grandparental factors, such as increased age or smoking, may increase the risk for ASD in the grandchild, but replication and molecular explanations for grandparental risk factors are currently lacking. One potential mechanism is multigenerational epigenetic inheritance (MEI), whereby DNA methylation marks on some loci are transferred across generations. Animal models have shown MEI in association with behavioral changes in the third and fourth generations, but this has not been studied in human cohorts. In order to investigate the potential impact of grandparental factors and MEI on the development of ASD, our study recruited participants from the CHARGE (Child Autism Risks from Genetics and the Environment) study, including grandparents, parents, and children (probands with ASD or typical development). We hypothesized that DNA methylation patterns associated with ASD in the probands may be detectable across generations and associated with grandparental environmental exposures. A questionnaire was used to gather information about the participants' exposure to environmental factors, including lifestyle exposures (smoking, drug use), medications, occupations, and other demographics. Biospecimens, including saliva samples collected from family members and newborn dried blood spots obtained from probands and parents from the California Newborn Registry, were collected. DNA was extracted from 349 saliva samples from 85 families and subjected to whole genome bisulfite sequencing (WGBS) to analyze DNA methylation. Sequence alignments and bioinformatic analyses are ongoing, utilizing DMRichR to identify individual genomic loci associated with ASD in each of the three generations and Comethyl to compare correlation patterns between methylation marks and selected variables, including grandparental exposures. Newborn blood spot collections of parents and probands are ongoing and will be used to identify potential ASD epigenomic signatures that are tissue and life-stage independent. This research will provide new insights into the increased prevalence and underlying etiology of ASD that should pave the way for future research in the field.

Session Title: Epigenetics Poster Session I

PB2306 Neuronal DNA double-strand breaks lead to chromosomal structural variations and 3D genome disruption in neurodegeneration

Authors:

V. Dileep[1], C. Boix[1], H. Mathys[1], A. Marco[1], G. Welch[1], H. Meharena[1], A. Loon[1], R. Jeloka[1], Z. Peng[1], D. Bennett[2], M. Kellis[1], L-H. Tsai[1]; [1]Massachusetts Inst. of Technology, Cambridge, MA, [2]Rush Alzheimer's Disease Ctr., Chicago, IL

Persistent DNA double-strand breaks (DSBs) in neurons are an early pathological hallmark of neurodegenerative diseases including Alzheimer's Disease (AD), with the potential to disrupt genome integrity. We used single-nucleus RNA-seq in human post-mortem prefrontal cortex samples and found that excitatory neurons in AD were enriched for mosaic gene fusions. Gene fusions were particularly enriched in excitatory neurons with senescence and DNA repair gene signatures. In addition, chromosomal structural variations and gene fusions were enriched in neurons burdened with DSBs in the CK-p25 mouse model of neurodegeneration. Neurons enriched for DSBs also had elevated levels of cohesin along with progressive multiscale disruption of the 3D genome organization aligned with transcriptional changes in synaptic, neuronal development, and histone genes. Overall, this study demonstrates the disruption of genome stability and the 3D genome organization by DSBs in neurons as pathological steps in the progression of neurodegenerative diseases.

Session Title: Epigenetics Poster Session II

PB2307 Novel Mitochondrial DNA Variations reveal insights from the Taiwanese Population

Authors:

T. Chou[1,2], P-M. Chien[2], P-X. Chen[2], T-H. Tseng[2], J. Hsu[3], P-L. Chen[1]; [1]Nation Taiwan Univ., Taipei city, Taiwan, [2]Graduate Inst. of Med. Genomics and Proteomics, Natl. Taiwan Univ., Taipei, Taiwan, [3]Natl. Taiwan Univ., Taipei City, Taiwan

Analyzing population mtDNA variation profiles can provide insights into the clinical significance, which plays a crucial role in energy homeostasis. Despite the availability of certain databases, there is a notable scarcity of data specifically for mitochondrial variants in East Asian populations. Furthermore, it is worth noting that the Taiwan Biobank does not encompass mtDNA variant information. To address this gap, our study utilized whole genome sequence data from 1,484 individuals in the Taiwan Biobank. We utilized the GATK Best Practices for SNP/Indel Variant Calling in Mitochondria (V.4.1.8.0) to identify mtDNA variants and their corresponding allele frequencies. We also classified subjects based on high-digit mtDNA sub-haplogroups. Among the Taiwanese population, we observed a total of 2,361 variants, with the most predominant haplogroup being M, followed by D and F. Comparing variant allele frequencies with gnomAD East Asian population revealed significant similarity between Taiwanese and East Asian populations. Notably, our analysis uncovered 77 novel variants in the Taiwanese population compared to the gnomAD v3.1 database. Furthermore, a comparison with MitoMap unveiled 63 novel single nucleotide variants (SNVs). We also made intriguing observations related to pathogenic variants. One individual out of 184 carried a confirmed pathogenic variant with a heteroplasmy level exceeding 10%. The m.1555A>G variant, known for its association with antibiotics-induced hearing loss, exhibited a relatively higher frequency in the Taiwan Biobank (0.204%) compared to gnomAD (0.112%) and MitoMap (0.141%). Furthermore, we identified a unique variant, m.4300A>G, associated with Maternally Inherited Cardiomyopathy (MICM), in a single individual from the Taiwan Biobank, which was not detected in either gnomAD or MitoMap. Our study provides valuable mtDNA variant data with allele frequencies among the Taiwanese population that supports the integration of mtDNA analysis alongside nuclear variants and facilitates diagnostic interpretation and research studies.

Session Title: Epigenetics Poster Session III

PB2308 PHF20 regulates starvation-induced autophagy by epigenetic control

Authors:

J. Kim[1], S. Park[2], S. Oh[2], S. Baek[2], D. Park[1]; [1]Ajou Univ., Suwon 16499, Korea, Republic of, [2]Seoul Natl. Univ., Seoul 08826, Korea, Republic of

Autophagy is a vital cellular process that eliminates unnecessary proteins and damaged organelles to maintain cellular balance. Defective autophagy can contribute to the development of serious human diseases, including type 2 diabetes, neurodegenerative disorders and cancer. To ensure optimal autophagic flux under cellular stress conditions, it is crucial to epigenetically regulate the transcription of autophagy components. In this study, we utilized genome-wide approaches and unveiled the role of PHF20, a member of the PHF family, in autophagy during glucose starvation. Initially, we observed a reduction of autophagy in Phf20−/− mouse embryonic fibroblasts (MEFs) compared to WT MEFs under both glucose and amino acid starvation conditions. To elucidate the role of PHF20 in autophagy at transcriptional level, we performed RNA-seq to compare the expression of autophagy genes between wild-type (WT) and *Phf20* KO MEFs under glucose starvation. By *k*-means clustering followed by functional analysis and gene set enrichment analysis, we revealed the genes related to the autophagic process were expressed in a PHF20-dependent manner. To investigate how PHF20 regulated autophagy at the epigenetic levels, we conducted ATAC-seq and explored the alteration of chromatin states in the absence of PHF20 during autophagy. We globally defined chromatin states by chromHMM and found that the non-promoter regions showed relatively strong dependency on PHF20 for starvation-induced chromatin opening. Moreover, we confirmed that H3K4me1 and H3K4me2, which were an active *cis*-regulatory element, increased at PHF20-dependent open chromatin in WT MEFs, but not in *Phf20*$^{-/-}$ MEFs upon glucose starvation.Furthermore, we also found that PHF20-dependent chromatin opening at enhancer regions was co-occupied by H3K36me2 and H3K4me1/2 in chromHMM. To further confirm the role of PHF20 as an epigenetic reader of H3K36me2, we tested the binding of GST-PHF20 Tudor 1 and 2 domains to various histone modifications including H3 and H4 methylation and acetylation. The peptide binding array and in vitro peptide binding assay revealed specific binding of PHF20 Tudor 1&2 to H3K36me2 as well as other di-methylated lysine peptides. Our findings suggest a novel insight into the role of PHF20 in regulating autophagy genes under glucose starvation. As an epigenetic reader and activator of autophagy genes, the recognition of H3K36me2 by PHF20 was essential to control the expression of autophagy genes. Moreover, these findings will be applicable in drug development research to target autophagy-related diseases.

Session Title: Epigenetics Poster Session I

PB2309 † Placental DNA methylation changes as predicting markers for gastroschisis

Authors:

M. Narechania[1], D. Jhala[2], U. Ratnamala[3], S. Nath[4], U. Radhakrishna[5,6]; [1]Gujarat Univ., Ahmedabad, India, [2]Sch. of Sci., Gujarat Univ., Ahmedabad, India, [3]Univ. school of Sci., Gujarat Univ., Ahmedabad, India, [4]Oklahoma Med. Res. Fndn., Oklahoma City, OK, [5]Green Cross Pathology & Molecular Lab., Ahmedabad, India, [6]Univ. of Pittsbourgh, Pittsbourgh, PA

Background: Gastroschisis, is the most common congenital abdominal wall defect in which babies are born with their abdominal organs outside of their bodies. It is primarily considered a complex multifactorial condition. Although the exact cause of gastroschisis is unknown, research suggests that genetic factors may contribute to the disease, though no study has examined epigenetic factors. This exploratory study aimed to identify mechanisms involved in gastroschisis pathogenesis that are epigenetically dysregulated. **Method:** We conducted a genome-wide methylation study (GWMS) of placental tissues using the Illumina Human Methylation EPIC BeadChip array on 24 placental tissues with gastroschisis and 24 age, ethnicity, and sex-matched controls. Statistical, bioinformatics, computational biology, and analysis were performed to distinguish gastroschisis from controls. Genes containing top differentially methylated CpGs (FDR $P<0.05$) were analyzed for pathway and network. **Results:** Significant alterations in cytosine nucleotide ('CpG') methylation were found at 7581 CpG loci in gastroschisis cases (FDR $p\leq0.05$) (AUC$\geq$0.75) including 6398 hypomethylated 1184 hypermethylated. Among these, 83 markers had high diagnostic accuracy (AUC$\geq$0.90) for gastroschisis detection, representing 83 discrete genes. Pathway analysis revealed the involvement of several dysregulated signaling pathways including Wnt/β-catenin signaling, Hippo signaling pathway, VEGF signaling pathway, and Ras signaling pathway, which are thought to play a role in gastroschisis pathophysiology.**Conclusions:** This is the first study that employed a robust assessment of placental global DNA methylation in several placental samples. The study provides substantial evidence that DNA methylation variation in placental DNA may contribute to gastroschisis, which opens the door to more accurate and early detection of the disorder.

Session Title: Epigenetics Poster Session III

PB2311 Regional principal components enhance cell type-specific detection of Alzheimer's disease-associated DNA methylation changes.

Authors:

T. Eulalio, D. Nachun, S. B. Montgomery; Stanford Univ., Stanford, CA

DNA methylation, a key regulator of gene expression, influences downstream physiological changes and is intricately associated with genetic variation and complex phenotypes. Despite abnormal methylation patterns being observed across various complex diseases, their mechanistic implications often remain elusive. We addressed this challenge using *in silico* cell type deconvolution to identify cell type-specific differentially methylated genes associated with genetic variants and Alzheimer's Disease (AD). Further, we introduce regional principal components (rPCs), a novel feature-level summary derived through principal components analysis (PCA), to improve gene-level methylation measurement. The rPCs enhanced the detection of AD-associated methylation changes, outperforming traditional CpG-level methylation and average-based methods, as evidenced by an over three-fold increase in significant differentially methylated features. We next explored the biological relevance of the observed methylation changes through cell type-specific methylation quantitative trait loci (meQTL) mapping and colocalization analysis with AD genome-wide association study (GWAS) data. This analysis identified 31 genes with significant evidence of shared causal variants regulating AD risk and feature-level methylation changes, implying a shared genetic basis for AD susceptibility and DNA methylation. There was strong evidence of colocalization between AD risk and methylation of the *AP2A2* gene in oligodendrocytes, supporting its link to AD pathogenesis. Notably, this colocalization was found using rPCs but not with averages. We conducted a transcriptome-wide association study (TWAS) to identify causal links between cell type-specific feature-level methylation and AD risk and integrated these results with colocalization using INTACT. We identified 27 loci, including genes like *RELB* in neurons and *DEF6* in oligodendrocytes, showing strong evidence of causal links between methylation and AD risk in specific cell types. In comparison to averages, rPCs identified a marginally higher number of putative causal genes and displayed increased cell type-specificity. In summary, our study demonstrates the effectiveness of rPCs for gene-level DNA methylation summarization, improving the detection of cell type-specific AD-associated methylation changes. By integrating genetic information through cell type-specific meQTL mapping, colocalization analysis, and probing potential causal pathways, we identified relationships between DNA methylation, genetic variation, and AD risk.

Session Title: Epigenetics Poster Session I

PB2312 Relationships between asthma biomarkers and immune cell DNA methylation profiles in a deeply phenotyped twin-family cohort

Authors:

A. Van Asselt[1], J. Beck[1], B. Johnson[1], C. Finnicum[2], N. Kallsen[1], L. Ligthart[3], J- . Hottenga[3], A-H. Maitland-van der Zee[4], S. Vijverberg[4], E. De Geus[5], D. Boomsma[3], E. Ehli[1], J. van Dongen[3]; [1]Avera McKennan Hosp. and Univ. Hlth.Ctr., Sioux Falls, SD, [2]Avera McKennan Hosp. & Univ. Hlth.Ctr., Sioux Falls, SD, [3]Dept. of Biological Psychology, Vrije Univ.it, Amsterdam, Netherlands, [4]Dept. of Pulmonary Med., Univ. of Amsterdam, Amsterdam, Netherlands, [5]Vrije Univ.it, Department of Biological Psychology, Vrije Univers, Netherlands

Asthma is a complex respiratory disease characterized by chronic airway inflammation and variable airflow obstruction. The etiology and pathogenesis of asthma involve intricate interactions between genetic and environmental factors, leading to extreme amounts of disease heterogeneity from both a phenotypic and molecular perspective. Due to this variability, investigating biomarkers associated with asthma may provide better clarity to its origins and drivers. Recently, DNA methylation has emerged as a crucial player in gene expression regulation and disease development. Understanding the relationship between DNA methylation and asthma biomarkers can provide insights into the mechanisms underlying the disease. The objective of this study was to investigate the relationships of DNA methylation patterns from whole blood samples (measured via the Illumina Infinium EPIC methylation array) in 341 individuals of twin families from the Netherlands Twin Register (MZ Twins = 76, DZ Twins = 95, Parents = 157, Siblings = 13). Participants were adult parents with their young adult twin offspring where at least 1 family member suffered from severe asthma. We studied baseline lung function, bronchial hyper-responsiveness, positive skin prick tests (SPT) to 11 allergens, IgE, positive specific IgE tests to four allergens, and eosinophil counts. In an exploratory analysis, we assessed the association of the measured endo-phenotypes to the top 10 principal components derived from the DNA methylation data, which showed correlations between principal component 2 and several measures of lung functionality including forced expiratory volume and total lung volume. The number of independent dimensions in the asthma endo-phenotype data was estimated by MatSpd. We then performed epigenome-wide association studies (EWAS) for four lung function measurements over 759,263 CpG's correcting for familial structure and adjusting for multiple covariates (sample cellular proportions, technical batch effects). After Bonferroni correction, these analyses identified 22 and 78 CpG's associated with peak flow and forced expiratory volume, respectively. Next steps for this study include performing additional EWAS's for each of the measured biomarkers and further investigation of these identified CpG's.

Session Title: Epigenetics Poster Session II

PB2313 Relative Telomere Length shortening and association with DNA promoter differential methylation in Thyroid Cancer.

Authors:

F. Jasmine[1], B. Aschebrook-Kilfoy[1], B. Pierce[1], M. Kamal[2], H. Ahsan[1], M. Kibriya[1]; [1]Univ. of Chicago, Chicago, IL, [2]The Lab., Dhaka, Bangladesh

Background: Telomere and telomerase regulation may contribute in cancer. Thyroid cancer (TC) is the most common malignancy of the endocrine system. In some study, telomerase reverse transcriptase (*TERT*) expression was detected in TC and some suggest *TERT* promoter mutation correlate with *BRAF V600E* mutation in TC. We have recently shown the association of *BRAF* mutation status and differential methylation in TC. In this study, we investigate (a) the change in Relative Telomere Length (RTL), a surrogate marker of telomere maintenance, in TC tissue compared to corresponding normal thyroid tissue and (b) if there is any association of RTL shortening and promoter methylation. **Material & Methods:** We included 40 differentiated TC patients (male=12, female=28). Twenty-nine of them had papillary thyroid carcinoma (PTC), 7 had follicular variant of papillary thyroid carcinoma (FVPTC) and 4 had follicular thyroid carcinoma (FTC). In paired TC tissue and surrounding healthy thyroid tissues from same patient, we measured the RTL using a Luminex-based assay. We also looked for the *BRAF V600E* mutation and genome-wide DNA methylation. **Results:** The RTL was measured as a ratio of intensity from telomeric region and that from a reference gene region. The RTL was found to be significantly shorter in the TC tissue compared to normal tissue (0.73 ±SD0.28 vs. 1.11 ±SD0.31, p= 2.01x10-10). This magnitude of RTL shortening (by ~0.38) in tumor tissue is equivalent to RTL shortening seen in human leukocytes over more than 20 years of aging measured by the same assay and is ~3 fold more marked than it is seen in colorectal carcinoma. The RTL shortening in TC tissue was seen irrespective of gender, age group, histological type, presence/absence of capsular invasion, lymph node involvement and BRAF mutation status. We divided the patients by the magnitude of RTL shortening (above and below median). The frequency of greater TEL shortening was higher among the *BRAF* wild type compared to *BRAF V600E* mutants (68% vs 28%, p=0.026). We compared the list of significantly differentially methylated loci (FDR 0.05) for both the groups. The genes differentially methylated only in patients with greater TEL shortening showed significant enrichment ($p<0.05$) in *MAPK* signaling, *ErbB* signaling, *Ras* signaling, thyroid cancer and thyroid hormone signaling pathways. However, none of the 109 promoter loci in telomere maintenance genes showed significant differential methylation. **Conclusion**: In TC tissue, there is significant RTL shortening compared to normal tissue from same patient. This shortening is associated with differential methylation changes in different pathways involved in cancer.

Session Title: Epigenetics Poster Session III

PB2314 † Reproducible human methylation variation fine-mapped to epigenetic haplotypes by 5-base HiFi sequencing

Authors:

E. Grundberg[1], B. Ge[2], W. Cheung[1], M. Lathrop[2], T. Pastinen[1]; [1]Children's Mercy Res. Inst., Kansas City, MO, [2]McGill Univ., Montreal, QC, Canada

Large-scale population profiling of DNA methylation (DNAme) by microarrays has identified genetic influences on epigenetic variation with evidence of overlap with genetic control of common diseases. However, the sparse coverage of CpGs on arrays prevents insight into coordinated patterns of DNAme variation as well as for the identification of genetically or environmentally controlled epigenetic landscapes. We present a comprehensive, high-resolution atlas of DNAme variation from ~3500 unrelated blood samples assayed across different platforms covering ~3.5 million CpGs: methylC-capture sequencing (MCC-Seq, N~2000), whole-genome bisulfite sequencing (WGBS, N~1000) and long-read 5-base HiFi genome sequencing (5-base HiFi-GS, N~500), respectively. We identify strong correlation of effects sizes from genetic association of single CpGs (mQTL, P<10-8) across short-read data sets (MCC-Seq vs WGBS: r=0.96) with less concordance when contrasting correlation of effects from significant mQTLs (P<10-8) detected by consortia (GoDMC) utilizing microarray (GoDMC vs MCC-Seq: r=0.75). We define a stringent orthogonally validated set of mQTLs requiring both effect size concordance and peak genetic signal (colocalization) concordance. Using these criteria, we identify ~350,000 CpGs in MCC-Seq that show same genetic signal in WGBS or 5-base HiFi-GS. Of note, only 7.5% of these curated genetically controlled CpGs overlap array-based mQTLs and concordance of genetic association among these is only 55% between sequencing and microarray mapping. In the MCC-Seq dataset we generate genome-wide DNAme correlation blocks (n~42,000, average CpGs/block=15) to clusters (~500bp) and demonstrating high reproducibly (>60%) using similar correlation matrix in independent WGBS. Around half of coordinated DNAme in the blocks is explained by local *cis*-regulatory genetic effect, whereas remainder is due to *trans*-acting and/or environmentally impact on coordinated DNAme. Finally, using the genetically controlled subset of DNAme coordination we utilize the 5-base HiFi GS data to develop high-resolution "haplotypes" of DNAme variation in long, continues reads (>12kb) to reveal interaction of DNA variation to local epigenetic and regulatory control where individual, constrained *cis*-regulatory sequences are strongly enriched. In summary, our layered approach to develop a validated map pf genetic and environmental impacts on population methylome reveals hundreds of thousands of previously uncharacterized links between methylome and DNA variation and uncovers their organization to biologically interpretable epigenetic features.

Session Title: Epigenetics Poster Session I

PB2315 Sex-specific DNA methylation marks in Rab-regulatory genes underly sex-biased risk of recurrence in unprovoked venous thromboembolism

Authors:

O. Bezerra[1], M. Rodger[2], M. J. Kovacs[3], G. Le Gal[4], P-E. Morange[5], G. Munsch[6], D-A. Trégouët[6], C. M. T. Greenwood[7], F. Gagnon[1]; [1]Univ. of Toronto, Toronto, ON, Canada, [2]McGill Univ., Montreal, QC, Canada, [3]Univ. of Western Ontario, London, ON, Canada, [4]Univ. of Ottawa, Ottawa, ON, Canada, [5]Aix-Marseille Univ., Marseille, France, [6]INSERM, Bordeaux, France, [7]Jewish Gen. Hosp., Montreal, QC, Canada

Deciding whether to stop oral anticoagulants beyond initial treatment (3-6 months) after an unprovoked venous thromboembolism (VTE) is challenging and controversial, partially due to an intriguingly higher risk of recurrence in men after therapy discontinuation compared to women. In preliminary work, we observed sex-specific blood DNA methylation (DNAm) marks in genes of the vitamin K cycle and its regulation (*VKORC1* and *UBIAD1*), relevant to the coagulation cascade and vascular integrity; vitamin K cycle plays a pivotal role in VTE biology. We hypothesized that sex-specific DNAm are associated with the observed sex-biased VTE recurrence (rVTE). Here, using the Illumina EPIC array, we performed a sex-stratified epigenome-wide association study (EWAS) with rVTE using peripheral blood samples of 417 European ancestry unprovoked VTE patients (REVERSE I study). Cox hazard regression was adjusted by age, principal components, and cell type heterogeneity. A follow-up methylation quantitative trait locus (meQTLs) analysis was performed within 250Kb (+/-) around the DNAm associated sites in the REVERSE I cohort. We identified two male (*TBC1D22B*-cg01060850 and *ZHX2*-cg07808424) and one female (*DENND3*-cg03401656) hypomethylated CpG sites associated with rVTE (P value < $7\times10-8$). Sensitivity analysis applying adaptations of the Cox model and repeating the EWAS while adjusting for other covariates revealed similar estimates of association. In a replication attempt, the female association displayed the same direction of effect in 139 French VTE women, although not significant (MARTHA study). *TBC1D22B* and *DENND3* are regulators of the Rab family proteins involved in vesicle trafficking, corroborating previous findings on the participation of Rab proteins - acting on exocytosis of Weibel-palade bodies - in VTE recurrence. Moreover, we identified common *cis*-meQTL variants known to modulate DENND3 methylation, also reported as eQTL (expression QTL) in whole blood by GTEx. Allele A carriers of the synonymous variant rs1045303 (G>A) significantly decreased DNAm of probe cg03401656 in VTE recurrent patients (Kruskal-Wallis, P value=0.004). Our results showing several independent DNAm sites associated with sex-specific rVTE may optimize decision-making on prophylaxis after a first VTE event. Further replication and functional analyses can expand insights into the molecular mechanisms driving sex-biased rVTE.

Session Title: Epigenetics Poster Session II

PB2316 Single-cell integrated RNA-sequencing and ATAC-sequencing in human motor cortex identifies cell-specific genetic drivers of ALS

Authors:

J. Cooper-Knock[1], S. Zhang[2], A. Weimer[3], E. Marie Monte[4], L. Jiang[4], X. Yang[4], T. Moll[5], K. Kenna[6], R. Highley[1], M. Snyder[7]; [1]Univ. Of Sheffield, Sheffield, United Kingdom, [2]Univ. Of Florida, Gainesville, FL, [3]Broad Inst., Boston, MA, [4]Stanford Univ., Stanford, CA, [5]Univ. of Sheffield, Sheffield, United Kingdom, [6]UMC Utrecht, Utrecht, Netherlands, [7]Stanford Dept. of Genetics, Palo Alto, CA

Amyotrophic lateral sclerosis (ALS) is an archetypal complex disease with a polygenic architecture. Despite high heritability, the latest ALS genome-wide association study (GWAS) identified a genetic risk factor in <10% of patients. Moreover, where the field has identified genomic regions containing disease-associated genetic variants, we often lack a biological understanding of the effect of the genetic variation, which is a necessary requirement to allow development of therapeutic interventions. We previously applied ATAC-seq to profile chromatin accessibility within normal motor neurons (MNs). By focusing on regulatory regions which are open and active within MNs, our machine learning method reduced the search space by >90% and uncovered 690 ALS risk genes, corresponding to 35% of SNP-based heritability. Here we have applied single-cell integrated RNA-sequencing and ATAC-sequencing to nuclei extracted from human motor cortex from ALS patients and controls. We have discovered cell-specific risk genes across neuronal subtypes and glia thereby increasing further the proportion of explained SNP-based heritability. Trajectory analysis of gene expression within microglia and astrocytes provides a data-driven method for discovering physiological glial phenotypes in vivo. Using this analysis we have identified ALS-associated glial expression of known ALS risk genes and provided previously missing functional annotations for known ALS GWAS loci.

Session Title: Epigenetics Poster Session III

PB2317 Single-nucleus chromatin accessibility variations implicate microglia and deep-layer excitatory neurons for major depressive disorder.

Authors:

A. Chawla[1], D. Cakmakci[1], W. Zhang[1], M. Maitra[2], M. Suderman[3], X. Zhang[1], Y. Li[1], C. Nagy[4], G. Turecki[5]; [1]McGill Univ., Montreal, QC, Canada, [2]Douglas Hosp., Montreal, QC, Canada, [3]Univ. of Bristol, Bristol, United Kingdom, [4]Douglas Mental Hlth.Univ. Inst., Verdun, QC, Canada, [5]Mc Gill Univ, Montreal, QC, Canada

Motivation: Major Depressive Disorder (MDD) is a debilitating psychiatric disorder and a major risk factor for suicide. However, its biological underpinnings are scarcely understood. While many brain cell-types have been associated with MDD, investigation of genetic and epigenetic factors holds promise to uncover a more coherent picture. To this end, MDD genome-wide association studies (GWAS) identified hundreds of genetic variants residing primarily in regions of non-coding genome whose functions remain to be determined. Single-nucleus ATAC sequencing (snATAC-seq) [1] provides an unprecedented opportunity for interrogation of such variants through the lens of accessible chromatin. **Objectives:** We aimed to identify cell-type specific accessibility differences in MDD and determine transcription factors (TFs) disrupted by accessibility variations influenced by MDD-related epigenetic and genetic factors. **Methodology:** Using snATAC-seq, we sequenced accessible regions of the genome in more than 200,000 nuclei extracted from dorsolateral prefrontal cortex of 84 MDD individuals or age- and sex-matched healthy controls. Next, we integrated snATAC-seq with snRNA-seq from matching subjects to identify gene-expression changes associated with MDD linked accessibility variations. Finally, we trained gapped-kmer SVM (gkmSVM) [3] and fine-tuned an attention-based deep learning model called DNABERT [4] to predict cell-type specific accessibility from genomic sequences, and pinpointed MDD SNPs causing significant allelic variations via in-silico mutation. **Results:** Majority of accessibility differences in MDD were found in grey matter microglia (58%) showing decreased binding of immune-regulatory TFs and demonstrating suppressed immune-functions. Additionally, NR4A2+ deep-layer excitatory neurons showed prominent accessibility variations (21%) and specifically enriched for MDD GWAS SNP-based heritability using LD score regression. Moreover, in-silico mutagenesis via sequence-based cell-type-specific classifiers revealed potential disruption of several neurodevelopmentally-important TFs in NR4A2+ excitatory neurons and a lineage-specifying TF in microglia. Finally, we identified significant differences in ligand-receptor interactions between these subtypes, implicating phagocytosis-related signaling pathways in MDD. **Conclusions:** Overall, we elucidate specific subtypes of neurons and microglia showing significant accessibility variations in association with MDD SNPs. 1. Buenrostro, J., *Nature* **523,** 486-490 (2015). 2. Lee, D. *Bioinformatics* **32**, 2196-2198 (2016). 3. Ji, Y. *Bioinformatics* **37**, 2112-2120 (2021)

Session Title: Epigenetics Poster Session I

PB2318 SMYD5 is a novel epigenetic gatekeeper of the mild hypothermia response

Authors:

S. Rafnsdottir[1], K. Jang[1], S. T. Halldorsdottir[1], A. Tomasdottir[1], M. Vinod[1], K. Möller[1], T. Reynisdottir[1], L. H. Atladottir[1], K. E. Allison[1], J. He[2], L. Zhang[3], F. J. Northington[3], R. Chavez-Valdez[3], K. J. Anderson[1], H. Bjornsson[1]; [1]Univ. of Iceland, Reykjavik, Iceland, [2]Michigan State Univ., East Lansing, MI, [3]Johns Hopkins Univ., Baltimore, MD

Physicians widely employ a strategy called targeted temperature management, which involves cooling patients to ~32°C to decrease the risk of neurological damage following asphyxia. Three genes (*SP1*, *CIRBP* and *RBM3*) are specifically upregulated at this exact temperature as part of the mild hypothermia response (MHR). Recent work suggests these genes mediate the neuroprotective effect of hypothermia. We have recently created mild hypothermia indicators (MHI) which translate transcriptional activity of the three key genes into fluorescent signals. Using these MHIs, we have shown that: 1) the MHR is specific as all show maximal activity at 32°C, with much less activity at other temperatures; 2) the MHIs mirror response at endogenous loci; 3) this is a generalized phenomenon even in cells of diverse origins; 4) the response is particularly strong in cells with neuronal characteristics including HEK293; and 5) the maximal response occurs after 6 hours for SP1-MHI, followed by RBM3- and CIRBP-MHIs, respectively. We have used the MHIs for two strategies: 1) screening an FDA-approved drug library on cells carrying SP1-MHI, which yielded a compound (Entacapone) that is able to activate the SP1-MHR at 37°C; and 2) a CRISPR-Cas9 genome-wide library screen which yielded multiple potential regulators of the SP1-MHR. The screen caused a dramatic shift in fluorescence of mutagenized cells, supporting the idea that there must be a number of factors upstream of SP1. One such candidate repressor from our screen was SMYD5, a histone lysine methyltransferase that places H3K36me3, H3K9me3 and H4K20me3 marks. In mouse embryonic stem cells (mESC) there is SMYD5 binding at *SP1* and *CIRBP* promoters (CUT&TAG) and upon knockout (KO) of SMYD5 in mESC there is an increase in *SP1* mRNA. Upon SMYD5 KO in HEK293 cells we see a significant increase of SP1 protein at 37°C and SMYD5 knockdown shows increased fluorescence of the SP1-MHR, supporting that SMYD5 acts as a repressor at euthermia. We also observe that SMYD5 levels show temperature dependence with less SMYD5 binding (CUT&RUN) and SMYD5 amounts *in vitro* and *in vivo* after 6 hours at 32°C. This repression is mirrored by temperature-dependent levels of H3K36me3 at the *SP1*-locus and globally indicating that the mammalian MHR is regulated at the level of histone modifications. We identified 45 additional SMYD5-bound temperature dependent genes suggesting a broader MHR-related role for SMYD5. Our study provides an example of how the epigenetic machinery integrates environmental cues into the genetic circuitry of mammalian cells and suggests novel therapeutic avenues for neuroprotection after catastrophic events such as neonatal asphyxia.

Session Title: Epigenetics Poster Session II

PB2319 Systematic genome-wide association scans across binary variables in the Canadian Longitudinal Study on Aging (CLSA).

Authors:

M. Kazemi, D. Taliun; McGill Univ., Montreal, QC, Canada

Systematic genome-wide association scans across binary variables in the Canadian Longitudinal Study on Aging (CLSA). The Canadian Longitudinal Study on Aging (CLSA) is a large 20-year follow-up cohort study on 50,000 residents of Canada aged 45-85 and recruited between 2010-2015. The participants underwent genome-wide genotyping from blood samples, and a large amount of phenotypic information was collected, including questionnaire data and physical measurements. In this project, we aim to perform systematic genome-wide association studies (GWASs) across all available phenotypes in this study and share generated summary-level association results with the broad research community for replication, validation, and other downstream analyses including the construction of polygenic risk scores and Mendelian randomization. Here we present the results from the first step in our project, where we focused on binary phenotypes. From the 765 baseline binary variables in the CLSA, we identified 35 corresponding to diverse health conditions, including neurological (e.g., multiple sclerosis, 178/25,029 cases/controls), mental health (e.g., depression, 4,170/20,976), cardiovascular (e.g., high blood pressure, 9,291/15,830), digestive (e.g., bowel disorder 2,453/22,708), endocrine (e.g., type 2 diabetes, 2,214/20,855), respiratory (e.g., asthma, 3,342/21,811), genitourinary (e.g., kidney disease, 706/24,461), sensory disorders (e.g., macular degeneration, 1,062/24,047), immune system conditions (e.g., allergies 9,606/15,432), and various types of cancers. For these phenotypes, we performed GWASs on 43,232,256 genetic variants imputed with the state-of-the-art TOPMed reference panel (imputation quality >0.3, minor allele count >20). We used the Regenie generalized mixed model association test to account for relatedness with sex, age and the first 20 principal components as covariates. At the moment of writing, the genotype data were available for 26,622 CLSA participants, and 25,254 of them were of European genetic ancestry based on our PCA analysis. Thus, we limited our GWASs to Europeans to maximize statistical power. We identified 1,502 statistically significant associations (P-value < 5×10-8) distributed across 23 independent loci. Our next steps include expanding GWASs to more binary phenotypes and setting up the public interactive browser for easy results lookups, comparisons and visualizations.

Session Title: Epigenetics Poster Session III

PB2320 Systems biology approach reveals intricate DNA methylation dynamics in skeletal muscle with age.

Authors:

A. Khamis[1,2], M. Boissel[1], L. Ning[1], L. Maurin[1], V. Raverdy[3], M. Derhourhi[1], F. Pattou[3], A. Bonnefond[1,2], P. Froguel[1,2]; [1]Inserm U1283, CNRS UMR 8199, EGID, Inst. Pasteur de Lille, Univ. of Lille, Lille, F-59000, France, Lille, France, [2]Dept. of Metabolism, Digestion and Reproduction, Imperial Coll. London, London, United Kingdom, [3]Univ. of Lille, Inst. Pasteur de Lille, Inserm U1190, EGID, CHU Lille, Lille, France

Maintaining human glucose homeostasis requires intricate complementary mechanisms, and skeletal muscle is the main regulator of glucose expenditure. We assessed whether this metabolic process is impacted by genetics and acquired factors, such as aging. Here, we assessed the effects of age on gene expression and DNA methylation in skeletal muscle biopsies from 84 individuals that underwent bariatric surgery. Using transcriptomics data, we performed a transcriptome wide association study (TWAS) we did not identify any gene expression changes associated with age (when adjusting for sex, body mass index [BMI] and ancestry). We then performed an epigenome wide association study (EWAS), adjusting for sex, BMI, ancestry and cellular composition, and identified 303 CpGs associated with age (bacon-corrected FDR < 0.05), of which ~ 60% have been previously identified in skeletal muscle. Consistent with other studies, none of the present CpGs were associated with nearby gene expression within ± 1 Mb window. We then hypothesised that genetic variants, known to influence DNA methylation, might be potential confounders of methylation-expression (ME) associations. Therefore, using genotyping data in the same individuals, we developed and applied a systems biology approach: ME were adjusted for genotype, which allowed the identification of robust functional CpGs associated with age and gene expression independently of individual's genotype. We removed SNPs associated with nearby gene expression (eQTLs) and/or nearby CpGs (mQTLs). Using this model, we found that 46, out of the 303 CpGs, were associated with a change in nearby gene expression, whereby the vast majority (87%) did not implicate the nearest gene. Pathway analysis of differentially expressed genes linked to the CpGs and aging revealed that one of the top pathways was the "glucose metabolic process", including key genes in glucose uptake: the down-regulation of INSR encoding the insulin receptor and TBC1D1; and the up-regulation of glucose and lipid metabolic enzymes TPI1, FADS3, and DCXR. In contract, a TWAS and EWAS with glucose and insulin measures (HbA1C, fasting glucose, fasting insulin and HOMA2-IR) revealed that none of these traits were associated with DNA methylation or gene expression changes. Therefore, our comprehensive, multi-omic pipeline revealed the importance of studying DNA methylation dynamics in their cis-regulatory context and highlights that the effects on glucose homeostasis in skeletal muscle is not directly influenced by glucose measures, but occurs indirectly through age.

Session Title: Epigenetics Poster Session I

PB2321 TDP-43 chronic deficiency leads to genome-wide dysregulation of the transcriptome and transposable elements by affecting R-loop and 5hmC crosstalk.

Authors:

Y. Hou[1], Y. Li[1], J-F. Xiang[1], K. Tilahun[1], J. Jiang[1], V. G. Corces[1], B. Yao[2]; [1]Emory Univeristy, Atlanta, GA, [2]Emory Univ., Atlanta, GA

TAR DNA-binding protein 43 (TDP-43) is an RNA/DNA-binding protein involved in numerous aspects of RNA metabolism, but its molecular roles in regulating gene and transposable elements (TEs) at the genomic level is not extensively explored. TDP-43 loss-of-function due to progressive cytoplasmic aggregations serves as a pathological hallmark and potential causality for several neurodegenerative diseases. Recent evidence suggest acute knockdown of TDP-43 affects the formation of R-loops, nuclear DNA:RNA hybrid structures implicated in transcription regulation. However, how stable and chronic functional perturbation of TDP-43, which more closely resembles age-related neurodegeneration, impact global transcriptome via R-loop dysregulation remains unclear. Here we demonstrate that stable and prolonged TDP-43 knockdown impairs cell proliferation and cell response to DNA damage. At the molecular level, TDP-43 chronic deficiency impacts gene expression either locally or distally by concomitantly altering the crosstalk between R-loops and 5-hydroxymethylcytosine (5hmC) in gene bodies and long-range enhancer-promoter interactions. Furthermore, TDP-43 knockdown induces substantial disease-relevant TE activation by influencing their R-loop and 5hmC homeostasis in a locus-specific manner. Taken together, our findings highlight the previously understudied genomic roles of TDP-43 in modulating R-loop/5hmC coordination in coding genes, distal regulatory elements, and TEs, and present a general molecular mechanism underlying the contribution of proteinopathies to the etiology of neurodegenerative disorders.

Session Title: Epigenetics Poster Session II

PB2322 Temporally specific single-cell epigenomic and chromosomal architecture reconfiguration in the developing human hippocampus and prefrontal cortex

Authors:

M. Heffel; UCLA, Los Angeles, CA

The human hippocampus (HPC) and prefrontal cortex (PFC) play critical roles in learning and cognition, yet the dynamic molecular characteristics of their development remain enigmatic. Here we investigated the epigenomic and 3D genome reorganization during the development of the HPC and PFC using more than 53,000 joint single-nucleus profiles of chromatin conformation and DNA methylation (sn-m3C-seq). The remodeling of DNA methylation predominantly occurs during late-gestational to early-infant development and is temporally separated from chromatin conformation dynamics. Neurons have a unique Domain-Dominant chromatin conformation that is different from the Compartment-Dominant conformation of glial cells and non-brain tissues. We reconstructed cell-type regulatory programs across developmental and differentiation trajectories, identifying temporal and lineage specific molecular markers idiosyncratic to each data modality. Integrating our data with previous GWAS studies, we find putatively causal common variants for schizophrenia strongly overlapping with chromatin loop-connected, cell-type-specific regulatory regions. Our data demonstrate that single-cell 3D-epigenomics is a powerful approach for dissecting neuropsychiatric risk loci.

Session Title: Epigenetics Poster Session III

PB2323 The atypical parent-of-origin effects dependent methylome is associated with the rate of aging in humans

Authors:

Y. Zeng[1], C. Gao[1], P. Navarro[2], A. McIntosh[2], C. Haley[2], R. Walker[3], C. Amador[2], A. Campbell[2], R. Madden[2], M. Adams[2], X. Bai[1], Y. Liu[1], M. Li[1], C. Hayward[2], D. Porteous[2], X. Shen[2], K. Evans[2]; [1]Sun Yat-Sen Univ., Guangzhou, China, [2]The Univ. of Edinburgh, Edinburgh, United Kingdom, [3]Univ. of Exeter, Exeter, United Kingdom

Aging at adulthood may be related to early events acting through the genomic regions that are both sensitive to early environments and related to health phenotypes in adulthood. The parent-of-origin-effect (POE)-regulated methylome includes regions targeted by genetically-controlled imprinting effects (the typical type of POE) and regions targeted by parental environmental effects (the atypical POE). The “early event/environment sensitive nature” of this part of the methylome making it a potential hub connecting early exposures, the epigenome and aging. In this study, a phenome-wide association analysis is performed for the POE-influenced methylome using GS:SFHS ($N_{discovery}$=5,087, $N_{replication}$=4,450). A total of 92 significant POE- CpG - phenotype associations are statistically identified and replicated. The POE-CpGs belonging to the atypical class contribute most of the associations and the most enriched and strongly associated phenotypes are aging (DNAmTL acceleration), intelligence and parental (maternal) smoking exposure phenotypes. The atypical- POE-CpGs could also form co-methylation networks (modules) which are associated with these phenotypes, one of the modules associated with aging shows an elevated level of methylation connectivity as age increases. We also reported additional aging-related features of the atypical POE-CpGs: high levels of methylation entropy, fast entropy loss with age and a high correlation with constituent CpGs of epigenetic clocks. These findings uncover the association between the part of methylome influenced by atypical-POE and the aging process, offering new evidence supporting the "early development of origin" theory for aging in humans.

Session Title: Epigenetics Poster Session I

PB2324 The methylome and transcriptome of maturing iPSC-derived retinal 3D organoids over a 2-year period of culturing

Authors:

C. Kiel[1], T. Strunz[1], P. Berber[1], B. H. F. Weber[1,2]; [1]Univ. of Regensburg, Regensburg, Germany, [2]Univ. Hosp. Regensburg, Regensburg, Germany

Purpose: 3D *in vitro* organoids have become popular in recent years as they provide complex and highly specialized model systems structurally most often related to the native tissue. Whether such models indeed also reflect a comparable functional context during developmental maturation processes remains unclear at present. Specifically, the context of DNA methylation and global RNA expression is of interest to better understand the possible use and the optimal time point to investigate normal and diseased states of an organoid. Here, we focused on retinal organoids (ROs) as a model system to clarify methylation pattern and RNA expression profiles over an extended period of culturing time.
Methods: ROs from three unrelated donors were analyzed at different time points and compared with profiles of naïve human retinae. For methylation analysis (Human MethylationEPIC BeadChip) ROs were harvested after 0.5 years (n = 19, 3 donors), 1 year (n = 11, 3 donors), 1.5 years (n = 5, 1 donor) and 2 years (n = 5, 1 donor) of cultivation and compared with 12 human retinae (mean age 55 years). Furthermore, transcriptomes (RNA-Seq) were generated and compared for 9 human retinae (mean age 57 years), as well as ROs with a cultivation time of 0.5 years (n = 6, 2 donors), 1 year (n = 21, 3 donors), 1.5 years (n = 11, 2 donors) and 2 years (n=6, 1 donor).
Results: PCA analysis of methylation data including human primary tissues such as PBMCs or fibroblasts, iPSCs and iPSC-derived cell lines (RPE and endothelial cells) highlighted clustering of ROs and human retinae. Interestingly, ROs can be distinguished on the basis of their methylation and gene expression pattern by cultivation time. While the three later time points (1 year, 1.5 and 2 years) differ only marginally and overall showed most similarity with adult retinae, ROs at the age of 0.5 years are significantly distinct. Overall, with an increasing cultivation time methylation increases, while gene expression decreases accordingly. Enrichment analysis of differentially expressed genes as well as methylated regions displayed most significant differences for pathways participating in anatomical structure development and visual perception.
Conclusion: During a 2-year maturation process the methylome and transcriptome of ROs converge to human adult retinae. Notably, our data suggest that the maturation process of ROs is almost completed after about 1 to 1.5 years while further changes in methylation and transcription, specifically of many visual perception genes are slowly but constantly occurring over time.

Session Title: Epigenetics Poster Session II

PB2325 The reninness score: integrative analysis of multi-omic data to define renin cell identity.

Authors:

B. Xue[1], S. Medrano[1], J. Smith[1], A. Martini[2], N. Sheffield[1], M. Sequeira Lopez[1], R. Gomez[1]; [1]Univ. of Virginia, Charlottesville, VA, [2]Univ. of Brasília, Brasilia, Brazil

BACKGROUND: Renin cells are crucial for survival. They are well-known for their function in blood pressure regulation, fluid-electrolyte homeostasis, kidney development, and tissue morphogenesis. In the fetal kidney, renin cells are progenitors for multiple cell types that retain epigenetic memory of the renin phenotype in adults. Renin cell descendants regain the endocrine renin phenotype to overcome homeostatic threats, a process known as recruitment. The determinants that control the identity, fate, and recruitment of renin cells are not well understood. **We aim to 1) define the chromatin pattern that determines renin cell identity; and 2) develop a computational tool to score samples by similarity to that pattern.**

RESULTS: We compared open chromatin regions (ATAC-seq) between non-renin-expressing and renin-expressing samples from three different sources and stimulation states: 1) primary renin cells in basal state (WT); 2) primary renin cells chronically stimulated to produce renin (recruited); and 3) constitutively renin-expressing tumoral cells. Differential analyses between the three types of renin cells and the non-renin cells revealed 1,525 unique open chromatin regions shared in at least two renin groups, including regions associated with renin cell-specific genes: *Ren1* and *Akr1b7*. Motif enrichment analyses showed that the bZip family was the most enriched motifs in the tumoral cells, and MEF2 family was the most enriched in primary cells, which aligns with our previous research. Co-enriched motifs found in *Ren1* and *Akr1b7* regions suggest a group of TFs govern their co-expression. In addition, we identified unique open chromatin regions and enriched motifs in recruited renin cells, suggesting associated TFs involved in recruitment.

Next, we used a machine-learning approach to calculate a "reninness" score by training a genomic region model with renin and non-renin ATAC-seq data. We identified a cluster of regions (n = 112,992) around the trained renin label, which included 89% of the increased-accessibility regions from the differential analyses. We calculated the reninness score based on the distance between the trained label embedding and sample embedding using the trained model. The results reflected the known renin transcription level: the homogeneous WT renin cells have the highest score, followed by the homogeneous renin-expressing tumoral cells and the renin-expressing primary cells. Conversely, non-renin-expressing cells have the lowest scores.

SUMMARY: In conclusion, we identified the chromatin configurations defining the identity and recruitment of renin cells and developed a computational score to quantify it.

Session Title: Epigenetics Poster Session III

PB2326 The role of exogenously induced mitochondrial DNA variation on the nuclear epigenome and transcriptome

Authors:

P. Win[1], S. Zeng[2], J. Xie[2], C. Newcomb[2], D. E. Arking[2], C. A. Castellani[1,2]; [1]Western Univ., London, ON, Canada, [2]Johns Hopkins Univ Sch. of Med., Baltimore, MD

Background: Mitochondrial DNA copy number (mtCN) reflects the quantity of mtDNA and serves as a proxy for mitochondrial (mt) function. Exogenous environmental stimuli leads to mtCN variation, which in turn, has been shown to influence nuclear DNA (nDNA) methylation and gene expression. We show that exogenous environmental chemical alteration of mtCN modfies nDNA methylation and gene expression in a dose-dependent manner and is often reversible. Characterization of these relationships allows for the identification of mtCN associated CpG sites and genes as well as their underlying biological mechanisms.
Methods: Exogenous mtCN altering chemical compounds were applied to HEK293T cells in culture, these included resveratrol (Resv), acetaminophen (Acet) and ethidium bromide (EtBr) at multiple doses (N=111). EtBr removal following exposure for four recovery timepoints was also tested. DNA and RNA was extracted, and methylation was profiled via the Illumina Methylation EPIC Beadchip. RNA sequencing was performed on the Illumina NovaSeq. Linear mixed and negative binomial regression analyses were used to assess the relationship between mtCN and methylation/gene expression, with nDNA measures as the dependent variable. Integrated analysis between differentially methylated and expressed genes was performed to determine significant CpG-Gene pairs. *In silico* enrichment analyses was performed to identify relevant shared pathways.
Results: Acet and Resv exposure resulted in dose-dependent mtCN increases, and EtBr led to dose-dependent mtCN decrease. 332 CpGs were associated with mtCN ($p<1e-7$). 2686, 1895 and 170 genes were differentially expressed after Acet, Resv and EtBr treatment, respectively ($p<1e-6$). EtBr recovery CpGs/gene variation correlated with mtCN changes and in many cases, mtCN, nDNA methylation, and gene expression recovered to initial levels within 96 hours of drug withdrawal. Integration of methylation and gene expression identified 423 pairs including associations between cg0918651 and nearby *NUDT3*, *UQCC2*, and *C6orf1* ($p<6e-9$). Voltage-gated ion channel activity ($p<6e-7$), and RHO GTPase cycle ($p<4e-6$) pathways were overrepresented in integrated analyses.
Conclusion: Environmentally induced changes in mtCN levels correlate with nDNA methylation and gene expression changes suggesting that the mechanism driving these changes is not independent. Further, we identify that in many cases, these changes are reversible to baseline mtCN and methylation/gene expression levels and are overrepresented by ion channel and cell cycle related pathways.

Session Title: Epigenetics Poster Session I

PB2327 Tissue-specific DNA methylation signatures of alcohol use disorder in human brain.

Authors:

M. Minto[1], C. Willis[1], B. Quach[1], S. Han[2], R. Tao[2], A. Deep-Soboslay[2], B. T. Webb[1], R. D. Mayfield[3], E. O. Johnson[1], J. E. Kleinman[2], D. B. Hancock[1], L. J. Bierut[4], J. D. White[1]; [1]RTI Intl., Research Triangle Park, NC, [2]Lieber Inst. for Brain Dev., Baltimore, MD, [3]Univ. of Texas at Austin, Austin, TX, [4]Washington Univ Sch Med, St Louis, MO

Alcohol Use Disorder (AUD) is a common psychiatric disorder in the United States and has a growing number of associated genetic loci. SNP-based h^2 for AUD is 5-10%, and long-term alcohol use has been associated with epigenomic-driven neuroinflammatory responses that may promote reward in addiction. Previously, we have shown differential DNA methylation (DNAm) at CpG sites across the genome in the nucleus accumbens (NAc) and dorsolateral prefrontal cortex (DLPFC) of AUD cases and controls. We observed high heterogeneity for many of the differential DNAm sites, suggesting the potential for both tissue-specific and shared associations with AUD. To help tease apart tissue-specific vs. shared associations, we used a linear mixed modeling approach on Illumina EPIC methylation array data from 115 decedents (58 cases with 2+ DSM-5 AUD symptoms, 57 controls with no AUD symptoms) of European ancestry to simultaneously model within-subject differences in DNAm across brain tissues and between-subject differences in AUD status. DNAm array data were pre-processed separately for NAc and DLPFC with 768,826 CpGs common across both regions. While no CpGs surpassed FDR < 0.05, at nominal $p < 0.001$, we identified 171 CpGs associated with AUD with shared effects across brain regions and 542 CpGs with evidence of interaction between brain region and AUD status, i.e., tissue-specific associations. Differentially methylated regions were identified using comb-p and ipDMR, which revealed a tissue-specific AUD-associated DMR near *TRIM10* which is important in cytokine signaling and immune response to inflammation. Future studies would benefit from a larger sample size with more ancestrally diverse representation to detect robust, statistically significant associations. Nonetheless, these findings suggest that distinct DNAm sites may influence the neurobiological state of AUD status within and across addiction-relevant brain regions.

Session Title: Epigenetics Poster Session II

PB2328 Tissue-specific epigenetic differences in adults born preterm with very low birth weight compared with their same-sex. siblings

Authors:

H. Hauta-alus[1,2,3,4], J. Ronkainen[1], J. Kuula[2,3], D. Czamara[5], A. Heiskala[1], S. Sandboge[6,2], J. Björkqvist[2], N. Kaseva[2], K. Räikkönen[4], K. Pietiläinen[4], S. Sebert[1], E. Binder[5], E. Kajantie[7,1,8]; [1]Univ. of Oulu, Oulu, Finland, [2]Finnish Inst. for Hlth.and Welfare, Helsinki, Finland, [3]Univ. of Helsinki and Helsinki Univ. Hosp., Helsinki, Finland, [4]Univ. of Helsinki, Helsinki, Finland, [5]Max-Planck-Inst. of Psychiatry, Munich, Germany, [6]Univ. of Tampere, Tampere, Finland, [7]Natl. Inst. for Hlth.and Welfare, Helsinki, Finland, [8]Norwegian Univ. of Sci. and Technology, Trondheim, Norway

Preterm birth and very low birth weight (VLBW; ≤1500 g) increase the long-term risks for poor health and chronic diseases, however, the mechanisms remain unknown. One suggested pathway is epigenetic modifications such as DNA methylation (DNAm), but few studies have assessed other tissues than blood. We examined whether DNAm in blood or fat differs between VLBW adults and their siblings. The current analysis is based on the Adults Born Preterm Sibling Study and included 78 adults born preterm (<37 gestational week) with VLBW with 76 same-sex sibling-controls born at mean gestational age 40 weeks with mean birth weight of 3361g, all born between years 1976-1996. Epigenome-wide DNAm at cytosine-guanine dinucleotide (CpG) sites were examined from blood and fat tissue by Illumina EPIC 850K at mean age of 29 years. Mixed model was conducted adjusting for age, sex, batch, estimated cell composition and maternal smoking. None of the DNAm differences in blood attained epigenome-wide significance. In fat tissue 440 CpG sites were differentially methylated with false discovery rate [FDR] $p<0.05$ and 86 CpG sites with epigenome-wide significance of $p<9.4 \times 10^{-8}$ between VLBW and their sibling-controls after adjustments. Top sites were annotated to genes related to fat metabolism (e.g. *FADS2*, cg00264176: coefficient (standard error) 0.08 (0.01), FDR $p=9.1 \times 10^{-15}$; *ACSL3*, cg14157824: 0.05 (0.01), $p=6.1 \times 10^{-11}$) and neural development (e.g. *KIF26A*, cg08277679: 0.05 (0.01), $p=1.7 \times 10^{-12}$; *QPRT*, cg05127221: 0.06 (0.01), $p=4.9 \times 10^{-11}$) among other functions. Our results suggest tissue-specific DNAm differences in VLBW adults compared with their siblings with pronounced associations in fat tissue rather than blood. The differentially methylated CpG sites observed were related to genes that have central roles in fatty acid and cholesterol metabolism, cell growth and nervous system development indicating possible pathways for adverse metabolic disturbances of VLBW adults.

Session Title: Epigenetics Poster Session III

PB2329 Transcription factor activation induces re-replication and site-specific copy gain of the interferon receptor cluster

Authors:

J. Wilson, M. P. Ludwig, L. Porritt, J. C. Black, K. D. Sullivan; Univ. of Colorado Anschutz Med. Campus, Aurora, CO

Transient site-specific copy gains (TSSGs) are an evolutionarily conserved phenomenon in which specific genomic loci undergo targeted re-replication during S phase in response to stress conditions or alterations in the local epigenetic environment prior to resolution in G2. Furthermore, TSSGs are frequently found in repetitive gene families and regions amplified in tumors, suggesting a potential role as predecessors to inherited copy number alterations. However, the connection between the cellular stress signaling and generation of the TSSG remains unclear. Here, we report a novel TSSG locus at the interferon receptor (*IFNR*) cluster on chromosome 21, as well as the identification of the signaling cascade leading to the induction. We stimulated cells with various activators of the innate immune response, screened for copy-number gains using fluorescent in-situ hybridization (FISH), and determined that, regardless of stimuli, cells underwent rereplication of the *IFNR* locus at a rate of 10-15%, consistent with other TSSG. To identify the cause of this induction, we analyzed the induced pathways that were conserved across tested stimuli at both the RNA and protein level and identified three putative pathways. Next, we used a combination of pharmacological and genetic experiments to determine the components of innate immune activation required to generate TSSGs at the *IFNR* cluster. The results of these experiments identified a key transcription factor that is both necessary and sufficient for TSSG induction. Cut & Run sequencing analysis for this transcription factor under two stress conditions sufficient to induce TSSGs revealed a strong peak directly upstream of the *IFNR* cluster in both. Furthermore, pharmacological inhibition of the pathway upstream of the transcription factor prevents this binding and suppresses TSSGs. Together, these data suggest that activation and binding of this transcription factor to the locus is required for TSSG generation at the *IFNR* cluster.

Session Title: Epigenetics Poster Session I

PB2330 Transcriptional and Epigenetic Consequences of Early-life Seizures

Authors:

S. Townsend[1], J. J. Westfall[1], S. Sran[1], T. A. Bedrosian[2]; [1]The Ohio State Univ., Columbus, OH, [2]Nationwide Children s Hosp., Columbus, OH

Early-life seizure neuronal activation induces brain plasticity and long-term behavioral changes. In particular, the developing brain is highly plastic compared to the adult brain. Neuronal activity is known to remodel the epigenome and alter gene expression. Yet, little work has been done to correlate transcriptional and epigenetic changes in distinct neuronal populations of the developing brain to behavioral phenotypes. In this proposal, we will bridge this gap in knowledge by using single-cell multiomics techniques to investigate epigenetic and transcriptional changes observed in developing hippocampal neuronal populations post-seizure induction. Widespread neuronal activation can be reliably induced using pentylenetetrazol (PTZ), a GABA-A receptor antagonist that rapidly induces generalized seizures by suppressing the function of inhibitory synapses. I will assay hippocampal and cortical cells for gene expression and chromatin accessibility (10x Genomics Multiome) at multiple timepoints (1-hr, 1-day, 1-month) post-seizure induction to identify differentially expressed genes (DEGs) and differentially accessible chromatin regions (DARs) within cell types. So far, we have processed and begun to analyze data from our acute seizure paradigm where mice were exposed to PTZ or Saline on postnatal day 7. Following seizure induction, brains were harvested at the 1hr and 24hr timepoints. The hippocampus and surrounding cortex were harvested from each brain at either 1hr or 24hr post-injection, giving rise to four experimental groups: Saline 1hr, Saline 24hr, PTZ 1hr, and PTZ 24hr. Tissue from multiple mice were pooled into one aggregate sample per experimental group and processed using the 10x Genomics Single-Cell Multiome ATAC+ Gene Expression kit. Libraries were sequenced on an Illumina NovaSeq 6000 and processed via the Cellranger-ARC pipeline according to manufacturer recommendations. Downstream analysis was performed following standard workflows in Seurat and Signac. Mice exposed to PTZ had increased neuronal activity as denoted by increased Immediate Early Gene (IEG) expression in the treated group at the 1hr timepoint. From the go enrichment analysis, we observed changes in key pathways associated with synapse organization, sodium ion transport, and cytoplasmic translation. I am now analyzing chromatin accessibility data and expect to see changes in transcription factor binding that correlate to the changes in biological processes observed from the go-analysis.

Session Title: Epigenetics Poster Session II

PB2331 Transcriptional networks in placenta and fetal brain reveal extensive sex differences in a mouse model of polychlorinated biphenyl mixture exposure.

Authors:

K. Chau, K. Neier, A. Valenzuela, R. Schmidt, P. Lein, J. Lasalle; Univ. of California, Davis, Davis, CA

Neurodevelopmental disorders (NDDs), including autism spectrum disorders, affect as many as 1 in 10 children born in the United States, with males being at highest risk. Because complex interactions between genetic, hormonal, and environmental factors contribute to the risk of NDDs especially during prenatal life, understanding dysregulated gene pathways affected by multiple risks is important. In an NDD enriched-risk human prospective study “Markers of Autism Risk in Babies: Learning Early Signs” (MARBLES), detectable concentrations of multiple polychlorinated biphenyls (PCB) congeners in maternal gestational blood samples were associated with adverse neurodevelopmental outcomes and alterations to methylation patterns of neurodevelopmental genes in term placenta. In a mouse model of prenatal exposure to this mixture of PCB congeners observed in the MARBLES cohort, there were enhanced repetitive behavior and significant deficits in sociability and ultrasonic vocalizations at the lowest dose (0.1 mg/kg) specifically in males. To understand the molecular genetic basis of sexual dimorphism and dosage effects in this human-relevant PCB risk model, we exposed C57Bl/6J females to varying doses (0, 0.1, 1, 6 mg/kg) of the humanized PCB mixture prior to timed mating. At embryonic day 18, brain and placenta samples were collected for RNA sequencing and Whole Genome Bisulfite Sequencing (WGBS). Differential expressed genes (DEG) analysis of PCB effects in RNA-seq data using limmaVoom showed sex-specific dose responses, with females uniquely showing non-monotonic dosage effects that were greatest at 0.1 mg/kg, but males showing more DEGs overall. A weighted gene co-expression network analysis (WGCNA) was applied to investigate groups of highly correlated genes affected by PCBs and revealed that broad pathways of cellular and nucleotide metabolism including DNA repair are disrupted with exposures to PCBs and show sexual dimorphism and non-monotonic dosage effects in both placenta and fetal brain. Comethyl, a novel pipeline developed by the lab for weighted gene correlation network analysis of WGBS data, is in progress to identify comethylated regions correlated with PCB exposures. Lastly, a multi-omics integration of gene expression and DNA methylation will be performed to identify epigenomic signatures of gene expression and DNA methylation in the placenta that will be compared to human NDDs for their predictive value.

Session Title: Epigenetics Poster Session III

PB2332 Transcriptome and acetylome profiling identify crucial steps of neuronal differentiation in Rubinstein Taybi syndrome

Authors:

J. Van Gils[1], S. Karkar[2], A. Barre[2], S. Nothof[3], R. Chevalier[3], N. Broucqsault[3], C. El Yazidi[3], P. Fergelot[1], F. Magdinier[3], D. Lacombe[1]; [1]Dept. of Med. Genetics, Univ. Hosp. of Bordeaux and INSERM U1211, Univ. of Bordeaux, Bordeaux, France, [2]Bordeaux Bioinformatic Ctr. CBiB, Univ. of Bordeaux, Bordeaux, France, [3]Marseille Med. Genetics, INSERM U 1251, MMG, Aix Marseille Univ., Marseille, France

Rubinstein-Taybi syndrome (RTS) is a rare and severe genetic developmental disorder characterized by multiple congenital anomalies and intellectual disability. *CREBBP* and *EP300*, the two genes known to cause RTS, are transcriptional coactivators that possess a catalytic lysine acetyl transferase (KAT) domain and play a major role in chromatin remodeling. Loss of CBP or p300 function results in a deficit in KAT activity. Altered histone acetylation leads to inappropriate modulation of gene activity, in particular during development. In RTS patients, nothing is known on the functional link between histone acetylation and transcriptomic profiles during neuronal differentiation. To address this question, we derived induced pluripotent stem cells (hiPSCs) from RTS patients carrying a recurrent *CREBBP* mutation known to inactivate the KAT domain. We differentiated these cells into cortical and pyramidal neurons. We compared their acetylome by LS-MS/MS and transcriptome by RNA-seq to cells derived from healthy donors at three different timepoints during neuronal differentiation. No morphological differences were observed throughout the neuronal differentiation procedure between RTS patients and controls, but maturation appeared as impacted at the molecular level. We identified 25 specific acetylated histone marks that are specifically altered in RTS. All four core histones are regulated by CBP but H2B and H3 sites are the most impacted. Altered acetylation occurs at a very early stage and is maintained throughout neuronal differentiation for the majority of these 25 specific histone marks. By transcriptomic analyses, we identified the transition between neural progenitors and immature neuron as a critical step during the differentiation process, with an increase of differentially expressed genes (DEGS) in RTS compared to controls. More specifically, these analyses showed an impact on genes involved in neuronal differentiation and brain development, illustrating a delayed maturation process in RTS. The identification of pathophysiological mechanisms, particularly at the neuronal level in RTS, remains a major challenge for a better understanding of the cognitive impairment in patients. Our data highlight specific histone acetylation changes and a crucial phase of differentiation as a hallmark of RTS neuronal development. Overall, this opens new perspectives in the definition of epigenetic biomarkers for RTS, whose methodology could be extended to all chromatinopathies.

Session Title: Epigenetics Poster Session I

PB2333 Understanding complex trait susceptibilities and ethnical diversity in a sample of 4,145 Russians through analysis of clinical and genetic data

Authors:

D. Usoltsev[1], N. Kolosov[1], O. Rotar[1], A. Loboda[1], M. Boyarinova[1], E. Moguchaya[1], E. Kolesova[1], A. Erina[1], K. Tolkunova[1], V. Rezapova[1], O. Melnik[1], O. Freylikhman[1], N. Paskar[1], A. Alieva[1], E. Baranova[1], E. Bazhenova[1], O. Beliaeva[1], E. Vasilyeva[1], S. Kibkalo[1], R. Skitchenko[1], A. Babenko[1], A. Sergushichev[2], A. Dushina[3], E. Lopina[3], I. Basyrova[3], R. Libis[3], D. Duplyakov[4], N. Cherepanova[4], K. Donner[5], P. Laiho[6], A. Kostareva[1], A. Konradi[1], E. Shlyakhto[1], A. Palotie[5], M. Daly[7], M. Artomov[8]; [1]Almazov Natl. Med. Res. Ctr., St. Petersburg, Russian Federation, [2]ITMO Univ., St. Petersburg, Russian Federation, [3]Orenburg State Med. Univ., Orenburg, Russian Federation, [4]Samara State Med. Univ., Samara, Russian Federation, [5]Inst. for Molecular Med. Finland FIMM, Helsinki, Finland, [6]Finnish Inst. for Hlth.and Welfare, Helsinki, Finland, [7]Massachusetts Gen. Hosp., Boston, MA, [8]Nationwide Children's Hosp., Columbus, OH

Introduction: Linking inherited DNA variation to the disease risks is one of the main goals in modern predictive medicine. Large-scale projects such as the UK Biobank, FinnGen and Biobank Japan have made a substantial contribution to the understanding of human biology and the advancement of personalized medicine. The growing ethnic diversity of genetic studies resulted in the discovery of population-specific susceptibility loci that could not be identified with studies limited to a single ancestry. Inclusion of previously underreported populations into biobank initiatives improves fine-mapping accuracy of previously identified GWAS signals and novel risk gene discovery efforts. **Methods:** A cohort of 4,800 residents of three areas in Russia - St. Petersburg (N=1,600), Orenburg (N=1,600) and Samara (N=1,600) were recruited in 2012-2013 through an ambulatory visit to local hospitals and polyclinics. Additional independent cohort of 138 samples were recruited in 2017-2018 for participation as controls in a local study of early childhood starvation effects. DNA was extracted from blood samples of 4,723 individuals and genotyped using a custom FinnGen ThermoFisher Axiom microarray. Additional genotypes were imputed using Haplotype Reference Consortium (HRC) data. Directly genotyped and imputed variants with DR2 > 0.8 were included in the downstream analysis. Final dataset included 4,145 individuals and 11,077,763 variants. **Results:** We identified multiple distinct subpopulations within the Russian population, displaying notable genetic affinities towards both Finns and Asians. Furthermore, we observe a substantial prevalence of unique Finnish variants throughout Russia and Asia, indicating their likely origin as common ancestral Finno-Ugric variants. Consequently, we establish the feasibility of replicating these distinctive Finnish variants using a relatively modest cohort of Russian individuals. Finally, we provide access to allele frequencies and genetic associations in the Russian population across 465 phenotypes through the Biobank Russia portal. We present replication of many previous associations and report new significant results in smoking initiation, blood pressure increase during pregnancy and abdominal obesity. **Conclusion:** In this study, we present the first analysis of a diverse population spanning across many admixed ancestries. Our comprehensive analysis encompasses over 465 distinct phenotypes, providing a foundational framework for understanding origins of Finno-Ugric populations and transferability of genetic risks between European and Asian populations.

Session Title: Epigenetics Poster Session II

PB2334 Unique genome wide methylation patterns within HIV disease

Authors:

V. Ramsuran, T. Aurumugam, W. Smidt; Univ. OF KWAZULU-NATAL, Durban, South Africa

Despite more than 40 years of research, the human immunodeficiency virus (HIV) is still a major health concern. Advances in the field have helped us understand various aspects of the virus, and this has led to the development of antiretroviral therapy (ART), which has extended life expectancy and it is starting to decrease new infections. However, getting rid of HIV with ART treatment in a large epidemic, such as in South Africa, is a costly and complex activity. In addition, drug resistance levels are increasing at alarming rates. New treatment options, such as dolutegravir, are promising but appear to have some serious side effects. In these circumstances, it is increasingly important to focus attention on developing an HIV vaccine and a cure. HIV host genetics has made significant advances in HIV cure research. In our recent work, we demonstrated the importance of an epigenetic mechanism, DNA methylation, plays on HIV disease. We previously showed specific human genes have methylation patterns that control HIV disease progression. In this study, we explored a rare cohort of females (N=120) followed up from pre-HIV infection to post-HIV infection and followed throughout various fiebig stages and are HIV treatment naïve. We then have these individuals followed post-ART treatment. This cohort from about 15 years of followup through various HIV stages of disease was used to explore the role genome-wide DNA methylation has on HIV infection and disease progression, with a total of 400 samples. Using the Infinium MethylationEPIC v2.0 array, which covers more than 850,000 methylation sites. Our results show specific unique genes associated with either HIV infection or HIV disease. We report unique methylation patterns within the HIV-negative timepoint compared to the HIV-positive timepoint from the same patients. We also developed a prediction algorithm that uses methylation levels as a determinant of viral load. Using our dataset, we can predict viral load (R^2=0.92; p=0.00001). These data provide several novel target sites for potential innovative therapeutic strategies that could be explored in the future.

Session Title: Epigenetics Poster Session III

PB2335 Using Long-Read Sequencing to Identify Methylation Differences Related to Alzheimer's Disease

Authors:

R. Genner[1,2], M. Meredith[3], K. Billingsley[1], P. Alvarez[1], L. Malik[1], B. Paten[3], W. Timp[2], M. Jain[4], C. Blauwendraat[1]; [1]Ctr. for Alzheimer's and Related Dementias, NIH, Bethesda, MD, [2]Johns Hopkins Univ., Baltimore, MD, [3]UC Santa Cruz Univ. Genomics Inst., Santa Cruz, CA, [4]UC Santa Cruz, Santa Cruz, CA

DNA methylation is an epigenetic mechanism that involves adding a methyl group to cytosine residues in DNA and is generally associated with transcriptional repression. Methylation differences affecting gene transcription levels have been shown to play a role in the development and progression of neurodegenerative diseases including Alzheimer's disease (AD). Methylation levels have traditionally been measured using bisulfite conversion followed by methylation arrays and short read sequencing techniques, however these methods detect only short-range methylation patterns, are not optimized for identifying methylation in traditionally challenging genomic regions, and do not allow for phased haplotype specific methylation detection.

New "long-read" sequencing approaches have been developed by PacBio and Nanopore that can directly sequence reads ranging from 10kb to 1Mb+ in length. This results in increased sequencing accuracy and resolution which is particularly helpful for analyzing complex regions of the genome. Long-read sequencing can also directly detect base modifications such as 5MC methylation, allowing for additional genetic information to be extracted such as haplotype phasing for the detection of allele-specific methylation differences. NIH's Center for Alzheimer's Disease and Related Dementias (CARD) has developed protocols designed to streamline and automate the tissue processing and long-read sequencing of thousands brain samples from individuals with and without AD. Generation of this sequencing data provides a unique opportunity to analyze genome-wide, population scale methylation patterns and assess the methylation levels of poorly resolved genomic regions in human brain tissue.

We are currently using computational methods and algorithms to identify allele-specific, cell type-specific and genome-wide methylation differences in long-read sequencing data from the first ~250 brain samples. We are also conducting benchmarking experiments to compare methylation detection in samples run on Nanopore R9 flow cell series with the updated R10 series, which reports improved read accuracy. We found that methylation levels are comparable across CpG islands, promoter regions, and genome-wide methylation levels for three cell lines and one brain sample.This information will provide new insights into the epigenetic mechanisms of the brain and the underlying etiology of AD that could eventually lead to the development of novel and improved targets for treatment strategies.

Session Title: Epigenetics Poster Session I

PB2336 Utility of DNA methylation episignatures in neurodevelopmental disorders: from variant classification to new diagnoses, interactions and novel episignature discovery.

Authors:

S. Trajkova[1], J. Kerkhof[2], L. Pavinato[1], M. Rossi Sebastiano[1], A. Mussa[3], S. Banka[4], C. Campbell[4], C. Giovenino[1], V. Pullano[1], S. Carestiato[1], S. Cardaropoli[3], E. Ferrero[1], E. Shukarova-Angelovska[5], A. Petlichkovski[6], T. Pippucci[7], P. Dimartino[8], S. De Rubeis[9], J. Buxbaum[9], L. van der Laan[10], A. Silva[11], K. Rooney[12], H. McConkey[13], G. Ferrero[1], B. Sadikovic[14], A. Brusco[15]; [1]Univ. of Turin, Turin, Italy, [2]London Hlth.Sci. Ctr., London, ON, Canada, [3]Dipartimento di Scienze della Sanità Pubblica e Pediatriche, Torino, Italy, [4]Manchester Univ. NHS FT, Manchester, United Kingdom, [5]Dept. of Endocronology and Genetics, Skopje, Macedonia, The Former Yugoslav Republic of, [6]Inst. for Immunobiology and Human Genetics, Skopje, Macedonia, The Former Yugoslav Republic of, [7]Sant'Orsola Malpighi Univ. Hosp., Bologna, Italy, [8]Dept. of Med. and Surgical Sci., Univ. of Bologna,, Bologna, Italy, [9]Icahn Sch. of Med. at Mount Sinai, New York, NY, [10]Amsterdam UMC, Amsterdam, Netherlands, [11]London Hlth.Sci. Ctr., London, ON, Canada, [12]London Hlth.Sci. Ctr., London, ON, Canada, [13]Western Univ., London, ON, Canada, [14]LHSC, London, ON, Canada, [15]Univ. of Torino, Turin, Italy

Unique DNA methylation (DNAm) patterns, known as “episignatures”, are associated with specific genes/pathways in neurodevelopmental disorders (NDD). We analyzed episignature profiles (EpiSignTM classifier v.4) in 164 NDD cases from four cohorts: (i) validation (n=71), cases with pathogenic/likely-pathogenic variants in genes with known episignatures; (ii) uncertain (n=13), cases with VoUS; iii) unsolved (n=24), cases with uninformative CMA/ES; (iv) discovery (n=56), aimed at identifying novel disease-associated episignatures. We found the expected episignature in 65/71 cases in the validation cohort. Exceptions included a *SMARCA2*:M856V variant in the Nicolaides-Baraitser syndrome-associated domain, with a methylation profile for the allelic disease Blepharophimosis-impaired intellectual development syndrome (BIS). The patient was later re-diagnosed with BIS. Interestingly the extreme case of the unique *SMARCA4*:M886V methylation profile overlapped with the *SMARCA2*:M856V. In silico tools showed that M856 and M886 are structural homologs, further supporting that their mutation to Valine exert analog effects. Another interesting finding was the DNAm pattern of R37H-Kleefstra variant and G11R in the N-terminal domain of *SMARCB1* protein that matched the C-terminal ARID1A/B:c.6200 sub-signature. Clinical presentation of the latest corresponded to the described R37H-*SMARCB1*.All these variants reside very near in the BAFopahty complex. In a patient with Rubinstein Taybi syndrome 2 (EP300), we found a GNAS methylation pattern suggestive of Pseudohypoparathyroidism 1B, as secondary finding. In the uncertain cohort, 8/13 cases did not match the methylation profile of the reported VoUS, suggesting that these were benign variants. Among these, a VoUS in *ARID1B* (Coffin-Siris 1) showed a Cornelia de Lange (CdL) episignature, the initial clinical diagnosis. In the unsolved cohort, a second CdL-case with uninfrometive-ES matched the CdL episignature. Another showed an ATRX-methylation profile, subsequently revealing a deletion of *ATRX* ex3-4 missed by ES. TRIP12-episignature was useful to underpin the genetic diagnosis in family with two similarly affected brothers having two different *de novo* variants (*TRIP12*:L1044Ffs*3, *FBN*:A1728V). In the discovery cohort, we defined three robust novel episignatures for NDD genes: *TLK2*, *CAPRIN1*, *ZMYM3*.Episignature analyses have improved variant classification, supported new disease associated gene as functional analyses, and unraveled novel gene interactions in the BAFopahty complex.

Session Title: Epigenetics Poster Session II

PB2337 Whole-genome bisulfite sequencing of cell-free fetal DNA from low volume plasma samples facilitates investigations of methylomic signatures associated with autism risk factors

Authors:

L. Williams[1,2,3,4], O. Sharifi[1,3,4], R. J. Schmidt[1,2,3], J. M. Lasalle[1,2,3,4]; [1]Univ. of California Davis Sch. of Med., Davis, CA, [2]Perinatal Origins of Disparities Ctr., Davis, CA, [3]MIND Inst., Sacramento, CA, [4]Univ. of California Davis Genome Ctr., Davis, CA

Background: The etiology of autism spectrum disorder (ASD) is complex, involving an interplay of genetic and environmental factors. Many proposed environmental risk factors for ASD have been implicated to act through epigenetic mechanisms, such as DNA methylation, during the perinatal development window. Cell-free fetal DNA (cffDNA), which can be found in maternal plasma, has shown promise as an accessible source of fetal DNA to investigate the perinatal period. However, the investigation of patterns of perinatal epigenetic dysregulation associated with environmental exposures and ASD has been historically challenging. Previous efforts to interrogate the perinatal fetal epigenome required high volumes of maternal plasma, and relied on methylation arrays of narrow genomic breadth.
Objective: A comprehensive understanding of ASD etiology requires the tools for scalable, genome-wide methylation profiling of fetal DNA from the perinatal period. We aim to perform low-pass whole genome bisulfite sequencing (WGBS) and comethylation network analysis on 1mL of maternal plasma as a proof of principle for the application of these methods in studies of ASD.
Methods: Twenty 1mL blood plasma aliquots were collected from pregnant women during 3rd trimester or at delivery. Cell-free DNA was isolated from these aliquots and subjected to bisulfite conversion. Libraries were generated from bisulfite converted cffDNA and pooled for sequencing on the NovaSeq S4. The CpG_Me pipeline was employed to perform QC and generate cytosine reports, which were used as input for comethylation network analysis via the Comethyl package.
Results: An average of 4.30ng of cfDNA was isolated from each blood plasma sample, with 14 of these samples used for WGBS. An average of 411.4M total reads for each sample was observed, with a mean coverage of 85.1% and average depth of 10.9X. These data were sufficient to generate comethylation networks from ~260k CpG-containing regions. Multiple comethylation module were found to have significant associations with the gestational timepoint at which plasma was collected and such modules were annotated.
Conclusion: This pilot study validates the feasibility of performing WGBS on cffDNA collected from low volumes of maternal plasma, and empowers researchers to identify associations between differences in fetal DNA methylation in the perinatal period and neurodevelopmental outcomes. Our group is in the process of applying this technique to cffDNA collected from a high-risk ASD cohort with known perinatal exposure data, which will enable us to identify patterns of epigenetic dysregulation associated with both perinatal exposures and ASD outcomes.

Session Title: Epigenetics Poster Session III

PB2338 Whole-genome bisulfite sequencing of motor neurons reveals cell specific enhancers and enables the identification of motor neuron derived cell-free DNA

Authors:

C. Harvey[1], E. Hanon[2], A. Weimer[3], C. Caggiano[4], T. Moll[1], C. Dos Santos Souza[1], S. Zhang[5], A. Barcons[1], K. Ning[1], K. Kenna[6], I. Timpanaro[6], C. van Dijk[6], N. Zaitlen[4], L. Ferraiuolo[1], J. Mill[2], M. Snyder[3], P. Shaw[1], J. Cooper-Knock[7]; [1]Univ. of Sheffield, Sheffield, United Kingdom, [2]Univ. of Exeter, Exeter, United Kingdom, [3]Stanford Dept. of Genetics, Palo Alto, CA, [4]UCLA, Los Angeles, CA, [5]Stanford Univ., Palo Alto, CA, [6]Univ. Med. Ctr. Utrecht, Utrecht, Netherlands, [7]Univ. Of Sheffield, Sheffield, United Kingdom

Background
The cardinal feature of amyotrophic lateral sclerosis (ALS) is the loss of upper and lower motor neurons, which causes progressive weakness and death. Vital to understanding the mechanism of ALS and finding successful treatments is understanding the selective vulnerability of motor neurons. A key regulator of cell specificity is cell specific epigenetic modifications which enable expression of cell specific genes.
In order to understand the unique epigenetic regulation in motor neurons we performed whole genome bisulfite sequencing (WGBS) in induced pluripotent stem cell (iPSC) derived motor neurons to capture genome wide methylation state. We then analysed these methylation profiles alongside a DNA methylation atlas of cell types from Loyfer et al. (2023) to identify motor neuron specific hypo and hypermethylated regions. We then analysed these regions to identify motor neuron specific regulatory regions such as enhancers.
A key use of a reference methylome is that it can be used with a deconvolution algorithm to infer the cellular composition of a mixture from its methylation state. Cell free DNA (cfDNA) in blood plasma is composed of DNA from recently lysed cells, and quantifying the motor neuron derived DNA within that has previously been identified as a promising biomarker in ALS (Robichaud et al. 2021). We used our data to test the feasibility of motor neuron-derived cfDNA as a biomarkerObjectives
To identify motor neuron specific hyper and hypomethylated regions, then to use these regions to identify motor neuron specific enhancers and regulatory regions.
Test the feasibility of motor neuron derived cfDNA as a biomarker of ALS.MethodsWe integrated whole genome bisulfite sequencing of iPSC-derived motor neurons derived from three healthy donors with the methylation atlas from Loyfer et al. (2023) to identify motor neuron specific regions.
We optimised a deconvolution algorithm using synthetic mixes of cfDNA from healthy donors and motor neuron derived DNA and describe its accuracy.
Finally, we will use our optimised deconvolution algorithm on WGBS sequenced cfDNA from 12 ALS patients and 12 controls, previously published in (Caggiano et al. 2021).ResultsWe identify 5281 regions uniquely hypomethylated and 5037 uniquely hypermethylated regions specific to motor neurons.
With our optimised deconvolution algorithm we demonstrate it is feasible to detect motor neuron derived DNA present at 1% of cfDNA

Session Title: Epigenetics Poster Session I

PB2339 Widespread age-associated changes in the chromatin architecture of skeletal muscle cell types.

Authors:

K. Moo[1], R. Albanus[2], A. Varshney[3], P. Orchard[1], N. Manickam[4], M. Laakso[5], J. Tuomilehto[6], T. Lakka[7], K. Mohlke[8], M. Boehnke[1], L. Scott[9], H. Koistinen[10], F. Collins[11], S. Parker[1]; [1]Univ. of Michigan, Ann Arbor, MI, [2]Washington Univ., Saint Louis, MO, [3]Univ. of Michigan, Canonsburg, PA, [4]10x Genomics, Pleasanton, CA, [5]Inst. of Clinical Med., Kuopio, Finland, [6]Dept. of Med., Kuopio, Finland, [7]Inst. of Biomedicine, Kuopio, Finland, [8]Univ North Carolina, Chapel Hill, NC, [9]Univ Michigan, Ann Arbor, MI, [10]Dept. of Publ. Hlth.and Welfare, Helsinki, Finland, [11]NIH, Bethesda, MD

Aging progressively alters the cellular epigenomic landscape and is associated with increased risk for complex diseases such as cancer, Alzheimer's disease, and type 2 diabetes. Genome-wide association studies have identified hundreds of genetic variants associated with these diseases, the majority of which occur in noncoding regions of the genome. Together, these findings suggest that aging-related epigenomic changes are one component of disease risk. Here we use skeletal muscle biopsies from 284 Finnish adults aged 20 to 79 years (mean 60 years) and single nucleus ATAC-seq to identify patterns in chromatin architecture that are associated with age. To find age-related chromatin patterns in an unbiased way without assuming a linear relationship we use Auto-Regressive Integrated Moving Average (ARIMA) models. Of the 13 skeletal muscle cell types identified via clustering, we selected the five most abundant (Type 1, 2a, and 2x muscle fiber, endothelial, and mesenchymal stem cells) for analysis. We then identified between 149,751 and 289,430 peaks within each of the five selected cell types and, using a generalized linear model, adjusted the sample-specific peak accessibility for body mass index and technical covariates such as batch. Then, we used the ARIMA modeling framework and found that between 6.2% and 15.5% of the tested peaks were age-associated. We used unsupervised clustering to classify age-associated peaks according to their pattern of association (increasing accessibility, decreasing accessibility, or dynamic with age). Breakpoint analysis of these three categories showed that in all cell types the chromatin landscape is changing at its greatest magnitude around 60 years of age. To determine if age-associated peaks are preferentially associated with particular biological pathways, we performed gene ontology (GO) enrichment analysis leveraging the GO terms assigned to nearby genes. These analyses showed both expected results, such as skeletal system development in skeletal muscle fiber, and several other robust signals in increasing, decreasing, and dynamic age-associated peaks. Analyses are ongoing to determine whether variants associated with cell type-relevant complex traits are enriched in age-associated peaks. We are also working to explore the associations of age-associated peaks with chromatin state (transcribed, enhancer, poised, etc.) in order to better understand the context of these peaks with non-linear age dependent dynamics. Together, these results highlight age-associated epigenomic differences in five skeletal muscle cell types, giving us insight into cellular development and aging across the human lifespan.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Session I

PB2340 *"A Digital Binder": How digital tools enable personalized and partnered care in genetics*

Authors:

D. Assamad[1], S. Saeedi[2], S. Grewal[2], V. Jha[2], S. Luca[1], M. Clausen[2], A. Shaw[2], W. Lee[1], Y. Bombard[2], R. Z. Hayeems[1], on behalf of the Genetics Navigator Team:, M. Aronson[3], F. Bernier[4], M. Brudno[5], J. C. Carroll[6], L. Chad[7], R. Cohn[7], G. Costain[7], I. Dhalla[8], H. Faghfoury[9], J. Friedman[10], S. Hewson[11], T. Jamieson[12], R. Jobling[7], R. Kodida[2], A-M. Laberge[13], J. Lerner-Ellis[14], E. Liston[7], M. Mamdani[15], C. Marshall[16], M. Osmond[17], Q. Pham[18], E. Reble[2], F. Rudzicz[19], E. Seto[18], S. Shastri-Estrada[20], C. Shuman[11], J. Silver[11], M. Smith[21], M. Smith[22], K. Thorpe[23], W. J. Ungar[1]; [1]Program in Child Hlth.Evaluative Sci., The Hosp. for Sick Children, Toronto, ON, Canada, [2]Genomics Hlth.Services Res. Program, Li Ka Shing Knowledge Inst., St. Michael's Hosp., Unity Hlth.Toronto, Toronto, ON, Canada, [3]Zane Cohen Ctr. for Digestive Diseases, Sinai Hlth.System, Toronto, ON, Canada, [4]Dept. of Med. Genetics, Alberta Children's Hosp., Calgary, AB, Canada, [5]HPC4Hlth.Consortium,, Toronto, ON, Canada, [6]Dept. of Family & Community Med., Sinai Hlth., Univ. of Toronto, Toronto, ON, Canada, [7]Div. of Clinical and Metabolic Genetics, The Hosp. for Sick Children, Toronto, ON, Canada, [8]Care Experience Inst., Unity Hlth.Toronto, Toronto, ON, Canada, [9]Fred A. Litwin Family Ctr. in Genetic Med., Univ. Hlth.Network and Sinai Hlth.System, Toronto, ON, Canada, [10]Dept. of Med. Genetics, Univ. of British Columbia, Vancouver, BC, Canada, [11]Dept. of Molecular Genetics, Univ. of Toronto, Toronto, ON, Canada, [12]Dept. of Med., Univ. of Toronto, Toronto, ON, Canada, [13]Div. of Med. Genetics, Dept. of Pediatrics, Ctr. Hosp.ier Univ.ire Sainte-Justine, Montreal, QC, Canada, [14]Lunenfeld-Tanenbaum Res. Inst., Sinai Hlth.System, Toronto, ON, Canada, [15]Dept. of Data Sci. and Advanced Analytics, Li Ka Shing Knowledge Inst., St. Michael's Hosp., Unity Hlth.Toronto, Toronto, ON, Canada, [16]Genome Diagnostics, Dept. of Pediatric Lab. Med., The Hosp. for Sick Children, Toronto, ON, Canada, [17]Children's Hosp. of Eastern Ontario Res. Inst., Univ. of Ottawa, Ottawa, ON, Canada, [18]Inst. of Hlth.Policy, Management, and Evaluation, Univ. of Toronto, Toronto, ON, Canada, [19]Vector Inst. for Artificial Intelligence, Toronto, ON, Canada, [20]Dept. of Occupational Sci. and Occupational Therapy, Temerty Faculty of Med., Univ. of Toronto, Toronto, ON, Canada, [21]Northwestern Univ., Chicago, IL, [22]Patient Partner, Canadian Organization for Rare Disorders, Toronto, ON, Canada, [23]Dalla Lana Sch. of Publ. Hlth., Univ. of Toronto, Toronto, ON, Canada

Introduction: Digital tools are emerging to support the delivery of genetic services, including pre-test counseling, consent, collection of family and medical history, education, results disclosure and post-test counseling. Research has shown that such tools generally improve patient-reported outcomes related to decision-making, knowledge, and psychosocial well-being. However, patients' perspectives and preferences for how these tools are used to facilitate patient-centered care are not well understood. To address this gap, we aimed to better understand patients' opinions and preferences towards the use of digital tools in genetic service delivery.

Methods: Semi-structured interviews were conducted with individuals who received genetic testing for themselves or for their child. Participants were recruited from the Canadian Organization for Rare Disorders (CORD) or were previously enrolled in a Toronto-based cancer genetics research study. Participants were asked about their experiences with and preferences for the use of digital tools in the delivery of genetic services and completed a demographics survey. Thematic analysis was used to interpret the qualitative data.

Results: A total of 30 participants were interviewed. Of these, 17 (57%) received genetic testing for themselves,13 (43%) for their child, and 20 (67%) identified as female. Overall, participants reported positive experiences with digital tools in medical settings and welcomed the shift towards the use of these tools. They desired more agency over their healthcare journey and control of their health information, including their genetic test results. Participants imagined that through the use of a digital platform that they referred to as a 'digital binder', they could access, review and share their health-related information including personal and family histories and test results. Participants anticipated that increased access would enable patients to feel like equal and empowered partners in their care. Balancing a desire for digitally-enabled agency and sustained personal connections with providers, participants anticipated that digital tools could be tailored to each patient's circumstances and needs, enabling a personalized and partnered model of care. They recommended that digital tools be used as a supplement to current models of care rather than as a replacement.

Conclusion: Digital tools, like a 'digital binder' can supplement traditional genetic service delivery and have the potential to enhance patient-centered care, if used appropriately. These findings can inform the implementation and evaluation of digital health tools in genetics service delivery.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Session II

PB2341 “I worry I don’t have control”: The psychosocial impacts of living with a hereditary cancer syndrome

Authors:

J. Sam[1], B. Sparkes[2], **M. Clausen**[1], E. Reble[1], C. Butkowsky[1], S. Rajeziesfahani[2], R. Gopalakrishnan[1], V. Aguda[1], M. Aronson[3], D. Bishop[4], L. Dawson[5], A. Eisen[6], T. Graham[6], J. Green[5], C. Mighton[1], J. Pauling[7], C. Pavao[8], P. Pechlivanoglou[9], C. Remocker[8], S. Savas[10], S. Sun[11], T. Tiano[7], A. Tilley[12], K. Thorpe[9], K. Schrader[13], H. Etchegary[2], Y. Bombard[9]; [1]St. Michael's Hosp., Toronto, ON, Canada, [2]Mem. Univ. of Newfoundland, St. John’s, NL, Canada, [3]Zane Cohen Ctr., Sinai Hlth.System, Toronto, ON, Canada, [4]Patient Partner, St. John’s, NL, Canada, [5]Mem. Univ. of Newfoundland, St. John's, NL, Canada, [6]Sunnybrook Hlth.Sci. Ctr., Toronto, ON, Canada, [7]Patient Partner, Toronto, ON, Canada, [8]Patient Partner, Vancouver, ON, Canada, [9]Univ. of Toronto, Toronto, ON, Canada, [10]Mem. Univ. of Newfoundland, St. John's, ON, Canada, [11]BC Cancer, Vancouver, BC, Canada, [12]Patient Partner, St.John's, ON, Canada, [13]BC Cancer / Univ British Columbia, Vancouver, BC, Canada

Introduction: Hereditary cancer syndromes (HCS) are one of the most common forms of inherited diseases, with Hereditary Breast and Ovarian Cancer Syndrome (HBOC) and Lynch Syndrome (LS) being the most prevalent. HCS patients are genetically more susceptible to developing cancer in their lifetime and often require consistent, lifelong screening and monitoring. Various facets of patients’ lives may be indirectly affected by their diagnosis and care. However, limited evidence describes the range of psychosocial impacts of HCS following a positive genetic diagnosis. The goal of this study is to describe the psychosocial & lifestyle impacts of HCS. Methods: Patients with a confirmed molecular diagnosis of HBOC or LS participated in semi-structured qualitative interviews. Patients were recruited from hereditary cancer clinics across 3 provinces with varying HCS health systems in Canada. Interpretive description was used to analyze the data. Results: Qualitative interviews were conducted with 73 patients (51 females, 21 males, 1 gender-diverse; age ranges 25-80yrs) diagnosed with HBOC (n= 39) or LS (n= 34). Cancer worry, the fear of oneself or one’s family members developing cancer, was a common concern for many patients rooted in a loss of control: “I worry I don’t have control”. As a result of their increased risk for cancer, patients described heightened symptom monitoring and concerns that unlikely symptoms (e.g. cold symptoms) were cancer-related. This cancer worry expanded to family members, and many parents expressed carrier guilt over possibly passing on their HCS to future generations. To cope with their cancer worry, patients described changes to their outlook to focus on the positive aspects of their diagnosis. To take control of their cancer, patients noted changes to their lifestyle, such as improving diet, exercise, and social activities. Additionally, some patients sought risk-reducing prophylactic surgeries to reduce their cancer risk, though they noted subsequent psychosocial and lifestyle impacts, such as challenges with body image and delaying family planning. Other strategies to gain control included joining patient groups, seeking professional therapy, or discussing their HCS journey with friends and family with similar lived experiences. Conclusion: A HCS diagnosis has broad psychosocial and lifestyle implications. Cancer worry is commonly expressed by many patients. In turn, this fear of developing cancer triggers changes to patients' lifestyles to take control of their cancer risk. This work emphasizes the need to provide multidisciplinary support to patients and families living with HCS beyond the point of diagnosis.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Session III

PB2342 "The reason you have anxiety is because you have this genetic disease": sources of anxiety and depression among individuals with inborn errors of immunity.

Authors:

K. Lewis[1], B. Beers[2], H. Davidson[2], M. Setzer[3], R. Gore[1], M. Kamen[4], C. Jodarski[5], H. Orellana[2], N. Reynolds-Lallement[2], D. Snyder[6], J. Yan[2], M. Walkiewicz[4], M. Similuk[7]; [1]Med. Sci. & Computing, NIH, Bethesda, MD, [2]NIH, Bethesda, MD, [3]NIH, Uniformed Services Univ. of the Hlth.Sci., Bethesda, MD, [4]NIAID, Bethesda, MD, [5]NIH, Washington, DC, [6]Natl. Inst.s of Mental Hlth., Bethesda, MD, [7]NIH, Bethesda, MD

Background: Inborn errors of immunity (IEI) affect as many as 1/2,000 live births. Genomic testing is often completed to ensure unequivocal diagnosis, optimal treatment, and accurate risk assessment. Roughly half of individuals with an IEI may report mild levels of anxiety, depression, or both. We aimed to describe sources of patients' anxiety and depression to inform the way providers address these concerns. **Methods:** Eligible participants were adults diagnosed with IEI who underwent genomic sequencing and had scores on a screening assessment in a range consistent with mild depression and/or anxiety (T-score >52.5 on Patient Reported Outcome Measurement Information System (PROMIS) 29 Profile v2.1) and above average adaptation (average >3.7 on Psychological Adaptation Scale) on a prior survey. Potential participants were contacted three times. Interviews were conducted by two interviewers using a semi-structured guide with questions about sources and experiences of anxiety and depression. Interviews were recorded and transcribed. Transcripts were coded and analyzed using an inductive, semantic thematic approach. **Results:** Twenty individuals (57.1%) participated. They were 32.3 years old on average. The most frequent diagnoses were common variable immunodeficiency (n = 5, 25%) and GATA2 deficiency (n = 5, 25%). Roughly half were women (n = 13, 65%) and seven had undergone hematopoietic stem cell transplant (35.0%). The most common theme was a bidirectional relationship between mental health and physical health. Most participants emphasized that uncertainty about their symptoms and prognosis caused feelings of anxiety and depression. Many felt lonely, isolated, or experienced social disruption due to their IEI (i.e., missing school or work for care or social distancing) which led them to feel anxious or depressed. Some recommended that providers assess patients' mental health, particularly because of the interrelatedness of their IEI and their mental health. As one participant said, "...They would help insofar as associating the fact that the depression had to do more with my fear of my disease than anything else. For a long time, I just felt like I separated those two things in my head."**Conclusions:** Participants endorsed symptoms of anxiety and depression as important considerations. It may be helpful for clinicians to acknowledge that aspects of IEI (i.e., chronic symptoms, uncertain prognosis, rarity) can contribute to feelings of anxiety and depression. We suggest that providers routinely screen patients with IEI for anxiety and depression and refer for further care, as needed. This study was limited by sample size and participant self-selection.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Session I

PB2343 A playbook for the implementation of genetics services and research in a developing country: a 10 year-experience from DR Congo

Authors:

A. Lumaka Zola[1,2], P. LUKUSA[1], K. Devriendt[3]; [1]Univ. of Kinshasa, Kinshasa, Congo, Democratic Republic of the, [2]CHU LIEGE, Liege, Belgium, [3]Univ Hosp Leuven, Leuven, Belgium

Genomic medicine is an emerging discipline of medical science that requires accessing the genomic information of an individual and interpreting this information in relation to the present or future phenotype. Despite the enthusiasm and expectations generated by this discipline, the uptake in Low and Middle-Income Countries (LMIC) remains limited or very slow.
Since 2009, the Center for Human Genetics of Kinshasa (Democratic Republic of Congo, DRC) and the Center for Human Genetics of the KULeuven (Belgium) engaged in the implementation of genomic medicine in the DRC. The following steps were undertaken: (1) establishing both intra and extra-mural clinical service with pediatricians, (2) establishing a manual DNA extraction laboratory, (3) training clinicians and laboratory personnel both in the country and overseas, (4) creating visibility through publications in international peer-reviewed journals, (5) establishing international collaborations with research and philanthropic partners, while (6) applying for international grants.
As a result, a local network of hospitals and specialized schools is developed in the DRC; DNA extraction is supporting multiple research and clinical activities; nine Ph.D. and one lab technician have been trained in human genetics; more than 40 publications in genetics made in international peer-reviewed journals; a partnership is established with philanthropic programs such the Illumina iHope program (about 100 families have received clinical Whole Genome Sequencing), the UDNI and Wilhelm foundation (Champions Initiative), H3Africa (150 families tested with Whole Exome Sequencing through the NIH-funded DDD-Africa project); and international grants have supported the training and the purchase of additional laboratory equipment.
Altogether, this effort has created access to genomic services and provided the country with professionals that are able to handle genomic data and integrate genomic information into care in a LMIC. Without being perfect, our playbook will encourage those that are hesitant and lead to better uptake of genomic medicine.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Session II

PB2344 † A roadmap for precision prevention in primary care using a PRS enabled preventive intervention to promote lung cancer screening and smoking cessation

Authors:

L-S. Chen[1], T. Chen[2], G. Pham[3], L. Bierut[4], C. Amos[5], R. Hung[6], X. Lin[7], H. Zhang[8]; [1]Washington Univ. Sch. of Med., St. Louis, MO, [2]Harvard T.H. Chan Sch. of Publ. Hlth., Boston, MA, [3]Washington Univ. Sch. of Med., St. Louis, MO, [4]Washington Univ Sch Med, St Louis, MO, [5]Baylor Coll. of Med., Houston, TX, [6]Sinai Hlth., Univ. of Toronto, Toronto, ON, Canada, [7]Harvard T.H. Chan Sch Pub Hlth, Boston, MA, [8]Natl. Cancer Inst., Cambridge, MA

Background: Lung cancer (LC) poses a significant global health challenge. Early detection and targeted intervention are crucial to reduce its morbidity and mortality. Polygenic risk scores (PRS) offer a robust tool for stratifying disease risk among populations and identifying high-risk individuals. However, their use remains limited in prevention practices in primary care with diverse populations. We aim to develop a PRS-based strategy for a precision prevention intervention to reduce lung cancer mortality. **Methods**: We developed a LC PRS for clinical application, calibrated on 2,504 individuals from the 1000 Genomes Project Phase 3 (1000G) and validated on over 340,000 patients of diverse ancestry from the UK Biobank (UKBB) and Genetically Informed Smoking Cessation Trial (GISC). We identified 101 SNPs out of 128 predictive SNPs from Hung et al. (2020) and constructed PRS using multi-ancestry summary statistics from Byun et al. (2022). We performed a Principal Component Analysis in 1000G and adjusted the PRS for genetic ancestry by regressing out the top five genetic PCs. We estimated the global cutoff points for adjusted PRS in 1000G, further validated in UKBB and GISC, and calculated the corresponding odds ratios (ORs) of LC for risk stratification across a diverse patient population. **Results**: Our methodology harmonizes the PRS distribution among different ancestries for equitable risk assessment without categorizing patients into discrete ancestry groups. To motivate lung cancer screening in eligible patients, we categorized their PRS into “risk”, “high risk” and “very high risk” groups based on the 1000G PRS distribution. Relatively to the “at risk” group, the overall ORs of LC, were 1.30 and 1.68 for the other two risk groups in UKBB. Notably, the OR for the “very high risk” group was significantly larger among non-European ancestry patients, though the case count was potentially insufficient to fully support this distinction. **Conclusion**: Our research offers an innovative PRS-enabled precision prevention intervention framework in primary care, especially for high-risk patients with actionable recommendations such as lung cancer screening and smoking cessation. The implementation and effectiveness of the PRS-enabled behavior change intervention, *RiskProfile*, will be evaluated in 1,625 diverse primary care patients and 140 physicians via two clinical trials. We also explore the translational roadmap for PRS, discuss barriers, potential solutions, and provide an innovative perspective on genetic risk-based health interventions.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Session III

PB2345 A systematic review of international adult Neurofibromatosis 1 surveillance frameworks.

Authors:

J. Chiang[1], S. Rana[2], C. Low[2], J. Koh[3], J. Ngeow[1]; [1]Natl. Cancer Ctr. Singapore, Singapore, Singapore, [2]NUS Yong Loo Lin Sch. of Med., Singapore, Singapore, [3]KK Women's and Children's Hosp., Singapore, Singapore

Background: Neurofibromatosis 1 (NF1) is an autosomal dominant tumour-predisposition condition commonly diagnosed in childhood and fully penetrant by adulthood. Long-term surveillance is cost-prohibitive and may cause variations in the monitoring of NF1 patients in high and low-income countries. Instrumental to adult NF1 care is the implementation of a surveillance framework by a multidisciplinary clinic. We aim to systematically review international surveillance strategies to 1) investigate any association between a country's socioeconomic status and differences in strategies used, 2) extract any existing surveillance framework for the implementation of a multidisciplinary adult NF1 clinic.
Methods: We searched PubMed, Embase, Web of Science and Cochrane. Relevant clinical information on the surveillance of clinical manifestations in adult NF1 patients worldwide was reviewed, extracted, and synthesised.
Results: We identified 51 papers reporting on 7724 individuals. Multiple imaging modalities are actively employed in developed countries for surveying adult NF1 patients. Interestingly, we did not find any relevant papers from less-developed countries. Internationally, there is also a lack of data on a consolidated surveillance framework to set up a multidisciplinary adult NF1 clinic. To address this, we provide a surveillance framework set by the National Cancer Centre Singapore for implementation in a multidisciplinary adult NF1 clinic in Asian countries.
Conclusion: This systematic review suggests that there is robust data on surveillance techniques for adult NF1 in high-income countries, but not for lower-income countries. There is also a lack of data internationally on surveillance frameworks to set up multidisciplinary adult NF1 clinics. Our provided framework will facilitate more effective monitoring of underserved patients in Asian countries. In addition, more research is warranted into the monitoring and surveillance of adult NF1 in lower-income countries.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Session I

PB2346 † A systematic way to find children with undiagnosed genetic diseases.

Authors:

Z. Tang, L. Jeffries, S. Lakhani; Yale Univ., New Haven, CT

Background: Although individually uncommon, rare diseases collectively affect 3-5% of the population and account for approximately half of total pediatric inpatient spending. Despite advances in molecular genomics, many patients still have descriptive diagnoses based on clinical signs and symptoms (i.e. infantile spasms or heterotaxy syndrome) and lack clear molecular genetic explanations. These patients may further be lost in "diagnostic odysseys" in search of a precise diagnosis that can lead to better prognostics, management, and family planning. Several advanced genetic diagnostic programs like Yale's Pediatric Genomics Discovery Program (PGDP) are freely available, but recruitment is patient-initiated and thus vulnerable to various healthcare access barriers inherent to the diagnostic odyssey. Thus, we sought to develop a reproducible, electronic health record (EHR)-based screening method to identify children with potentially undiagnosed rare diseases.
Methods: We developed a screening tool based on approximately 400 ICD-10-CM diagnostic codes, then queried the Yale New Haven Health System (YNHHS) EHR for hospitalized children between 2017-2022 with at least one screening code. Then, we manually reviewed patient charts to exclude patients with likely non-genetic or already confirmed genetic diagnoses. We used Pearson chi-square for categorical data, a multinomial regression model for predictors of PGDP enrollment, and Kruskal-Wallis one-way analysis of variance with pairwise comparisons with Bonferroni correction for multiple comparisons.
Results: Our EHR screen identified 1,648 potential undiagnosed patients. Manual chart review of the first 169 revealed that 4% were already enrolled in PGDP and 42% were previously unknown patients with potentially undiagnosed genetic disorders. A higher proportion of our screened cohort were Black (24%) and Hispanic (31%) relative to the general population or those enrolled in PGDP (p=0.033).
Conclusions: Our EHR screen identified a large number of previously unknown undiagnosed children as appropriate candidates for advanced diagnostic genetic evaluations. The missed group had over-representation of Blacks and Hispanics, suggesting that these groups are more likely to miss receiving precise genetic diagnoses.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Session II

PB2347 Adherence to risk management guidelines in individuals with hereditary cancer conditions in Singapore

Authors:

S. Jonnagadla[1], A. Lee[2], J. Yuen[1], J. Ngeow[1,2]; [1]Natl. Cancer Ctr. Singapore, Singapore, Singapore, [2]Nanyang Technological Univ., Singapore, Singapore

Background: Cancer remains the leading cause of death in many countries including Singapore. Up to 10% of cancer is associated with an inherited predisposition. There are globally established cancer risk management guidelines for inherited cancer conditions that include earlier, more frequent screening to detect cancer early or risk-reducing surgeries to prevent cancer. A preventive model of healthcare can reduce cancer morbidity, mortality and improves an individual's health-related quality of life. Assessing adherence to risk management is essential for service improvement and economic benefits to the patients and healthcare systems. This study aims to report adherence to risk management in individuals identified with inherited cancer conditions at the Cancer Genetics Service at the National Cancer Centre Singapore.
Methods and results: Medical charts of individuals with inherited cancer conditions identified between January 2014 to June 2022 were reviewed. Information on cancer screening, preventative surgeries and cancer detected post-genetic testing was collected for 489 eligible patients. The overall adherence to cancer screening and preventative surgeries was 88.9%, comprising of 308 (62.9%) fully adherent and 126 (25.9%) partially adherent patients. Adherence was higher in females (91.8%), individuals without family history of cancer (94.2%), individuals older than 60 years (96.3%), and individuals previously diagnosed with cancer (90%). Twenty-seven individuals (5.5%) were diagnosed with tumours/cancer through screening recommendations post-genetic testing. There were 226 individuals (46.2%) with hereditary breast and ovarian cancer syndrome. Of which, 38.2% females with this condition opted for bilateral mastectomy, 80.4% were managed with annual breast screening and 138 females (85.2%) underwent bilateral salpingoophorectomy. There were 74 (15.1%) Lynch syndrome carriers who were similarly adherent to colonoscopy and gastroscopy (77.5% and 75.9% respectively). Fifteen females (30%) with Lynch syndrome underwent prophylactic surgery to reduce their risk of endometrial and/or ovarian cancer.
Conclusion: The findings from this study indicate high adherence rates to recommended risk management among individuals with inherited cancer conditions in Singapore. This demonstrates value in identifying these high-risk individuals through genetic testing as carriers can avoid cancer morbidity and mortality. Genetic testing and counselling with emphasis on the benefits and utility of risk management strategies should be adopted to increase adherence in males and less adherent groups.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Session III

PB2348 An approach for standardizing and optimizing the process for genomic test order creation, sample intake, report generation, and result delivery for clinical whole genome sequencing including integration with tertiary analysis platforms.

Authors:

S. Lagu, S. Gabriel, N. Lennon, T. Desmet, J. Medrim, R. Onofrio, S. Matthews, K. Lewis, J. Tamulynas, D. Toledo, H. L. Rehm; Broad Inst. of MIT and Harvard, Cambridge, MA

The Broad Clinical Labs (BCL) is a CLIA-licensed and CAP-accredited clinical laboratory that to-date has provided over 200,000 clinical whole genomes (cWGS) to partner laboratories and clinical research programs. As genomics services become highly utilized and BCL's customer base increases, there is a need to be able to scale operations and to provide an improved user experience to our users across the clinical research and clinical diagnostic space. The solution also needs to be compliant with stringent patient privacy and information security regulations that ensure safety of patient data.

The vision behind our novel Sample Intake, Delivery and Reporting System (SIDR) is to provide an optimized, secure and compliant, self-service solution for users. This solution will provide a standardized process for order creation, sample intake and delivery of end results. The users will also be able to track real time sample status as it is being processed and data is analyzed thus minimizing repetitive manual touchpoints. The SIDR user-interface (UI) will provide both internal and external customers the ability to place their orders directly or organizations can use the API backend that will provide an automated ordering process for bulk orders. SIDR is a key component of the end to end data delivery system and will provide integrations with external platforms such as the terra.bio data repository and the Fabric genomics platform for clinical interpretation and reporting.

SIDR v1 was released earlier this year, allowing users to place orders for a variety of clinical whole genome sequencing based tests. These tests, broadly defined by their intended use include: a technical clinical genome (data and variants created under the clinical quality system) for users who are doing clinical research or have their own tertiary analysis capabilities; a screening genome that incorporates board-certified geneticist interpretation and report sign out of variants in defined gene lists such as actionable findings recommended by ACMG, and a diagnostic genome that ingests patient (and parental) phenotypic information to provide an indication-driven interpretation. A proprietary, genome-backed newborn screening panel has also been launched through the platform. We will present here the technical architecture of SIDR and the business processes that have been created to deliver clinical data files and reports to authorized users. This framework, and the lessons learned in building it, is broadly applicable to any group offering genomic services to the community.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Session I

PB2349 APHMG Undergraduate Medical Education Core Competencies in Genetics and Genomics: A framework and implementation example.

Authors:

L. Massingham[1], S. Nunez[2], J. A. Bernstein[3], D. P. Gardner[4], A. Parikh[5], E. T. Strovel[6], M. Ashfaq[7], R. J. LeClair[8], F. Quintero-Rivera[9]; [1]Warren Alpert Med. Sch. of Brown Univ., Providence, RI, [2]Yale Univ. Sch. of Med., New Haven, CT, [3]Stanford Univ Sch Med, Stanford, CA, [4]Marian Univ. Coll. of Osteopathic Med., Indianapolis, IN, [5]Univ. Hosp. Cleveland Med. Ctr., Cleveland, OH, [6]Univ Maryland, Baltimore, MD, [7]UTHlth.Houston McGovern Med. Sch., Houston, TX, [8]Virginia Tech Carilion Sch. of Med., Roanoke, VA, [9]Univ. of California Irvine (UCI), Irvine, CA

Discoveries in the field of genetics and genomics are expanding at a remarkable pace. This gained scientific knowledge is being translated into clinical practice with genomic information now being routinely used for preventative, diagnostic, and therapeutic decisions. As adoption of genomic medicine continues to evolve, health professionals will be required to stay abreast of genetic discoveries and technologies. Implementation of these advances within their scope of practice will be indicated. Therefore, medical school genetics curriculum must be periodically updated to ensure future physicians have the knowledge and tools necessary to provide optimal care for their patients. In 2022, the Association of Professors of Human and Medical Genetics (APHMG) updated the medical school genetics core competencies that were initially developed in 2013 (PMID: 36040446). The competencies were reviewed and updated through a structured approach incorporating a modified Delphi method. The updated APHMG core competencies will be presented to bring further awareness of this tool to those involved in medical genetics education (including teaching, leadership, and curriculum development, implementation, and assessment). The competencies are concise, specific, and assessable. Additionally, there are new updates that incorporate recent advances in clinical practice and promote equity and inclusion in clinical care. Competency based education provides the framework to shape the educational components into more clear and organized goals. One mechanism chosen by APHMG is to utilize these competencies as a framework to compile a Question Bank of board-style questions that are mapped to each of the 2022 competencies. This question bank is available on the APHMG Genetics Education Resource Exchange (www.aphmg.org/) for members to utilize in their courses in an easily accessible and searchable format. Use of these competencies across the undergraduate medical curricula will foster the knowledge, skills, and behaviors required in medical practice across a range of specialties.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Session II

PB2350 Assessing the Demographics of Clinical Genome Resource Members: an initial step toward enhancing diversity in an international consortium.

Authors:

E. Riggs[1], D. I. Ritter[2,3], B. VanHoute[1], J. Bulkley[4], M. DiStefano[5], M. Elnagheeb[6], I. McGregor[7], J. Berg[6], T. Klein[8], J. Morales[9]; [1]Geisinger, Danville, PA, [2]Baylor Coll. of Med., Houston, TX, [3]Texas Children's Hosp., Houston, TX, [4]Kaiser Permanente Ctr. for Hlth.Res., Portland, OR, [5]The Broad Inst., Boston, MA, [6]UNC Chapel Hill, Chapel Hill, NC, [7]Natl. Human Genome Res. Inst., Bethesda, MD, [8]Stanford Univ., Stanford, CA, [9]NHGRI, Bethesda, MD

The Clinical Genome Resource (ClinGen) is an NIH funded effort to create a central resource of clinically relevant genes and variants to improve genomic medicine. As part of the ClinGen Justice, Equity, Diversity and Inclusion (JEDI) initiative, we are committed to increasing diversity within our workforce. To identify areas for improvement and enhance JEDI principles across the consortium, we developed a survey to establish baseline characteristics of our consortium membership. Survey development was iterative and included evaluating multiple similar surveys, obtaining input from ClinGen working groups and external advisors, and guided interviews with both US-based and international participants. The survey was disseminated to 2098 ClinGen members in Nov - Dec 2022; 768 members responded (36.6% response rate). Overall, 64% (n=488) of ClinGen respondents are from the USA and 73% (n=544) self-identified as white. Sixty-five percent (n=497) of respondents identify as women, and 48% (n= 369) are between 25-44 years old. A broad range of professionals contribute to ClinGen, with the largest single group being clinical laboratory geneticists (21%, n= 145). The survey highlighted several gaps that we are actively addressing. For example, respondents self-identifying as Black/African-American (2%, n=17) or Hispanic/Latino (7%, n=49) were under-represented (using the 2020 US census data as a comparator), reflecting a systematic lack of representation in the ClinGen workforce. We are currently planning targeted outreach to organizations such as the Society for Advancement of Chicanos/Hispanics & Native Americans in Science with the goal of increasing participation. Though the majority of respondents identified as women, we identified a higher ratio of men in leadership roles; 45% of working group chairs were men though they made up 34% of total respondents. This constitutes ~30% fewer women chairs than expected (assuming that the proportion of women chairs should be equal to the proportion of women in ClinGen). We also found a lower ratio of Hispanic/Latino and Black/African American individuals in leadership roles (3% and 1%, respectively) compared to their overall representation in the consortium. We are issuing guidance to promote rotation of chairs so that under-represented participants can be promoted to leadership positions, as well as emphasizing education and awareness to ensure that expert panels are composed of diverse individuals at the outset. We aim for the leadership of ClinGen to reflect its participants , with the ultimate goal of enhancing diversity across the consortium in alignment with our JEDI Action Plan.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Session III

PB2351 Assessing the Diagnostic Odyssey in Pediatric Autism.

Authors:

K. Parsons[1], Z. Liu[1], M. K. Higashi[1], P. Kruszka[2], H. Raynes[3]; [1]Gene Dx, Stamford, CT, [2]GeneDx, Samford, CA, [3]Mount Sinai, New York, NY

Autism spectrum disorder (ASD) in the United States has an estimated prevalence of 21.5 cases per 1,000 children at 4 years of age (PMID: 36952289). The cumulative incidence of Autism Spectrum Disorder (ASD) up to 8 years suggests a prolonged diagnostic odyssey (PMID: 36952289). Identifying a genetic cause can inform patient care, making a compelling case for genetic testing. However, patients often endure a series of tests that may prolong the diagnostic odyssey. The purpose of this study was to describe the genetic testing workflow for ASD in a real-world setting and quantify cost of this diagnostic odyssey. A retrospective ASD cohort (2003-2021) was curated from a deidentified longitudinal dataset of 3.9 million patients from a major health system in New York State. Demographic and clinical variables were obtained from structured data. International Classification of Diseases (ICD-9/10) codes and Current Procedural Terminology (CPT) codes associated with a clinical ASD diagnosis and genetic tests were identified. Patients without procedures within 3 years prior to diagnosis and those older than 10 were excluded. Procedure charges were obtained from de-identified claims data for 2022.The cohort with an autism diagnosis (ICD-10 code) and a prior symptomatic code was 733. The most common CPT code was chromosomal microarray (141), followed by karyotype (120), FMR1 testing (112) and X-linked disability gene panel (94). Most patients had several testing procedures, with the majority of patients beginning their diagnostic journey with FMR1 testing (59%). The most frequent combination of tests was FRM1, Karyotyping and CMA. A sizable group of patients were tested by these methods as well as X-linked disability gene sequencing. The most expensive test was next generation sequencing (NGS) for X-linked disability genes (average charges $4,377 (95% CI $3,455-$5,408). Given that most ASD patients undergo numerous tests, total charges for several combinations were calculated. The most common pattern of tests in our cohort (FMR1, Karyotyping, and CMA) results in total charges of $3,410. Adding NGS increases this to $6,787. Our data suggests that a minority of children with ASD undergo genetic testing. Among those who do, most had tests with low diagnostic yields (0.5-2.0 % for FMR1 and 10-15% for CMA (PMID: 31182824)). Our results suggest that ASD patients and their caregivers face a significant diagnostic odyssey when stepwise testing is employed, representing a considerable cost burden. These results stress the need for clinical practice to align with current recommendations and to consider routine exome testing to reduce the ASD diagnostic odyssey.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Session I

PB2352 BAP1 tumor predisposition syndrome expert provider survey results.

Authors:

L. Byrne[1], C. Porteus[1], A. Kumar[1], H. Hanson[2], C. Cebulla[1], M. Abdel-Rahman[1]; [1]The Ohio State Univ., Columbus, OH, [2]SouthWest Thames Ctr. for Genomics, St. George's Univ. Hosp., London, United Kingdom

Objective: *BAP1* germline likely pathogenic/pathogenic variants (GPV) are associated with the hereditary tumor predisposition syndrome, BAP1-(TPDS) with an increased risk for uveal and cutaneous melanomas, mesothelioma, renal, and potentially other cancers. The phenotype and surveillance of BAP1-TPDS is still unclear. Thus, we created a survey to assess the current practices and challenges of medical providers of individuals with *BAP1* GPV. Methods: We invited healthcare providers from different specialties with expertise in the care of individuals with *BAP1* GPV to fill out a survey. Health care providers were selected based on either previous report(s) on *BAP1* GPV individuals and/or referring individuals with *BAP1* GPV to our study. Results: Out of 112 providers contacted, 51 (46%) completed the survey. Forty-three (84%) were physicians, 7 (14%) were Genetic counselors and one PhD geneticist. Twenty (39%) reported their specialty as Genetics and 11 (22%) Ophthalmology, and the majority work in an academic medical center and have 16 or more years' experience. The majority (55%) reported seeing ≥ 5 *BAP1* GPV individuals, with an additional 27% seeing more than one but <5. Regarding tumors that should be included in screening, there was high concordance on the four main cancers: uveal melanoma (94%), cutaneous melanoma (90%), renal cancer (88%) and mesothelioma (84%). However, there was less agreement for other cancers such as basal cell carcinoma (47%), bile duct cancer (25.5%), breast carcinoma (13%), ovarian carcinoma (13%), hepatocellular carcinoma (10%), and meningioma (6%). For cancer surveillance, variations in the age of onset, frequency and modality were reported. With respect to germline genetic testing of *BAP1*, 92% of providers would offer testing where there is known familial GPV, 90% for an individual with two BAP1 associated tumors, 86% for an individual with one BAP1 associated tumor and one in a first or second degree relative, 65% to an individual with early age of onset of a *BAP1* associated tumor (40 or younger), and 31% following identification of somatic *BAP1* mutation in a tumor. Conclusions: Although, most providers with experience in management of individuals with germline variants in *BAP1* agree on the association of *BAP1* GPV with four main cancers (uveal and cutaneous melanomas, renal cell carcinoma and mesothelioma), the contribution to predisposition to other cancers is still not clear. However, there is still variation in many of the aspects of preventive surveillance strategies for individuals with *BAP1* GPV including the age of onset to start surveillance for each cancer and the frequency and modality of such surveillance.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Session II

PB2353 † Barriers and Successes in Providing Genetic Cancer Risk Information to Families of Patients who are Deceased: Experiences of the Genetic Risk Analysis in Ovarian Cancer (GRACE) Study

Authors:

J. Hunter[1], Y. Prado[2], A. Reyes[2], J. Zepp[2], M. Maiyani[3], J. Sawyer[3], L. White[3], S. Vertrees[4], A. Rope[5], S. Weinmann[2], S. Lee[6], N. Henrikson[7], H. Feigelson[3]; [1]RTI Intl., Research Triangle Park, NC, [2]Kaiser Permanente Ctr. for Hlth.Res., Portland, OR, [3]Kaiser Permanente Colorado, Aurora, CO, [4]Kasier Permanente Ctr. for Hlth.Res., Portland, OR, [5]Genome Med., San Francisco, CA, [6]Columbia Univ., New York, NY, [7]Kaiser Permanente Washington Hlth.Res. Inst., Seattle, WA

The Genetic Risk Assessment in Ovarian Cancer (GRACE) study aims to identify families with an increased susceptibility for cancer by offering genetic testing to individuals with a prior diagnosis of ovarian cancer (OC). One goal of the study is to test the feasibility of providing genetic information to relatives of individuals who are deceased. Search of tumor registry data at two health care systems (Kaiser Permanente Northwest and Colorado) identified cases of OC diagnosed 1998-2020 who either did not have genetic testing or had genetic testing limited to *BRCA1/2* and could benefit from testing with a comprehensive panel of cancer risk genes. Of all patients with a diagnosis of OC identified, 943 were deceased (roughly half of the total 1874 cases), 72% of which had no genetic testing and an additional 10% had testing of BRCA1/2 only, representing a significant care gap for receipt of genetic risk information for these families. However, challenges arose in recruitment of families of individuals who are deceased. First, legal and regulatory guidance on family contact and health information disclosure impacted who the study team could contact and what information could be communicated about the patient's cancer history. Second, planned germline genetic testing of archived pathology specimens was not always successful. Despite these challenges, recruitment of families of individuals who are deceased has moved forward; when testing of the pathology specimen is not successful or feasible, first-degree relatives of the individual are directly offered germline testing. To date, relatives of 25 individuals who are deceased have been contacted, with at least 1 member from 12 of the families consenting to testing, a 48% recruitment success rate for families. This uptake of testing by families of patients who are deceased reflects a strong interest in genetic risk information. However, practical barriers must be addressed to provide genetic risk information on a broader scale. Overall, the GRACE study can inform implementation of future programs across health care systems, providing life-saving information to prevent and mitigate the burden of ovarian and other hereditary cancers.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Session III

PB2354 Building awareness, respect, and confidence about genetics through science education partnerships.

Authors:

R. O'Malley, N. Vincenten, R. Bowman, G. McNeil, H. Kim, T. Wu, M. Gelbart; Harvard Med. Sch., Boston, MA

The Building Awareness, Respect, and Confidence through Genetics (ARC) project (2016-2021) empowered high school teachers to facilitate dialogue about genetics advances and societal impacts within classrooms and in their communities. Components included a) professional development workshops, b) new curricula resources refined through input from teachers, c) building a community of practice to support teachers, and d) sharing and learning within schools and communities about recent advances in genetics, including their potential benefits and ethical, legal, and social implications. The five-year project supported 17 in-person and two virtual workshops. Teacher attendees (n=396) collectively reached an average of 5,100 students annually. Seven (7) lesson plans and 19 distance learning resources were created, the latter in response to educators' needs during the COVID-19 pandemic.

An evaluation based on pre- and post-workshop surveys as well as annual follow-ups found that a) teachers' interest in teaching genetics is very high; b) teachers' confidence in conversations about genetics topics increased after workshop participation; c) teachers' knowledge increased in all measured topic areas; and d) according to teachers, personal genetics topics interest and engage students. Anecdotal evidence indicates that lessons learned in class spread beyond the classroom setting through conversations with peers and family members. The most prevalent barrier to implementation by educators was lack of time.

The project resource library continues to grow with new or revised resources for educators, to ensure that up-to-date classroom-ready materials for teaching genetics are free and easily accessible online. Future studies would benefit from evaluating project impacts on the students themselves in addition to teacher perceptions.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Session II

PB2356 Carrier screening testing: Assessing advantages and limitations through a small consultant group.

Authors:

K. Rocha[1,2], M. Lazar[1,2], M. C. Varela[1,2], K. Bagatini[1,2], D. G. Franco[1,2], V. N. V. Takahashi[1,2], J. Y. T. Wang[1,2], M. R. Passos-Bueno[1,2]; [1]Human Genome and Stem Cell Res. Ctr., Univ. of São Paulo, São Paulo, Brazil, [2]Dept. of Genetics and Evolutionary Biology, BioSci.s Inst., Univ. of São Paulo, São Paulo, Brazil

In order to evaluate the benefit of offering a carrier screening investigation to larger segment of the Brazilian population, we have reassessed the results of individuals referred to our laboratory for carrier screening testing in the last years. Using a panel containing 234 to 340 genes associated with recessive and X-linked conditions we have performed 79 carrier-screening tests, comprising, altogether, 53 potential couples. For 27 couples, both partners were genotyped, while for the remainder, only one partner (21 female / 5 male) was tested. 62,3% of couples reported consanguinity or have been born in small nearby towns; 56,6% reported a previous, in most cases, undiagnosed genetic condition affecting themselves, their offspring, or another family member and, 24,5% informed both (consanguinity and a previous condition). Among consanguineous couples which both partners were tested, we have identified shared pathogenic variants in 47,1%, including all those who declared a deceased child or pregnancy loss. It was identified pathogenic variants in same gene in three out of nine non-consanguineous couples (both genotyped): one had reported a deceased child; one had likely pathogenic variants in *USH2A* and one was an IVF case with egg donation (*GJB2* gene). Though small, this cohort indicates the relevance of offering a broader program of carrier screening that might include not only at-risk families or manifestly consanguineous couples, but also couples from small towns. Although continentally large, Brazil has several isolated populations that might benefit from a carrier-screening program. Besides, the inclusion of fertility clinics should also be considered.
FAPESP/CEPID 13/08028-1

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Session III

PB2357 Cascade testing, reach, and acceptability or health system contact of at-risk relatives: pilot trial results.

Authors:

N. Henrikson[1,2], J. Zepp[3], P. Blasi[4], A. Scrol[4], M. Anderson[1], J. Ewing[1], J. Ralston[5], S. Fullerton[2], K. Leppig[6]; [1]Kaiser Permanente Washington Hlth.Res. Inst, Seattle, WA, [2]Univ. of Washington, Seattle, WA, [3]Kaiser Permanente Northwest Ctr. for Hlth.Res., Portland, OR, [4]Kaiser Permanente Washington Hlth.Res. Inst., Seattle, WA, [5]Kaiser Permanente Washington Hlth.Res. Instisute, Seattle, WA, [6]Kaiser Permanente of Washington, Seattle, WA

Cascade testing can improve outcomes for at-risk relatives of people with pathogenic variants associated with cancer risk, yet many relatives never learn of their risk. We piloted a new direct contact program where the care team contacts relatives to offer cascade testing for hereditary cancers. We co-designed the program with patients and families; key requirements included pre-test consent of probands and relatives to participate in the program and that the program should complement, not replace, existing patient-mediated disclosure. We assessed feasibility, acceptability, and limited effectiveness of the program. METHODS We conducted a single-arm, prospective mixed-methods intervention study at one U.S. integrated health system. We recruited adult probands awaiting pre-test genetic counseling for cancer risk; offered to contact at-risk relatives after testing; and contacted relatives to recommend cascade testing. We surveyed all study participants at 6-8 weeks to assess relative uptake of genetic testing and program acceptability. We assessed proband genetic testing outcomes using administrative data. We conducted semi-structured interviews with program participants and triangulated the results with survey data. RESULTS We invited 148 probands to participate in the study; 37% (n=55) enrolled. Enrolled probands were mean age 58; 75% female (n=41), 88% White race (n=45); and 7% Hispanic. The majority (69%, n=37) were married; 58% (n=31) reported college education or more. Half of probands (n=28, 51%) requested direct contact of relatives and provided consent to contact 101 relatives; 44% (n=45) of relatives consented to be contacted. We interviewed 32 participants from 22 families. Survey and qualitative data found high program acceptability for both probands and relatives. 83% of probands reported the program helped their family and 56% reported that relatives only learned of their risk because of the program. Of 35 probands with test results returned, 5 received pathogenic findings. At 6-8 weeks (n=40 relatives), 3 (8%) had received genetic testing; one of these three relatives reported a cancer risk-associated pathogenic variant. Five additional relatives (15%) reported plans to have genetic testing. CONCLUSIONS The direct contact program was acceptable and successfully reached at-risk relatives. The program worked as intended as a complement to patient-led disclosure. Future implementation efforts should seek to balance family preferences for pre-test consent and focusing outreach efforts on patients and families with confirmed pathogenic findings.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Session II

PB2359 Clinical genetics and genetic counselling external quality assessment (EQA) provision.

Authors:

M. Tabiner[1], K. Rack[1], R. Hastings[2], Z. Deans[2]; [1]GenQA, Oxford, United Kingdom, [2]GenQA, Edinburgh, United Kingdom

Background/Objectives: GenQA offers global external quality assessments (also known as proficiency testing) for the entire clinical genomics service. Clinical Genetics educational case scenarios allow an external assessment of patient consultations and counselling whilst over a hundred other programs assess laboratory processes, interpretation of results and reporting procedures. Since 2014, clinicians have been able to participate in a customised EQA program and subsequently review their clinical practice, learn from peers and demonstrate continuing professional development. EQAs for all clinical genetics disciplines are now available: dysmorphology; cardiovascular disorders; monogenic disorders; oncogenetics; inborn errors of metabolism. A separate program focussing exclusively on Genetic counselling is also offered. A new pilot EQA for Genomic multi-disciplinary team (MDT) working will be introduced in 2023.

Methods: Annual online multi-stage case scenarios which follow the patient pathway and reflect real clinical cases are provided for each EQA. Information regarding the clinical presentation of the proband is provided on a bespoke website, followed by specific questions about the initial consultation. Further details are provided sequentially and access to subsequent stages is blocked until all answers in a stage have been submitted. Participants in the Clinical Genetics EQAs must request appropriate genomic investigations and then interpret the results provided to establish a clinical diagnosis. There are also questions assessing the consequences of the diagnosis and content of the post-test appointment. Submissions are marked according to professional standards by an expert panel of international clinicians. A tailored score report with feedback comments and a participation certificate are provided to each centre.

Results: Satisfactory participation in these EQAs evidences training to a recognised level and demonstrates ongoing professional competency. The detailed EQA Summary Reports provide the expected answers and learning points for the case scenarios together with an overview of performance. These can be used for benchmarking the quality of clinical services provided.

Conclusion: These EQAs provide a mechanism for continuous assessment against agreed quality standards and therefore provide valuable training and educational opportunities for Clinical Geneticists and Genetic Counsellors. Participation in EQA can refine best practice and ultimately improve the quality of care for patients.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Session III

PB2360 Clinical Genome Resource Variant Classification and Curation Interface Workshops: Collaboration with Genetics Organizations in Low and Middle Income Countries.

Authors:

D. Ritter[1], F. Ali[1], A. Dalal[2], M. Hegde[3], R. Julius[4], A. Lumaka[5], A. Moirangthem[6], J. Morales[7], S. Phadke[6], A. Santani[8], K. Riehle[1], S-A. Yang[1], M. Mandell[9], M. Wright[9], T. Klein[9], S. Plon[1]; [1]Baylor Coll. of Med., Houston, TX, [2]Ctr. for DNA Fingerprinting and Diagnostics, Hyderabad, India, [3]PerkinElmer, Waltham, MA, [4]H3Africa CC, Univ. of Cape Town, Cape Town, South Africa, [5]Univ. of Kinshasa, Kinshasa, Congo, Democratic Republic of the, [6]Sanjay Gandhi PGIMS, Lucknow, India, [7]NHGRI, Bethesda, MD, [8]Lets Get Checked, Monrovia, CA, [9]Stanford Univ., Palo Alto, CA

The Clinical Genome Resource (ClinGen) is an NIH funded effort creating a central resource of clinically relevant genes and variants to improve genomic medicine. Dissemination of ClinGen's work is critical to ensure broad uptake. Recently ClinGen has focused on multiple educational approaches including educational workshops of genomics professionals to increase understanding of variant classification and use of ClinGen tools. Here we report the results of our recent outreach efforts to genetics professionals in low and middle income countries (LIMCs). In the last year, two dedicated workshops were held on variant classification using the ClinGen Variant Curation Interface (VCI). An in-person satellite workshop to the annual meeting of the Society for Indian Academy of Medical Genetics in India (83 participants) focused on ACMG/AMP variant classification and use of the VCI. Following an overview of classification guidelines and the VCI, participants were randomly divided into ten groups to work as teams on curating progressively challenging variants within test instances of the VCI. Although no post-workshop survey was given, feedback on better understanding of variant classification was received and we identified 14 new active user accounts from India in the production version of the VCI. We built on this experience for a virtual workshop with genetics trainees and professionals from Human Hereditary & Health Africa (H3Africa). The 3hr workshop provided a brief orientation to the ClinGen VCI and Allele Registry. The 61 participants were divided into country-based breakout rooms (from 6 countries) of 5-10 people with a "ClinGen lead" experienced in VCI curation. Information from the post-workshop survey (n=34, 55% response rate) revealed that participants were split between those with no or little experience versus those with substantial experience in variant classification (ACMG Classification: 12 inexperienced vs 22 experienced). Post workshop, 30/34 respondents would likely use the VCI, 30/32 requested annual VCI training, and 29/24 preferred future workshops be structured in shorter sessions over multiple days. Open text responses described a preference for teams based on prior experience. We identified 15 active user accounts established after the workshop, with a total of 3 affiliations from laboratories in Africa. These efforts showed that whether in-person or virtual, interactive training workshops resulted in VCI user uptake among 20% (29/144) of participants. Engagement with international genetic organizations has helped develop a workshop model to aid in uptake of genomic resources supporting genomic medicine in LMICs.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Session I

PB2361 † Clinical utility of all types of medically relevant secondary findings: An observational cohort study

Authors:

C. Mighton[1,2], R. Kodida[1], S. Shickh[2,1], M. Clausen[1], E. Reble[1], J. Sam[1], S. Grewal[1], S. Panchal[3,4], M. Aronson[5,4], S. Armel[6,4], T. Graham[7,8], N. Forster[9], J-M. Capo-Chichi[10,11], E. Greenfeld[12,11], A. Noor[12,11], I. Cohn[13], C. F. Morel[9,14], C. Elser[3,15], A. Eisen[7], J. C. Carroll[16,17], E. Glogowski[18], K. A. Schrader[19,20], K. K. W. Chan[21,2], K. E. Thorpe[22], J. Lerner-Ellis[12,11], R. H. Kim[6,14], Y. Bombard[2,1]; [1]Genomics Hlth.Services Res., St. Michaels Hosp., Unity Hlth.Toronto, Toronto, ON, Canada, [2]Inst. of Hlth.Policy, Management and Evaluation, Dalla Lana Sch. of Publ. Hlth., Univ. of Toronto, Toronto, ON, Canada, [3]The Marvelle Koffler Breast Ctr., Mount Sinai Hosp., Sinai Hlth., Toronto, ON, Canada, [4]Dept. of Molecular Genetics, Univ. of Toronto, Toronto, ON, Canada, [5]Zane Cohen Ctr. for Digestive Diseases, Mount Sinai Hosp., Sinai Hlth., Toronto, ON, Canada, [6]Bhalwani Familial Cancer Clinic, Princess Margaret Cancer Ctr., Toronto, ON, Canada, [7]Odette Cancer Ctr., Sunnybrook Hlth.Sci. Ctr., Toronto, ON, Canada, [8]Dept. of Molecular Genetics, Univ. of Toronto, Torono, ON, Canada, [9]Fred A. Litwin Family Ctr. in Genetic Med., Univ. Hlth.Network, Toronto, ON, Canada, [10]Genome Diagnostics Div., Toronto Gen. Hosp., Univ. Hlth.Network, Toronto, ON, Canada, [11]Lab. Med. and Pathobiology, Univ. of Toronto, Toronto, ON, Canada, [12]Pathology & Lab. Med., Mount Sinai Hosp., Sinai Hlth., Toronto, ON, Canada, [13]Div. of Clinical Pharmacology & Toxicology, Hosp. for Sick Children, Toronto, ON, Canada, [14]Dept. of Med., Univ. of Toronto, Toronto, ON, Canada, [15]Dept. of Med., Univ. of Toronto, Toronto, Toronto, ON, Canada, [16]Dept. of Family & Community Med., Univ. of Toronto, Toronto, ON, Canada, [17]Granovsky Gluskin Family Med. Ctr., Mount Sinai Hosp., Sinai Hlth., Toronto, ON, Canada, [18]Sanofi, North Bergen, NJ, [19]British Columbia Cancer Agency, Vancouver, BC, Canada, [20]Dept. of Med. Genetics, The Univ. of British Columbia, Vancouver, BC, Canada, [21]Div. of Med. Oncology & Hematology, Dept. of Med., Sunnybrook Hlth.Sci. Ctr., Toronto, ON, Canada, [22]Dalla Lana Sch. of Publ. Hlth., Univ. of Toronto, Toronto, ON, Canada

Introduction: Guidelines prioritize medically actionable secondary findings (SFs), but a broader spectrum of medically relevant results can be returned. We characterized the clinical utility of all types of medically relevant SFs by evaluating yield and changes to patient management. **Methods:** Observational intervention study in an RCT (Incidental Genomics RCT, NCT03597165). Adult cancer patients had genomic sequencing (GS) with return of primary cancer findings and option to learn multiple categories of SFs: medically actionable, Mendelian, early-onset neurodegenerative, carrier status, common disease risk, and pharmacogenomic (PGx) variants. Variants in patients' chosen categories were classified following ACMG/AMP criteria. P and LP variants were reported as SFs and returned by study genetic counselors (GC). GS reports and consult letters with recommendations from study GC and medical geneticist were sent to patients and their family doctors. Yield and clinical management changes were collected through chart review and patient-reported surveys up to 1 year after SF return. **Results:** Participants (n=139) were 85.7% female, average 55.1 years old, and 60.7% White/European. Overall, 100% of participants had ≥1 SF reported; 98.5% had multiple SFs. Yield was highest for PGx variants (97.8% of participants), followed by carrier status (90.1%), common disease risk (89.4%), Mendelian (27.4%), medically actionable (15.2%), and early-onset neurodegenerative (2.6%) variants. 1.4% of participants had SFs in ACMG-recommended genes. Overall, 19.4% of participants had a change in medical management attributed to their SFs, mainly appointments with specialists (11.5%) and family doctors (8.6%). Management changes were completed among 47.6% (10/21) of patients with medically actionable findings (e.g., medical genetics consultation), 29.4% (10/34) with Mendelian findings (e.g., imaging), 66.7% (2/3) with early-onset neurodegenerative findings (e.g., medical genetics consultation), 1.7% (2/118) with carrier status findings (e.g., family doctor appointment), 2.2% (3/135) with PGx variants (e.g., medication change), and 2.2% (2/118) with common disease risk variants (e.g., ophthalmologist consultation). **Conclusions:** Yield of SFs was high, largely due to reporting carrier status, PGx and common disease risk variants. SFs demonstrated clinical utility by prompting changes in patient management, including results that were not considered medically actionable. This implies potential benefits of returning a wide range of SFs; longer-term studies are needed to determine if management changes improve health outcomes.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Session II

PB2362 Clustering of heritability beliefs across diseases, traits, and individuals.

Authors:

S. Persky; Natl. Human Genome Res. Inst., Bethesda, MD

Introduction: Individuals' beliefs about the extent to which diseases and traits are heritable varies depending upon life experience, exposure to educational content, and broader biological and social belief systems. Heritability belief patterns can reflect intuitions and normative ideas underlying their formation and suggest ways beliefs might generalize based on shared underlying characteristics. In a practical sense, these beliefs are associated with behavioral and psychological response to health risks. Moreover, they have been tied to differential support for structural health policy programs with health equity implications.
Method: Data from an online survey panel (N=742) on heritability beliefs (rated 0-100%) for 30 traits and diseases were assessed for emerging clusters using hierarchical clustering. Individual cases were assessed using k-means cluster analysis for 4 participant clusters which were evaluated for demographic differences in membership.
Results: Four clusters of traits and diseases emerged with varying average perceptions of heritability; A) phenotypes with little perceived behavioral influence (type 1 diabetes; height), M=75, SD=18, B) common complex diseases (e.g., type 2 diabetes), M=49, SD=17, C) behavioral traits (e.g., tendency to eat in the absence of hunger) and cancers with well-known behavioral factors (e.g., melanoma), M=20, SD=19, D) biological characteristics (e.g., brain structure), M=60, SD=23. Participant cluster membership varied significantly by self-reported racial group, sexual minority status, gender, parental status, and age. Cluster membership did not differ by education, ethnicity, or BMI. Participant cluster 1 had high heritability beliefs across each trait/disease type and had disproportionately high gender and sexual minority and female membership. Cluster 2 was marked by relatively low heritability beliefs for biological characteristics alongside average heritability beliefs for other condition types and had high rates of female and parent membership. Cluster 3 had high heritability beliefs for trait clusters A and D, low for B and C and was disproportionately populated by White, male, non-parent individuals. Cluster 4 had the lowest levels of all heritability beliefs and had a high proportion of Black and male members.
Conclusion: Heritability beliefs across diseases and traits cluster relatively neatly in accordance with common conceptions of genetic and behavioral causal factors. Notably, participant clusters were unrelated to formal education and membership patterns suggest important sociocultural and lived experience forces underlying belief formation.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Session III

PB2363 † Community engagement in the development of genomic medicine research: Enrollment experience from the BabySeq Project

Authors:

S. Walmsley[1,2], B. Zettler[1,2], C. R. Horowitz[3,4], C. Genetti[5], S. Edwards[6], R. C. Green[1,2,7,8], I. A. Holm[5,8]; [1]Dept. of Med., Mass Gen. Brigham, Boston, MA, [2]Ariadne Labs, Boston, MA, [3]Dept. of Med., Icahn Sch. of Med. at Mount Sinai, New York, NY, [4]Dept. of Population Hlth.Sci. and Policy and Inst. for Hlth.Equity Res., Icahn Sch. of Med. at Mount Sinai, New York, NY, [5]Div. of Genetics and Genomics, The Manton Ctr. for Orphan Disease Res., Boston Children's Hosp., Boston, MA, [6]E.WE Fndn., Huntsville, AL, [7]Broad Inst., Cambridge, MA, [8]Harvard Med. Sch., Boston, MA

Introduction: Partnering with community and clinical stakeholders in research design and conduct may help develop studies that are practical, engaging and encourage diverse participation in research. We studied whether these stakeholder partnerships could increase diversity of participants in the BabySeq2 Project: a randomized controlled trial of genome sequencing (GS) in infants.
Methods: We convened a diverse Community Advisory Board (CAB) with representation from all recruitment cities (Boston, MA; New York, NY; Birmingham, AL) to provide input via virtual quarterly meetings with study staff. We also built partnerships with pediatricians in general practices, briefly trained them in genomics, and embedded research assistants from local communities at their practices to recruit families. CAB input was influential in several key domains. 1) We revised the protocol to require the consent of only 1 parent/guardian as the CAB felt a 2-parent requirement would be exclusionary. 2) We simplified and clarified consent language. 3) The CAB contributed to the formative research (qualitative and survey) used to develop protocols and recruitment procedures. 4) We changed the infant sample type from a venous draw to a heel stick. 5) We changed randomization to occur before sample collection and to not collect samples from the control group due to concerns of potential distress from unnecessary and uncomfortable testing, and skepticism over how secondary samples were used or destroyed.
Data Analysis: We compared the uptake rate and demographics (race; ethnicity; education; household income) of families recruited in BabySeq1, our prior study of newborn GS, with recruitment to date in BabySeq2. We analyzed between-group differences using Fisher's exact test.
Results: The overall uptake rate of BabySeq2 is significantly higher than BabySeq1 (13.8% vs. 6.9%, $p<0.001$). Additionally, families in BabySeq2 are significantly more likely to identify as non-White (91.9% vs. 17.6%, $p<0.001$), Hispanic (35.1% vs. 7.4%, $p<0.001$), have an education below a bachelor's degree (65.6% vs. 8.3%, $p<0.001$), and have a household income under $150,000 (92.9% vs. 43.7%, $p<0.001$).
Discussion: The uptake rate of the current BabySeq2 study is nearly double that of BabySeq1, and the currently enrolled cohort is significantly more diverse and nationally representative than in BabySeq1. Early recruitment experiences suggest that engaging with local communities and clinicians and recruiting in diverse practices can increase uptake and diversity in genomic research.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Session I

PB2364 Concept and Awareness of Genetic Counseling among Intersex Individuals in Pakistan.

Authors:

Z. Tariq, M. Rida, T. Ahmed, I. Yasin; Kinnaird Coll. for Women Univ., Lahore, Pakistan

The embodiment of gender with respect to the assigned biological sex plays an important role for a person to be recognized as male, female, or intersex i.e., atypical sex due to sexual variation at the chromosomal, gonadal, hormonal, or genital level. Biological perspective of intersexuality is not understood by many stirring extreme discrimination, violation, and stigmatization for the innate carnal ambiguities in intersex individuals. Altered genetic make-up might be the fundamental reason that might also be unknown to affected individuals whom could be informed, and supported through genetic counseling. This cross sectional study was conducted to determine the socio-economic factors, social dilemmas, depression, and genetic counseling scope among intersex individuals. A survey-based questionnaire having 50 questions including Hamilton Depression Rating Scale (HDRS) was filled by 90 intersex individuals. Descriptive Statistics, Univariate and Bivariate analyses, Pearson Correlation Tests, and Linear Regression Analysis using SPSS version 21 were applied. 33.33% male, female, and intersex genders each were assigned at birth. However, with the gonadal and hormonal changes, 6.7% considered themselves males, 40% females, and 53.3% intersex. One sample T-test indicated significant association of depression score (p=0.000), lack of support score (p=0.000), and awareness and need of genetic counseling score (p=0.00). One-way ANOVA confirmed significant association of marital status (p=0.006), societal discrimination (p=0.024), and familial support (p=0.02) in development of depression, and education (p=0.047), income (p=0.042), and governmental facilities (p=0.049) in lack of support. Genetic counseling awareness and need was significantly correlated with genetics knowledge (p=0.000), will to undergo genetic counseling (p=0.003), availability and accessibility (p=0.021), previous genetic counseling experience (p=0.000), and scope of genetic counseling (p=0.003). Linear Regression analysis confirmed the significant role of availability and accessibility (p=0.001), and previous genetic counseling experience (p=0.000) in the need and awareness of genetic counseling among the targeted population. To our knowledge, this is the first study on intersex individuals in this aspect worldwide. Severe gender-based societal discrimination and stigmatization of the intersex community leading to lack of support and development of depression in these individuals has been found. Poor genetic counseling services are another major drawback in providing relevant support to these individuals for which significant work has to be done.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Session II

PB2365 Consultagene: design and experience of a tele-genetic academic platform supporting genetic consultation, patient and provider education, and research engagement.

Authors:

D. Riconda, K. C. Worley, A. M. Wallis, C. A. Shaw, C. M. Holder, S. A. Nassef, S. A. Darilek, S. R. Lalani, B. Vuocolo, R. A. Sierra, P. L. Magoulas, S. M. Huguenard, B. H. L. Lee, Consultagene Group; Baylor Coll. of Med., Houston, TX

Introduction: Increasing use of genetic testing in medicine expands the need for genetic engagement and education for patients, caregivers, providers, and research participants. We designed an education platform for our academic medical department that is agnostic to use or configuration of Electronic Health Record systems and that supports different uses. **Methods:** The Consultagene platform is implemented with a secure computational infrastructure and supports different user roles (genetic provider, referrer, and client) with different types of interaction including content and activities for educational only, patient consultative, peer-to-peer and research engagement. Educational content was developed by our experienced geneticists and genetic counselors with focus group input from community members and reviewed for utility and understanding by platform users. **Results:** We have successfully implemented this platform in outpatient pediatric, adult, and prenatal genetic services. To date, 877 pediatric and 1456 prenatal/preconception educational journeys have been referred through the platform. We surveyed patients/parents on their satisfaction and understanding of two of our core educational videos. The majority (75 to 94%) of participants responding had positive assessments of the content and 73% reported improved understanding of the material. These numbers have been stable over time. Consultagene is being used for engaging in research studies including Project GIVE (UG3TR004047) where the platform is being used for outpatient genetic evaluation and return of WGS results for children living along the Texas-Mexico border. Since May of 2022, the PIs have engaged pediatricians at the University of Texas Rio Grande Valley (UTRGV) and community front-line providers in referring over 75 Hispanic individuals with undiagnosed rare genetic diseases and have accepted 50 pediatric patients for evaluation. Providers such as these who lack local access to geneticists are seeking peer-to-peer engagement with the platform as independent providers. **Conclusion:** The Consultagene platform supports genetic engagement in clinical practice, peer-to-peer consultation, and research applications with educational journeys that can be configured for different use cases. The materials and access are valued by patients and physicians. The platform enables genetic engagement that can reach to the populated but remote regions of Texas hundreds of miles from any genetics services.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Session III

PB2366 Curious but cautious: Patients' preferences for all types of clinically actionable genomic sequencing incidental results

Authors:

S. Shickh[1], R. Kodida[2], C. Mighton[3], M. Clausen[4], E. Reble[4], J. Sam[5], S. Grewal[2], S. Panchal[6], M. Aronson[7], S. R. Armel[8], T. Graham[9], C. Elser[10], A. Eisen[9], J. C. Carroll[11], E. Glogowski[12], K. A. Schrader[13], K. E. Thorpe[14], J. Lerner-Ellis[15]; [1]Univ. of Toronto, Toronto, ON, Canada, [2]Unity Hlth.Toronto, Toronto, ON, Canada, [3]St. Michaels Hosp. & Univ. of Toronto, Toronto, ON, Canada, [4]St. Michael s Hosp., Toronto, ON, Canada, [5]Unity Hlth.Toronto - St. Michael s Hosp., Toronto, ON, Canada, [6]The Marvelle Koffler Breast Ctr.-Mount Sinai Hosp., Toronto, ON, Canada, [7]Zane Cohen Ctr. for Digestive Diseases, Mount Sinai Hosp., Toronto, ON, Canada, [8]Bhalwani Familial Cancer Clinic, Princess Margaret Cancer Ctr., Toronto, ON, Canada, [9]Odette Cancer Ctr., Sunnybrook Hlth.Sci. Ctr., Toronto, ON, Canada, [10]The Marvelle Koffler Breast Ctr., Mount Sinai Hosp., Toronto, ON, Canada, [11]Dept. of Family & Community Med., Univ. of Toronto, Toronto, ON, Canada, [12]Sanofi, North Bergen, NJ, [13]Dept. of Med. Genetics, The Univ. of British Columbia, Vancouver, BC, Canada, [14]Dalla Lana Sch. of Publ. Hlth., Univ. of Toronto, Toronto, ON, Canada, [15]Pathology & Lab. Med., Mount Sinai Hosp., Sinai Hlth., Toronto, ON, Canada

Background: Genomic sequencing (GS) can generate a large volume of incidental results (IR), which have varying levels of clinical actionability. Although preference studies show that patients want all types of results, there is limited real-world evidence evaluating preferences of patients undergoing sequencing and actual return of IR. Here, we report preferences and changes in preferences among patients undergoing GS and return of all types of IR. Methods: We conducted analysis of preference data generated within the Incidental Genomics Trial (NCT03597165). Participants were adult cancer patients who previously received uninformative cancer panel results. Patients were offered GS and randomized to receive either cancer (control) or cancer+incidental results (intervention). Intervention participants chose between 5 IR categories: 1) actionable 2) common disease SNPs (i.e. polygenic risks), 3) Mendelian disorders, 4) early-onset brain disorders and 5) carrier status. Participants made selections at a pre-test counseling appointment and had the option to change preferences at a 1-week check in appointment. We analyzed preferences and changes using descriptive statistics. Results: 144 participants were randomized into the intervention arm, and were predominantly female (86%) of European ancestry (60%) with breast cancer history (57%). At the pre-test counseling appointment, participants were very interested in all categories, even those without clinical actionability. Nearly all participants (99%;142/144) chose actionable findings, 92% (133/144) chose carrier status, 90% (129/144) chose polygenic risks, 87% (125/144) chose Mendelian diseases and 81% (117/144) chose brain diseases. Three quarters (74%;106/144) chose all categories. At the 1-week follow up, 13% (18/144) changed preferences. The most common changes were for early-onset brain diseases: 8 participants added this category while 4 removed it. Conclusions: The majority of our participants selected to receive all types of IR, including those without actionability. A minority of patients changed selections at the follow-up appointment. Overall, the early-onset brain disease category had the least interest and most preference-sensitive changes, which aligns with current clinical practices that employ extensive genetic counseling protocols for patients undergoing testing for early-onset brain diseases. Our results emphasize the need for incorporating patient preferences, along with clinical evidence and costs, into future guidelines and highlight the importance of dynamic genetic counseling that provide patients an opportunity to reflect and change preferences.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Session I

PB2367 Cytogenomic characterization of breakpoints involving chromosome 8 in some Egyptian patients.

Authors:

S. Abd Allah[1], M. Mekkawy[1], A. kamel[1], E. Ashaat[2], M. Yousry[2], O. Eid[1], A. Mohamed[1]; [1]Human Cytogenetics, Natl. Res. Ctr., Cairo, Egypt, [2]Clinical genetics, Natl. Res. Ctr., Cairo, Egypt

Structural abnormalities of chromosome 8 have been reported in at least 1300 patients. Deletions of chromosome 8 may occur as an individual abnormality or with other chromosomal defects. Deletions of the short arm mostly occur with partial duplication for more proximal segments of 8p.
We report 5 patients with chromosome 8 rearrangements, presented to the Human Genetics Clinics and referred to the Cytogenetics department at the National Research Centre, Egypt.
The patients underwent clinical examination, IQ evaluation, anthropometric measurements, and imaging analysis. Karyotyping was performed for the patients and parents using GTG banding technique on peripheral blood lymphocytes. Fluorescence in situ hybridization (FISH) analysis was conducted using 8p & 8q subtelomeres probe set, whole chromosome painting 8 probe and other necessary probes when needed. Chromosomal microarray (CMA) analysis was performed using Genome-Wide Human SNP Array 6.
The patients' ages ranged from 4-10 years old. They were 3 males and 2 females. All exhibited distinctive dysmorphic features and intellectual disabilities (ID) ranging from mild to severe. They also had variable associated anomalies affecting the limbs, brain and heart. The karyotype analysis revealed in female and male siblings 46,XX,der(12) and 46,XY,der(12) respectively that were derived from maternal balanced translocation 46,XX,t(8;12)(q24.3;p13.3); add(18)(p11.3) in a female patient that was also inherited from maternal balanced translocation 46,XX,t(8;18)(p21;p11.3); add(8)(p23) in two male patients that were de novo.
CMA identified the extent and size of deletion and/or duplication in the affected patients; 15 Mb duplication in 8q24.3 that was associated with 3Mb deletion in 12p13.3 in two siblings, duplication of 8p21-8p23 that was combined by deletion in 18p11.3, another patient had 7.1Mb deletion in 8p23.3-8p23.1 followed by intermediate intact segment then 34 Mb duplication in 8p23.1- 8q21.3 (inverted duplication deletion). FISH analysis using the suitable probes confirmed CMA results in all patients and their parents.
We can conclude that breakpoints seem to be more concentrated at three intervals: interval 1 is at the telomeric end; intervals 2 and 3 are in 8p23 close to the beta-defensin gene cluster (DEFB) and olfactory receptor (OR) low-copy repeats (LCRs). CMA is a crucial tool for precise detection of breakpoints and the involved genes which are crucial for proper genotype-phenotype correlation. FISH analysis is essential in patients with chromosomal abnormalities due to parental balanced translocations for detection of copy number variants in the subtelomeres.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Session II

PB2368 Degree of uncertainty predicts emotional reaction to mock polygenic risk score results

Authors:

C. Jodarski[1], R. Gore[1], J. Yan[1], M. Kamen[2], K. Lewis[1], N. Reynolds-Lallement[1], S. Byers[1], B. Beers[1], M. Yavi[1], E. Ballard[1], C. Zarate[1], J. Austin[3], M. Walkiewicz[2], M. Similuk[1]; [1]NIH, Bethesda, MD, [2]NIAID, Bethesda, MD, [3]Univ British Columbia, Vancouver, BC, Canada

Psychiatric illnesses typically result from a complex interaction of genetic and environmental factors. Due to advances in the understanding of the many genetic factors contributing a small effect towards risk, polygenic risk scores (PRS) could be used to quantify an individual's risk of developing a psychiatric illness. While this technology is not currently being used on a clinical basis, it is important to evaluate how the individuals who may use PRS react to their results, so that appropriate pre- and post-test counseling can be provided. Therefore, this project aimed to identify factors that may be related to a positive or negative emotional response amongst individuals with treatment resistant depression (TRD), who received a hypothetical PRS result.

An online survey was sent to participants who underwent remote genetic counseling and genome sequencing one week after their initial appointment. The survey included a hypothetical scenario depicting a mock PRS for depression (randomized to low, average, or high risk PRS report). Affective response to the result was measured via the 12-item Feelings About Genomic Testing Results scale (FACToR), and potential domains of personal utility were evaluated using the 17-item Personal Utility Scale (PrU). Participants were also asked to identify which type of mock result they received, as a knowledge check.

We analyzed 64 responses from the 160 individuals who were invited to participate (40% response rate). Most participants (65.6%) correctly identified the type of mock PRS results received, and the remainder (34.4%) were unsure. The type of mock result received (low, average, or high risk) did not impact the degree of positive or negative emotions reported by participants ($p=0.17$ and 0.37, respectively). However, multiple linear regression (MLR) identified that increased uncertainty about PRS results and concerns about PRS technology were correlates of negative emotions towards PRS results ($\beta=0.54$, 0.29, respectively, all $p<0.01$). Additionally, MLR identified that greater perceived utility of results and less uncertainty were correlates of positive emotions towards PRS results ($\beta=0.42$, -0.25, respectively, all $p<0.05$).

These results suggest that while the degree of risk predicted by mock PRS result itself does not influence the type of emotional reaction experienced by participants in this study, the degree of uncertainty individuals feel about this result does. Given this knowledge, it will be important for those undergoing PRS testing to receive adequate counseling surrounding the test's utility and limitations so that results can be better understood within their personal and familial contexts.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Session III

PB2369 † Democratizing Gene Therapy Education for Patient Communities.

Authors:

V. Bonham, T. Jackson, H. Ramirez, K. Amini, R. Willard, A. Buscetta; NHGRI, Bethesda, MD

As novel and complex genetic therapies continue to emerge and gain approval from the US Food and Drug Administration, it is important for patient communities to have access to open-source, accurate, and understandable patient education materials (PEMs). Engagement with individuals living with the disease, support persons, and community advocacy groups is essential to understand their educational needs and gaps. Their voices should be heard by the research and scientific community, who are developing and implementing new genetic therapies. The Democratizing Education for Sickle Cell Disease Gene Therapy project used rational democratic deliberation and a community dialogue approach to engage stakeholders to create high-quality educational materials for gene therapy for sickle cell disease (SCD). A deliberative stakeholder engagement model was used to guide this project, employing both deliberative democracy and stakeholder engagement processes. Deliberative democracy is an inclusive approach to reaching consensus decision-making through participative and representative engagement. Community engagement encompasses individuals, constituent groups, and entities that play a role in developing and implementing new genomic technologies. We will examine the model employed by this project to engage patients, advocacy groups, researchers, clinicians, industry, and government representatives to create open-source PEMs for SCD gene therapy. Partnerships with community-based advocacy groups to disseminate these gene therapy education materials will be described. The method of using content syndication to republish to other websites and ensure regular updates that preserve their scientific accuracy will be presented. As the field of gene therapy advances, patient education content updates can be shared in real-time, thus providing the community with access to the most up to date and high-quality PEMs on gene therapy treatments and techniques.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Session I

PB2370 Designing a hybrid implementation-effectiveness study of clinical genome-wide sequencing in Ontario, Canada: integrating both geneticist and non-geneticist providers

Authors:

M. Mackley[1], G. Costain[1,2], T. Hartley[3,4], W. Lee[5,6], A. Szuto[1], M. K. Gillespie[3,7], C. Chisholm[3], K. Fooks[5], V. Venkataramanan[5], R. Mendoza-Londono[1], S. Sawyer[3,4], W. Ungar[5,6], K. Boycott[3,4], R. Z. Hayeems[5,6]; [1]Div. of Clinical & Metabolic Genetics, Dept. of Paediatrics, The Hosp. for Sick Children, Toronto, ON, Canada, [2]Program in Genetics and Genome Biology, SickKids Res. Inst., Toronto, ON, Canada, [3]Dept. of Genetics, Children's Hosp. of Eastern Ontario, Ottawa, ON, Canada, [4]Children's Hosp. of Eastern Ontario Res. Inst., Univ. of Ottawa, Ottawa, ON, Canada, [5]Program in Child Hlth.Evaluative Sci., SickKids Res. Inst., Toronto, ON, Canada, [6]Inst. of Hlth.Policy, Management and Evaluation, Univ. of Toronto, Toronto, ON, Canada, [7]Children's Hosp. of Eastern Ontario Res. Inst., Ottawa, ON, Canada

BACKGROUND: Genome-wide sequencing (GWS; including exome and genome sequencing) is increasingly recommended as a diagnostic test for rare diseases (RD), but timely access in many jurisdictions remains limited to those referred to medical geneticists. Enabling access to GWS for non-geneticist providers has the potential to expedite diagnosis for affected families and reduce wait times for the geneticist-led model of care that has limited capacity. However, evidence to guide the implementation of GWS into the practice of non-geneticists is lacking.
SETTING: Genome-wide Sequencing Ontario (GSO) is a harmonized multi-institutional service developed to deliver GWS for RD across Ontario, Canada. GSO's two-year pilot demonstrated the effectiveness of a hub-and-spoke model of delivery—where sequencing and informatics is done centrally with distributed analysis and reporting at two institutions—ensuring quality, consistency, efficiency, and sustainability province-wide. A diagnostic yield of 31% was achieved across sequencing strategies (singleton, duo, trio), and was significantly higher when the patient phenotype included developmental delay/intellectual disability. The mean overall turnaround time was 54.4 days (SD: 29.6), with 92% of cases reported in less than 12 weeks.
DESIGN: A mainstreaming model of care that enables a selected group of non-geneticist subspecialists to order clinical GWS will be developed, implemented, and evaluated by the GSO study team. To enable co-design, non-geneticist providers will be drawn from specialty areas that care for patients who meet eligibility criteria, including pediatric neurology and developmental pediatrics. Patient partners and knowledge users will be included in the sampling frame. Using a multi-phase mixed methods design, we will: (i) conduct semi-structured qualitative interviews and expert consultations to explore barriers and facilitators to the uptake of GWS among non-geneticist providers; (ii) co-design, test, and implement a mainstreaming model of care; and, (iii) measure both GWS outcomes (i.e., diagnostic utility, clinical utility, and timeliness) and implementation outcomes (i.e., acceptability, feasibility, fidelity, costs, sustainability, and penetration of the new delivery model) at the level of both providers and patients.
INTERPRETATION: Findings from this work will provide evidence related to the performance of GWS in a service delivery model that includes non-geneticist care providers. Outcomes will inform provincial and cross-provincial policy and implementation planning related to the organization and delivery of clinical-grade genome diagnostics for RD.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Session II

PB2371 Developing a programmatic evaluation for programs focused on improving access to genetic services for underserved populations

Authors:

M. Lyon[1], D. Maiese[1], L. Creel[2]; [1]American Coll. of Med. Genetics and Genomics, Bethesda, MD, [2]Univ. of Colorado Anschutz Med. Campus, Sch. of Med., Dept. of Med., Div. of Hlth.Care Policy and Res., Aurora, CO

Due to the scarcity of genetic services and resources in parts of the United States, the Health Resources and Services Administration (HRSA)/ Maternal and Child Health Bureau (MCHB)/ Genetic Services Branch (GSB) funds the seven Regional Genetics Networks (RGNs), the National Coordinating Center for the Regional Genetics Networks (NCC), and the National Genetics Education and Family Support Center (NGEFSC) to improve access to genetic services for underserved populations. Through a regional infrastructure that covers up to eight states per network, the system provides education and workforce development (e.g., general genetics and genetics policy) to individuals, families, and health care and public health professionals. The system also focuses on linking patients to genetic services and supports telehealth training and service delivery. The RGN/NCC/NGEFSC system uses common performance measures to assess the five objectives set forth in the RGN Notice of Funding Opportunity. In this evaluation, medically underserved is defined as living in a geographic region designated by HRSA as a Health Professional Shortage Area. ZIP Codes are collected to determine the reach of the system into medically underserved communities. The goal of this cross-regional evaluation is to assess the system's year-to-year progress toward reaching underserved populations and connecting individuals and families to needed genetic services. Since 2017, the RGNs, NCC, and the Family Center have collected data from a variety of sources through a mixed-methods approach. Data collected includes registration data, agreements with clinics to provide data, and follow-up surveys to RGN participants. Data are compiled by each program and submitted to NCC via REDCap to create annual reports to HRSA about the system's reach. This data set includes the number of individuals that received education or training on genetics; linkages created to connect patients to genetic services; providers trained or aided with TA in telehealth; and RGN-supported clinics that use telehealth modalities. Through a consensus-building process among the RGNs, NCC, Family Center, and HRSA, we have streamlined REDCap, and agreed upon definitions for each data element with reference documents to assist programs in understanding how data should be counted. Using the data collected through this evaluation, we provide publicly available data on the outcomes of our system (nccrcg.org/our-impact), which includes educating over 8,000 providers and reaching more than 4,000 individuals and families, thus reaching people in every state and territory.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Session III

PB2372 Development of a comprehensive framework for understanding the scope of undiagnosed disease

Authors:

D. Adams[1], E. Macnamara[2], C. Tifft[3], C. Toro[1], W. Gahl[4]; [1]NIH, Bethesda, MD, [2]NIH, Bethesda, MD, [3]NIH Office of the Clinical Director, Bethesda, MD, [4]NHGRI (NIH), Bethesda, MD

INTRO: The NIH Undiagnosed Diseases Program (UDP) has now been enrolling participants since 2008, and as part of the Undiagnosed Diseases Network (UDN) since 2015. Participants have diseases that remain undiagnosed after extensive standard of care evaluation. This period has also seen the emergence of a global undiagnosed disease community. The root causes for diagnostic failure are complex. Examples include undefined new diseases, lack of diagnostic resources and atypical phenotypes. A comprehensive system for categorizing the circumstances leading to failed diagnoses is needed, both to understand the scope of the associated medical need and to prioritize research and clinical resources.
METHODS: The study authors utilized their experience with the UDP and UDN to construct a framework for describing mechanisms leading to the failure to reach a diagnosis. Individual UDP cases were used to refine the framework by studying real-world examples. Mapping of 314 past, diagnosed UDP cases to causes of diagnostic failure generated preliminary estimates of the relative contribution of individual factors across the cohort.
RESULTS: The patients evaluated by the UDP and UDN represent a subset of all undiagnosed disease patients. There is a bias toward diseases with manifestations of sufficient duration to persist through standard of care and subsequent UDN clinical evaluations. Prospective participants may be excluded if they have not had a standard of care workup. Geographic biases arise from a largely US study cohort. Of the 314 cases, 257 (82%) were associated with a gene-associated disease—likely reflecting program enrollment priorities. For these reasons, quantitative information about the frequency of individual contributors to diagnostic failure was specific to the UDP/UDN cohort and not generalizable. Despite these limitations, the process allowed for the generation of a standardized vocabulary for use in prospective work.
CONCLUSIONS: We present a framework for categorizing pathways by which undiagnosed disease patients come to the end of available clinical evaluation without having achieved a diagnosis. Further work will be required to understand the relative contributions of framework elements in a broad spectrum of healthcare settings. To accomplish that goal, new collaborations and data sources will be needed to augment the UDP/UDN experience. We anticipate that further development and study of the framework will allow for refined targeting of efforts to mitigate delays and diagnostic failures.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Session I

PB2373 Disclosure of genetic test results for stroke among stroke-free West Africans

Authors:

A. Fakunle[1], A. Okekunle[2], O. Asowata[2], O. Akpa[2], J. Akinyemi[2], F. Sarfo[3], A. Akpalu[4], R. Obiako[5], K. Wahab[6], R. Akinyemi[2], B. Ovbiagele[7], M. Owolabi[2]; [1]Osun State Univ., Osogbo, Nigeria, [2]Univ. of Ibadan, Ibadan, Nigeria, [3]Kwame Nkrumah Univ. of Sci. and Technology, Kumasi, Ghana, [4]Univ. of Ghana, Accra, Ghana, [5]Ahmadu Bello Univ., Zaria, Nigeria, [6]Univ. of Ilorin, Ilorin, Nigeria, [7]Univ. of California, San-Francisco, CA

Background: Genetic testing is an essential component of human genetics. With the recent revolution in genetic testing and its promising contribution towards mitigating the burden of stroke in sub-Saharan Africa, it is crucial to assess Africans' perception, preference, and attitude towards the disclosure of stroke genetic test results.
Objective: This study evaluated the preference, perceptions, and attitudes towards the disclosure of genetic testing results for stroke among stroke-free controls in the SIREN-SIBS Genomics Study.
Methodology: We employed a quantitative data collection technique to assess the perception, preferences, and attitudes of stroke-free controls in Ghana and Nigeria towards disclosure or non-disclosure of different tiers of stroke genetic results using the SIREN questionnaire. Multivariable regression model was used to investigate sociodemographic factors that influence Africans' preferences, perceptions, and attitudes towards the disclosure of genetic testing results for stroke.
Results: A total of 351 stroke-free controls were included in the current study. Respondents' mean (SD) age was 51.26 (14.68) years, with a male preponderance (58.0%). The majority of the respondents (66.1%) were against the opinion that getting the results of the genetic test for stroke is of no use, while 95.6% claimed that they prefer a face-to-face disclosure of their genetic results. More than half (66.0%) had a negative attitude towards disclosing genetic test results for stroke. Age >70 years was associated with a negative attitude towards disclosure of genetic test results for stroke [odds ratio (OR) 4.19; 95% confidence interval 1.04-6.88].
Conclusion: These findings necessitate culturally sensitive interventions for continuing education to improve individuals' attitudes towards the disclosure of genetic test results for stroke in Africa.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Session II

PB2374 Disease-related and recreational genetic testing in Australia: insights into associated health and sociodemographic factors from an analysis of 45,061 participants of the 45 and Up Study.

Authors:

J. Steinberg, D. Goldsbury, H. Tanha, Y-J. Kang, A. Smit, K. Dunlop, L. Petelin, P. Ngo, C. Tang, H. Hui, N. Meagher, M. Merritt, M. Weber, A. De Fazio, A. Cust, K. Canfell; The Univ. of Sydney, Sydney, Australia

Genetic testing has considerable promise for precision health, with tests increasingly available for disease risk and tailored treatment. Australia presents an example of a high-income country with universal healthcare, but only limited nationally re-imbursed genetic testing, with many tests covered by health services or directly by individuals. To help realise the potential of genetics, it is thus important to understand current patterns of testing, including differences by socioeconomic status.

We conducted a large-scale analysis of self-reported genetic testing and its association with sociodemographic and health characteristics (including family history for multiple diseases) in Australia, using the 45 and Up Study (a longitudinal cohort of 267,357 participants aged 45+ years, recruited 2005-2009) and linked NSW Cancer Registry data.

Among 45,061 participants who completed follow-up questionnaires in 2020, 9.2% (95%CI 8.3-10.0%) reported ever having a genetic test, with 3.9% (3.0-4.8%) reporting a disease-related test (risk, diagnosis, or treatment) and 5.2% (4.3-6.1%) a recreational test (ancestry/diet/fitness/other). Estimates adjusted to match national education or income levels were similar (e.g. any genetic test: 8.4-9.5%).

Variation between population subgroups was substantial. For example, disease-related testing was significantly associated with relatively younger age (e.g. 5% age 60-69 vs 2% age 80+ years, adjusted odds ratio (aOR) 2.50), a record of breast (aOR 4.14) or bowel cancer (aOR 2.50), cardiovascular disease (aOR 1.37), family history of breast (aOR 1.67) or ovarian cancer (aOR 2.41), but also university education (aOR 1.48 vs school only; all above estimates adjusted for all characteristics analysed and significant after multiple-testing correction).

As cancer is a key area for disease-related testing, we also specifically considered participants with a linked cancer diagnosis record. Here, disease-related testing was strongly associated with a more recent cancer diagnosis (e.g. 2015-2019 vs pre-2005: aOR 1.95). Associations with family history increased strongly when multiple generations were affected, or when restricting the analysis to participants with the same cancer. However, there remained consistent evidence for an association between testing and higher education (disease-related testing: aOR 1.37 for university vs school only, $p<0.05$).

In summary, our results provide detailed insights into current access to genetic testing and how it differs between population subgroups in a high-income country, re-enforcing the need for further research on equitable access to genomic technologies.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Session III

PB2375 Dissecting racial disparities in the implementation of pediatric whole genome sequencing as a first-line diagnostic test: insights from SeqFirst.

Authors:

J-H. Yu[1], K. MacDuffie[1,2], O. Sommers[1], K. Anderson[1], M. Sikes[2], L. Kruidenier[2], H. Gildersleeve[1], A. A. Scott[2], K. Buckingham[1], A. Snook[3], P. Kruszka[3], D. E. Miller[1], J. X. Chong[1], D. Veenstra[1], K. M. Dipple[1,2], T. L. Wenger[1,2], D. Doherty[1,2], M. J. Bamshad[1]; [1]Univ. of Washington, Seattle, WA, [2]Seattle Children's, Seattle, WA, [3]GeneDx, Stamford, CT

Background: Racial disparities in access to genetic services, whole genome sequencing (WGS), and a precise genetic diagnosis (PrGD) are well-documented and contribute to inequitable outcomes among infants and children with a suspected genetic condition. The SeqFirst Study aims to improve equitable access to a PrGD by offering WGS as a first-line diagnostic test to families of infants and toddlers without an obvious environmental explanation for their condition. SeqFirst enrollment through a quaternary Children's Hospital NICU suggests that BIPOC patients/families with genetic conditions may be less likely than white families to receive genetic services. Such access disparities might be due to a range of structural and social determinants that manifest in interactions between family, providers, and institutions (e.g., family's distrust, provider bias). **Methods:** Semi-structured key informant interviews with n=9 neonatologists and n=10 neurodevelopmental providers involved in SeqFirst recruitment were conducted to investigate factors underlying observed racial disparities in access to a PrGD as well as attitudes toward genetic research, testing, and experience of SeqFirst recruitment. **Results:** Despite intentions to avoid discriminating between BIPOC and white families, providers articulated a range of considerations and practices that may lead to disparities. Providers were adamant that race does not and should not matter; yet thought it may influence their ability to detect dysmorphic features contingent on medical training and experience serving racialized populations. Race was thought to complicate evaluations along with infant age, size, development, and parent availability at bedside. Nearly all identified parent-provider language discordance and differences in "medical literacy" as challenges to offering genetic tests and services. Together these factors characterized newcomer families in contrast to those who are English-proficient, highly educated, wealthy, privileged, and self-advocating. Limited institutional resources (e.g., suboptimal interpreter services, availability of referral pathways) and social determinants (e.g., occupationally-constrained parent availability) left families vulnerable to access disparities. **Discussion:** A constellation of factors and practices that interact throughout a family's healthcare journey may contribute to racial disparities in access to genetics. Efforts are needed not only to address biases in provider training but also institutional resource constraints and barriers caused by social determinants to achieve equitable access to genetic testing and services.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Session I

PB2376 Do the Benefits Outweigh the Barriers?: Clinical Implementation of Parent-of-Origin Genomic Analysis

Authors:

A. Roston[1], A. Virani[1], K. Schrader[2]; [1]Univ. of British Columbia, Vancouver, BC, Canada, [2]BC Cancer / Univ British Columbia, Vancouver, BC, Canada

Background: Parent-of-origin assignment (POAga) is an emerging genetic testing modality, in which DNA can be reliably assigned to either paternal or maternal origin, even in the absence of parental testing. The ability to determine inheritance from a single genetic test represents a paradigm shift in clinical genetics, as it mitigates the need for the 'trio-based' testing approach during variant segregation. However, there are several barriers that may hinder its translation from the research laboratory to the clinic. Benefits: POAga promises numerous benefits for patients, families, and the healthcare system. First, POAga can inform screening and management recommendations for variants whose effects depend on the contributing parent. This will be especially beneficial to those without available parents, including patients who are adopted, elderly, or who have immigrated. Additionally, POAga prediction will likely assist in curating variants of uncertain significance (VUS), which are known to disproportionately affect patients of non-Western European ancestry. At the family level, POAga data helps direct cascade testing to one side of the family, and allows for more refined genetic counselling and screening recommendations. Finally, this ability to refine surveillance recommendations among patients may ultimately lead to decreased healthcare costs and improved resource stewardship. Barriers: Because POAga predicts the genotypes of untested parents, parental autonomy and confidentiality must be considered. As an additional concern, the ability to predict parental carrier status within a single test will lead to a duty to warn parents of pathogenic results, and on an unprecedented scale. POAga also provides the greatest direct benefit to a relatively small proportion of individuals (e.g. variants with parent-of-origin effects *and* unavailable parents). Compounding this issue is the fact that POAga relies on cutting edge, high-cost technologies; therefore, it is very likely that, at least in the initial stages, there will be unequal access to this testing modality. Finally, the initial implementation of POAga, which will require significant user education and policy changes, will likely lead to temporary increases in system costs. Conclusions: Overall, while a promising new technology with numerous clinical benefits, POAga warrants a careful discussion of its ethical and practical implications.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Session II

PB2377 Equity and human rights perspectives on multigene panel testing for cancer: Health insurance coverage in Japan and Switzerland.

Authors:

K. Nakasato[1], C. Manz[2], K. Kato[1]; [1]Osaka Univ. Graduate Sch. of Med., Suita, Japan, [2]Univ. of Lausanne, Lausanne, Switzerland

Advancements in genomic medicine have led to prospects in a wide range of clinical fields, including oncology. In particular, developments in next-generation sequencing (NGS) multigene panel tests have led to the possibility of tailoring treatment to the specific genomic markers of a patient's cancer. However, several practical challenges have yet to be fully addressed. One of these potential challenges is whether and under which circumstances multigene panel testing for cancer is covered by health insurance, which can have broader implications for health equity and human rights. Such discussions have often been centered around countries at the forefront of implementing precision cancer medicine into routine care such as the United Kingdom and the United States. Comparatively, less is known about the situation in countries like Japan and Switzerland, which have a strong foundation in genomic research and national health insurance but have not implemented genomic medicine to the same extent. As such, we investigated the approaches to insurance coverage for cancer multigene panel testing in Japan and Switzerland. Results have shown that while there is an overlap between the two countries regarding the use of multigene panel tests, there are noticeable differences in the circumstances under which these tests are covered by national health insurance. With these results, we will discuss the implications for policy-making decisions for equitable access to genomic medicine, as well as broader discussions on non-discrimination in health care.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Session III

PB2378 Ethical considerations for returning sex chromosome aneuploidy as a secondary finding.

Authors:

A. Buscetta, J. Aickareth, H. Ramirez, J. Floyd, V. Bonham; NIH/NHGRI, Bethesda, MD

The American College of Medical Genetics and Genomics (ACMG) has published recommendations for reporting secondary findings (SF) in clinical exome and genome sequencing. These recommendations include a minimal list of pathogenic and likely pathogenic variants, which is updated annually. The ACMG recommendations are often used as a guideline and educational resource for research groups who are returning results. The ACMG encourages the community to submit nominations of actionable genes to be considered for inclusion. They state these recommendations are not inclusive or exclusive of all procedures or tests and professional and clinical judgment is warranted in specific clinical circumstances. Currently, sex chromosome aneuploidies (SCA) are not included on the ACMG list. Approximately, 75-90% of SCA cases are undiagnosed during an individual's lifetime. With advances in prenatal testing, studies have shown better outcomes with earlier diagnosis of an SCA. Given the variation in phenotype and severity of SCAs, it is important to consider the diagnosis and the impact on the individual prior to returning the result. Three ethical principles to help guide decisions to return SF have been proposed: validity, value, and volition. Validity and value concern the accuracy and usefulness of the diagnosis to the person. Volition centers around whether or not the person wants to receive the information, i.e. informed consent. We present a case report of a research participant living with sickle cell disease (SCD), who consented to whole genome sequencing (WGS) and receiving SF. The participant was found to have a variation in the copy number of sex chromosomes, which indicated the presence of XXY or Klinefelter syndrome. Prior to contacting the patient to obtain a new sample for CLIA confirmation, an ethics consultation was received to weigh potential risks and benefits of returning this result. The following relevant considerations were discussed: clinical actionability, fertility issues, and psychosocial effects. The final ethical analysis determined there was enough direct clinical value and personal utility to support disclosing the results, however, not enough to deem this ethically obligatory. We examine the ethical dilemma of returning secondary results for sex chromosome aneuploidy.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Session I

PB2379 Evaluation of Genetic Counseling Services by Genetics Teams in Korea

Authors:

I. Choi[1], H. Shin[2], M. Kang[2], H. Bae[3], S. Hwang[4], G-H. Kim[2], B. Lee[4]; [1]Univ. of Ulsan Coll. of Med., Seoul, Korea, Republic of, [2]Asan Med. Ctr., Seoul, Korea, Republic of, [3]Asan Med. Ctr. Children s Hosp., Seoul, Korea, Republic of, [4]Asan Med. Ctr. Children's Hosp., Seoul, Korea, Republic of

In South Korea, the provision of clinical genetics services started with the opening of genetics clinics in the 1990s. Despite the increasing number of geneticists, the number of geneticists remains insufficient compared with the number of patients requiring genetics services. It is important for medical geneticists and genetic counselors to collaboratively provide genetic counseling services. This study aimed to examine Korean patients with rare genetic disorders and their families who visited the medical genetics center in a tertiary general hospital in terms of their perception and satisfaction regarding genetic counseling services. In addition, the genetic counseling service model of a genetics team, which involves collaboration between medical geneticists and genetic counselors, was evaluated. Among the study participants, 164 participants (80.8%) were aware of the genetic counseling services and 135 participants (66.5%) were aware of the role of genetic counselors. The score of satisfaction with genetic counseling services provided by the genetics team was 8.19 ± 1.68 out of 10. The satisfaction score was highest for the item informing about the medical expense support program (4.46 ± 0.93 points out of 5), followed by the items providing medical knowledge about diseases (4.34 ± 0.74 points), helping with making decision about diagnostic work-up/options for genetic testing (4.34 ± 0.82 points), explaining about the ways of being inherited and the recurrence risk within the family (4.24 ± 0.84 points), providing psychosocial support (3.96 ± 0.97 points), and informing about additional resources, including patient support groups, special education, and rehabilitation (3.43 ± 1.24 points). In conclusion, genetic counseling services provided by collaboration between medical geneticists and genetic counselors were effective in providing information and can be helpful in diagnosing, treating, and managing patients. As a genetic team approach, collaboration between medical geneticists and certified genetic counselors would be useful in providing information and in diagnosing, treating, and managing patients.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Session II

PB2380 Exploring precision medicine through the lens of the LGBTQ+ community.

Authors:

D. Diaz Caro[1], B. Rolf[2], D. Kochhar[3], M. Campion[4]; [1]Columbia Univ. Irving Med. Ctr., New York, NY, [2]Univ. of Washington, Seattle, WA, [3]Seattle Univ., Seattle, WA, [4]Stanford Univ., Stanford, CA

Background: For the benefits of precision medicine (PM) to reach all patients, initiatives must understand and address the needs of all underrepresented populations. Previous studies have explored the perspectives of ethnic and racial minorities regarding barriers, facilitators, concerns, and benefits of PM; however, no study has surveyed individuals that identify as lesbian, gay, bisexual, transgender, queer, and/or other marginalized sexual and gender identities (LGBTQ+). This study aims to address this gap by exploring the perspectives that LGBTQ+ community members have about PM. **Methods:** Participants were identified through a LGBTQ+ community center recruitment survey that gathered demographics and introduced PM via an educational video. Participants were randomly selected for interviews using a stratified sampling approach. Semi-structured interviews were conducted via videoconferencing with 12 participants that self-identified as LGBTQ+. The interviews gathered information on understandings of PM, perceptions of PM initiative trustworthiness, benefits, expectations, concerns, and foreseeable barriers and facilitators for PM participation. Interview recordings were transcribed, and content analysis was conducted to identify themes. **Results:** Interviewees had varying levels of understanding of PM prior to participating in our study. Expectations of PM encompassed addressing general and LGBTQ+ specific health concerns and social inequities that affect health outcomes. Barriers to the adoption of PM included negative past experiences in healthcare, negative social and personal implications of identity disclosure, and lack of access and resources. A major concern identified was the potential misuse of LGBTQ+ identity for discriminatory practices. Interviewees indicated that they would view PM initiatives as trustworthy if barriers and concerns were addressed through a demonstrated commitment to accountability, credibility, accessibility, inclusivity, transparency, and privacy. Facilitators included qualities that would demonstrate trustworthiness such as endorsement from trusted LGBTQ+ practitioners or institutions, inclusive content, and the dissemination of information through community trusted sources. **Conclusion:** The study findings identified specific areas that future PM initiatives need to address in order to foster inclusivity and recognition of the LGBTQ+ community. Further research should investigate the perspectives of LGBTQ+ individuals participating in PM initiatives and include a broader spectrum of intersecting identities not characterized in this study.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Session III

PB2381 Exploring the Impact of Secondary Genetic Findings on the Psychosocial Wellbeing of Caregivers of Cleft Children

Authors:

A. Oladayo[1], V. Sule[2], Y. Oshodi[3], A. Adekunle, A[4], W. Adeyemo[4], O. Ogunlewe[4], A. E. Ajadi[3], B. Aikomo[3], O. Aiyelomi[3], A. Babatunde[3], T. Busch[5], A. Alade[5], M. Olujitan[1], E. Aladenika[5], W. Awotoye[6], L. Gowans[7], M. Eshete[8], C. Campbell[5], D. J. Caplan[1], O. Adeagbo[1], P. Mossey[9], A. Adeyemo[10], J. Murray[5], A. Prince[5], A. Butali[1]; [1]Univ. of Iowa, Coralville, IA, [2]Uiowa, Coralville, IA, [3]Univ. of Lagos, Lagos, Nigeria, [4]Univ. of Lagos/ Lagos Univ. Teaching Hosp., Lagos, Nigeria, [5]Univ. of Iowa, Iowa City, IA, [6]Univ. of Iowa, Iowa, IA, [7]Kwame Nkrumah Univ. of Sci. and Technology, Kumasi, Ghana, [8]Addis Ababa Univ., Addis Ababa, Ethiopia, [9]Univ. of Dundee, Dundee, United Kingdom, [10]NIH, Bethesda, MD

Introduction: Caregivers with children affected by orofacial clefts (OFCs) in low- and middle-income countries often experience significant emotional and psychological problems, including social stigma. The discovery of secondary findings (SFs) of potential health importance from whole-genome sequencing (WGS) may further exacerbate this since genetics and heritability may raise additional ethical and social issues. However, little is known about the experiences of caregivers and the impact of receiving SFs in the context of OFCs in sub-Saharan Africa. Our study aimed to determine if identifying SFs will affect the psychosocial well-being of caregivers of cleft children in Nigeria. **Methods:** Two Focus Group Discussions were conducted with 22 purposively sampled caregivers of children who underwent WGS analysis and attended the cleft clinic at the LUTH, Lagos, Nigeria. We investigated participants' experience regarding genomic testing, knowledge about SFs, psychosocial impact and coping strategies. Audio recordings from discussions were transcribed verbatim and analyzed using content analysis. **Results**: The average caregiver age was 34.3 ± 11.3 years, and 19 out of the 22 participants were females. Four themes and thirteen subthemes emerged**.** These include **pervasive burden** {child care practices, financial hardship hindering follow-up on SFs, physical and mental health impact and effect on family unit}, **concern** {causes of cleft, acting on SFs and community treatment of affected child}, **coping strategies** {holding on through faith, family support and peer support at cleft clinic} and **need for psychological support** {coping with cleft diagnosis, associated stigmatization and requirement of caring for a cleft child}. Although WGS results sparked relief and concerns, caregivers reported that SFs might potentially exacerbate existing stress. **Conclusion:** Caregivers faced numerous difficulties, including psychosocial issues, as they cared for a child with a cleft. However, they were able to draw strength and optimism from peers at the cleft clinic. Promoting a healthy collaboration between care providers and caregivers could lay the foundation for improving their understanding and utility of testing results while linking them to appropriate resources that could provide needed psychosocial support and alleviate the negative caregiver experience associated with OFCs in limited resource settings. Similar studies could provide evidence to plan for mental health care for caregivers of cleft children in the presence of genetic risks. Thus, it is recommended that this study is repeated in other regions across the continent.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Session I

PB2382 Fighting genetics harms with education: a course on genetics, ethics, and society

Authors:

R. Ungar, R. Patel, A. L. Pyke, E. Greenwald, A. Adijoelja, J. Gomez-Stafford, A. Narain, N. Hunter, P. Goddard, T. Mackenzie, D. Cotter, M. Chakraborty, D. Martschenko; Stanford, Stanford, CA

Despite decades of scientific activism, we are still grappling with the misuse of genetics to justify racist, sexist, ableist, and transphobic ideology. However, few training opportunities provide students with the tools to understand how society and science are related, or to interrogate their own roles as part of the scientific system.
To address this gap, we as a group of trainees ideated, designed, and led a course using contemporary design principles with the following learning goals:
1. Connect the historical context of genetics research to its modern-day practice
2. Evaluate the social and ethical implications of genetics research
3. Analyze how societal norms and structures, along with personal identities, biases, and responsibility, impact the conduct of scientific research
We first designed this as a six-session, discussion-based mini-course in Spring 2022, and focused on the history of genetics; race and ancestry; forensic genetics; and reproductive genetics. We then expanded this to a ten-week course in Spring 2023, and added new topics, including behavioral genetics; sex and gender; genetics and (dis)ability; and scientific activism. We situated our discussions of these topics in the language of positionality, reflexivity, and social responsibility, which enabled students to ideate actionable solutions for social harms. We have released all of our learning goals, slides, and related materials here: bit.ly/stanford-genethics.
We conducted an IRB-approved survey to evaluate the impacts of this course. Following our class in Spring 2022, every student strongly agreed that "research is influenced by society", and many shared that it was important to them that research "does not re/create harm". Despite the discussion of difficult topics, this course empowered students; all students agreed they "can make a change towards a more equitable, inclusive, and just scientific research and structure"; one student mentioned they "learned how to be a better advocate for more ethical science". It is important to note 73% of our students identified as underrepresented in STEM, which highlights the need to better engage the broader academic community. We will provide additional insights gained from Spring 2023, as our course was still in progress at time of abstract submission.
To achieve the collective goal ASHG set to "consistently acknowledge and oppose harms and injustices" we propose an educational model that can be adopted and iterated on at other institutions or workshops. We argue this training is an important step to enable collective action towards equitable and inclusive scientific research.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Session II

PB2383 Fragile X Syndrome: Evaluation of a documentary used to share genomic research findings in Cameroon.

Authors:

K. Kengne Kamga[1,2], M. Lebogang[2], N. Seraphin[3], A. Wonkam[4], J. De Vries[2]; [1]Regional Hosp. Limbe, Limbe, Cameroon, [2]Univ. of Cape town, Cape town, South Africa, [3]Univ. of yaounde 1, Yaounde, Cameroon, [4]Dept. of Genetic Med., JHU, Baltimore, MD

Fragile X Syndrome (FXS) is the most common genetic cause of intellectual disability and Autism Spectrum Disorder (ASD). Recently, a large family in Cameroon with FXS was studied, and a documentary on FXS was created to share the research findings. This paper reports on the post-production evaluation of the FXS-documentary. Between March 2022 and January 2023, we surveyed individuals from communities in Cameroon's western and southwestern regions. Our survey consisted of a 17-question questionnaire. To participate, we obtained consent from individuals and let them watch the 10-minute documentary on FXS and complete the questionnaire. We used EPI info 7 for quantitative data, while Nvivo 12 was used for qualitative data. Our statistical significance was set at $p \leq 0.05$. Our survey gathered information from 42 individuals residing in the southwest and western regions of Cameroon. Most participants were female and had completed at least a high school education. The participants had diverse occupations, with healthcare workers (33.3%), students (19.0%), artists (16.7%), people in business (4.8%), and farmers (4.8%) being among the most well-represented. The survey revealed that 76.2% of participants watched the entire FXS-documentary, while 23.8% were unable to do so. The participants provided qualitative feedback on the FXS-documentary, highlighting three main themes: the documentary's ability to enhance genetic knowledge, the importance of community involvement, and its effectiveness as a counselling tool. Only 38.1% of participants expressed dissatisfaction with the documentary's language, background noise, and video quality. It's worth noting that non-English speaking viewers may face challenges accessing the FXS-documentary since it's only available in English. Nonetheless, the FXS-documentary received favorable reviews with an impressive rating of 4.4 stars. All participants were willing to share the FXS-documentary with their friends, family, and colleagues. This documentary on FXS can serve various purposes, including being a genetic counselling tool, lecture support material, and audiovisual material for community engagements. Hence documentaries could be used with personalized genetic counselling services to provide comprehensive support and guidance to affected individuals.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Session III

PB2384 Framing the conversation about prognoses for children with genetic neurodevelopmental conditions

Authors:

E. Turbitt[1], M. Bourne[1], A. Gokoolparsadh[1], A. McEwen[1], D. Amor[2]; [1]Univ. of Technology Sydney, Sydney, Australia, [2]Murdoch Children's Res. Inst., Parkville, Australia

Background: Conversations about prognosis for genetic neurodevelopmental conditions are becoming more frequent; however, there is a lack of evidence and guidance on how to approach these conversations.
Aim: We aimed to investigate what parents want to discuss about their children's prognosis with a focus on aspects such as expected intellectual functioning and autism features. We further aimed to investigate how parents receive prognostic information from clinicians, and online, and their preferences for these conversations.
Methods: This was a mixed-methods study, using qualitative interviews and an environmental scan of online information. The semi-structured interview data with parents were analysed using reflexive thematic analysis. We conducted a content analysis of online information.
Results: We interviewed 32 parents from across Australia. Parents had a child with a genetic neurodevelopmental condition, such as Fragile X syndrome (28%), 22q11.2 deletion syndrome (16%) or Angelman syndrome (16%). Parents found conversations about their child's prognosis stressful and emotional, with a preference to discuss their child's potential strengths as well as challenges. They reported that conversations about prognosis often focused on the child's potential deficits and that online information they encountered was similarly framed negatively. Our content analysis of online information confirmed parents accounts: 95.3% was coded as negative, while only 4.7% was positive/neutral.
Conclusion: Our data provide evidence of an over-emphasis of deficit-framed prognostic information about genetic neurodevelopmental conditions. The initial exposure to negative information may adversely affect parents' psychological well-being and expectations, which future research could address. Health professionals could consider strengths-based framing of prognostic information gained from current and emerging technologies when returning results to families.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Session I

PB2385 Generative methods for Pediatric Genetics Education.

Authors:

R. Waikel, T. Patel, A. Othman, S. Ledgister Hanchard, P. Hu, D. Dat, B. D. Solomon; Natl. Human Genome Res. Inst., Bethesda, MD

With the advancement of generative artificial intelligence (AI), there has been much effort to validate the effectiveness of these generated data (e.g., how well ChatGPT generates programming code). Analogous to these studies, to help address patient privacy concerns and the limited ancestral diversity of medical teaching images, we propose the use of generative AI images in medical education. Specifically, we sought to determine whether generative AI could help train pediatric residents to recognize two genetic conditions: Noonan and Kabuki syndromes. To generate fake images, publicly available images of real syndromic individuals were collected and used to finetune StyleGAN2-Ada. This finetuning used vector representations of the syndromic images. By varying these vector representations, we generated images of fake individuals transformed from unaffected to affected with the genetic condition of interest. Recruited pediatric residents provided opinions about facial diagnosis and classified 20 images following exposure to one of 4 possible educational interventions: text description of the genetic condition, text and real images, text and generative AI images, or text and generative AI transformation image strips, displaying a series of 5 images from unaffected to affected. When provided a text description of the facial dysmorphology, pediatric residents had a higher baseline average accuracy when classifying Noonan images (65.3%) compared to Kabuki images (48.2%), suggesting more familiarity with Noonan syndrome. The addition of real images significantly raised the accuracy for Noonan to 74.3% and Kabuki to 60.3%. Generative images also improved accuracy, particularly for Kabuki syndrome, with an accuracy of 57.0% (single image) and 59.6% (transformation image strip). Most participants reported educational images regardless of type to be helpful in classification, although generative images were rated as less helpful than real images. This difference could reflect an inherent bias against AI methodologies, as performance between groups was more similar than reported usefulness. We conclude that generative AI images can serve as an educational tool particularly for rarer conditions, such as Kabuki syndrome.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Session II

PB2386 Genetic knowledge and acceptability of genetic risk communication among hematologic malignancy survivors

Authors:

N. Sharafeldin, A. Lively, A. Grimm, N. Balas, L. Francisco, L. Hageman, S. Bhatia; Univ Alabama at Birmingham, Birmingham, AL

Background: Gene-based personalized interventions carry the potential to be more effective in inducing changes in outcome. Hematologic malignancy (HM) patients are at increased risk of adverse cognitive outcomes post-treatment. Clear evidence, however, is lacking on patient genetics knowledge, their receptiveness to genetic risk communication, and its influence on deciding to participate in interventions that mitigate adverse cognitive outcomes. Methods: HM patients enrolled in the Cognitive Training and Attitudes towards Genetics study completed a self-administered survey post-treatment. Genetic knowledge was measured using a previously published structured survey measuring factual and clinical genetic knowledge and scores were compared to historical cohorts. Attitudes towards risk communication were measured using a hypothetical scenario that randomized patients to having low or high genetic risk for cognitive impairment. Logistic regression was used to assess associations of sociodemographic variables with knowledge scores and measure likelihood of intervention uptake based on genetic risk information.
Results: A total of 133 HM patients completed the survey: median age 55y (23-74), 60% females, 73.7% non-Hispanic Whites (NHW) and 21% African American, 47.4% had a college degree or higher, and 20.3% had a household income of 100K or more. Average total correct score was 61.1%, 54.8% for factual and 76.5% for clinical knowledge. Scores were statistically lower among HM survivors compared to the general population ($p<0.001$) but higher than other chronic disease patient populations ($p=0.0013$). Older age was associated with lower total scores (OR=0.36, 95%CI: 0.15, 0.86, $p=0.021$), while higher education level was associated with higher factual (OR=3.79, 95%CI: 1.61, 8.88, $p=0.002$) and race other than NHW with lower clinical scores (OR=0.34, 95%CI: 0.13, 0.87, $p=0.024$). A total of 57% considered genetic testing for post-treatment outcomes somewhat or extremely important, 88% felt confident they understood the questions, and 76.5% found it not hard to use genetic information to make decisions on using interventions. Perceived high genetic risk for impairment was significantly associated with higher likelihood of affecting patients' decision to use an intervention (OR=2.36, 95%CI: 1.13, 4.93, $p=0.022$). Conclusion: Discrepancies in knowledge scores among hematologic cancer survivors present an opportunity to improve patient knowledge levels. These findings show patients are amenable to genetic risk communication and support the potential utility of incorporating genetics in risk-based management of cancer survivors.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Session III

PB2387 Genetic return of results in a population-based cohort: Enrollment experience from the PopSeq Project

Authors:

K. Christensen[1], S. Walmsley[2,3,4], C. Blout Zawatsky[2,3,4], G. Splansky[5,6], R. Bowles[5,6], L. Clayson[5,6], M. Valentino[5,6], K. Swenson[5], M. Lebo[7,4,8], K. Machini[7], H. Zouk[7,8], M. Hickingbotham[9], A. P. Carson[10], N. Heard-Costa[5,6], R. C. Green[2,3,4,8]; [1]Dept. of Population Med., Harvard Med. Sch. and Harvard Pilgrim Hlth.Care Inst., Boston, MA, [2]Dept. of Med., Mass Gen. Brigham, Boston, MA, [3]Ariadne Labs, Boston, MA, [4]Broad Inst., Cambridge, MA, [5]Boston Univ. Chobanian & Avedisian Sch. of Med., Boston, MA, [6]Framingham Heart Study, Framingham, MA, [7]Lab. for Molecular Med., Mass Gen. Brigham, Cambridge, MA, [8]Harvard Med. Sch., Boston, MA, [9]Dept. of Population Med., Harvard Pilgrim Hlth.Care Inst., Boston, MA, [10]Dept. of Med., Univ. of Mississippi Med. Ctr., Jackson, MS

Background: Genetic testing is often conducted in large-scale human studies; however, there is little empirical evidence around genetic return of results (gRoR) within the context of large-scale population studies. Most published data has been collected through biobanks affiliated with academic medical centers. Here we present data collected from the PopSeq Project: a prospective study of gRoR in population-based cohorts.
Methods: Framingham Heart Study (FHS) participants were sequenced as part of the NHLBI Trans-Omics for Precision Medicine (TOPMed) program and consented to re-contact for actionable results. Results were returnable if a pathogenic or likely pathogenic variant was identified in a gene included in the American College of Medical Genetic and Genomics v2.0 secondary findings list. Living participants with variants previously unknown to the study were contacted by a FHS affiliated geneticist (MD) by phone. If a participant elected to move forward, the MD facilitated CLIA confirmation and a clinical disclosure. If participants declined the reason was recorded.
Results: Of 4,197 FHS participants sequenced, 109 (2.6%) were identified to have returnable results, of which 42 (38.5%) were deceased. Living individuals with returnable results included 29 of 67 (43.3%) participants from the Original and Offspring cohorts, recruited in 1948 and 1971, respectively; and 38 of 42 (90.5%) participants from the Omni and Generation 3 cohorts, recruited in 1994 and 2002, respectively. Of the 67 living participants with returnable results, the study team was able to reach 39 (58.2%) for notification, of which 34 (87.2%) agreed to gRoR. Thus far, results have been clinically confirmed for 20 (58.8%) individuals. It took a median of 3 (1-8) contact attempts per participant to reach them for pre-test counseling. The reasons for decline were already being aware of their genetic result, no interest in the study, did not think testing was relevant due to already being affected, and wanting to prioritize non-medical issues.
Discussion: The PopSeq Project is the first study to conduct gRoR research in a population-based cohort. Analyses found that many FHS participants with actionable results had died by the time variants were identified. Of those still living, the majority were interested in clinical confirmation after being notified, and reported decliner reasons were similar to other published studies. Similar efforts are underway within the all-African American Jackson Heart Study. These results can help inform the development of protocols for population-based studies interested in gROR.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Session I

PB2388 Genomics return of results to participants in the All of Us Research Program

Authors:

P. Jain[1], M. Begale[1], M. Bryant[2]; [1]Vibrent Hlth., Fairfax, VA, [2]Vibrent Hlth., Mt. Pleasant, SC

"Participants have access to their information" is a core value of the All of Us Research Program (AoURP). "Retention", broadly defined as the number of enrollees actively participating in program activities, has emerged as a crucial program metric. Genetic Return of Results (gRoR) is a central tool to achieve both of these programmatic priorities. gRoR has two core components: health-related genetic results and engagement (non-health related) genetic results. This session will describe the program's strategy for returning both engagement and health-related genetic results through the Genomics Platform (GP). The main objective of returning genetic results is to provide value to participants by giving them access to insightful information based on their genetics. The return of genetic results is optional and affirmatively consented participants choose which types of results they want to receive. By providing both health-related and non-health-related (or engagement) genetic results as options, the AoURP enables participants to choose which results they find most interesting, valuable, or engaging.

The GP is an AoURP-branded, web-based user experience that houses the genetics results modules. Participants will be able to navigate from their participant portal to the GP, and from there, choose which types of results they would like to see within two categories: health-related results or genetic ancestry and trait results.

Aims of the program we will present: Aim 1: Develop and deploy the Genomics Platform (GP), a secure information and content management platform that will deliver genetic results to All of Us Research Program (AoURP) participants through a web-based user experience.

Aim 2: Make engagement and health-related genetic results available to eligible participants

Aim 3: Ongoing monitoring of the GP, including the Genetics Engagement Module (GEM) and Health-Related Genomics Module (HGM) return of results.

Additionally, the GEM content and informing loops have been subjected to comprehension testing, to assess whether participants understand key concepts or important takeaways, from the materials. Special attention was given to the genetic ancestry content.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Session II

PB2389 *Global genomic pathogen surveillance: online training and pandemic preparedness.*

Authors:

l. jackson[1,2], COG-Train; [1]Univ. of Exeter, Exeter, United Kingdom, [2]Univ. of Cambridge, Cambridge, United Kingdom

Training in genomics and bioinformatics is at peak demand due to its increased use in public health applications such as disease surveillance. The pandemic forced the development of innovative methods to continue providing training while generating more affordable and equitable access to scientists who are in great need for genomics skills development. COG-Train, a partnership between Wellcome Connecting Science and COG-UK, pioneered a truly global delivery of a SARS-Cov-2 genomics training programme focused on increasing capacity for application of pathogen genomic surveillance in emergency response at global scale. The programme brought together scientists based in Africa, Asia and Latin America to develop a series of free online-based educational initiatives covering topics from genomic sequencing techniques to bioinformatic analysis and application of genomics in public health. Our three main training courses (MOOCs, Distributed Classrooms and Virtual Courses) have provided training to over 13,000 participants from &gt150 countries, predominantly from our three target regions in the global south (Africa, Asia and Latin America). The majority of participants were early-mid career researchers working on pathogen genomics and microbiology, although there were also numerous individuals working in public health and in more senior roles. Week-long intensive courses in viral bioinformatics, delivered virtually in Asia and Latin America provided more advanced bioinformatics training, enabling 60 early career scientists to develop their own SARS-CoV-2 analysis pipelines and to establish best practices on how to share their skills with others. A blended learning format and distributed classroom model spanning Africa, Asia and Latin America, supported scientists with intermediate experience to learn concurrently on how to manage a viral genome sequencing pipeline, extract and interpret results. This mixed format facilitated engagement from scientists regardless of their location, experience and resources. Such a global approach to pathogen genomics and bioinformatics is key to building capacity and regional networks which may be crucial for future epidemic and pandemic response to infectious diseases. Lessons from this program can inform the teaching of human genomics and bioinformatics in the Global South.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Session III

PB2390 How do patients with hereditary cancer syndromes navigate the healthcare system? A qualitative comparative study

Authors:

C. Butkowsky[1,2], J. Sam[1], E. Reble[1], M. Clausen[1], S. Rajeziesfahani[3], B. Sparkes[3], R. Gopalakrishnan[1], V. Aguda[1], M. Aronson[4], J. C. Carroll[2], D. Bishop[5], L. Dawson[3], A. Eisen[6], T. Graham[6], J. Green[3], C. Mighton[1,2], J. Pauling[7], C. Pavao[8], P. Pechlivanoglou[2,9], C. Remocker[8], S. Savas[3], S. Sun[10], T. Tiano[7], A. Tilley[5], K. Thorpe[1,2], K. Schrader[10], H. Etchegary[3], Y. Bombard[1,2]; [1]St. Michael's Hosp., Toronto, ON, Canada, [2]Univ. of Toronto, Toronto, ON, Canada, [3]Mem. Univ. of Newfoundland, St. John's, NL, Canada, [4]Zane Cohen Ctr., Sinai Hlth.System, Toronot, ON, Canada, [5]Patient Partner, St. John's, NL, Canada, [6]Sunnybrook Hlth.Sci. Ctr., Toronto, ON, Canada, [7]Patient Partner, Toronto, ON, Canada, [8]Patient Partner, Vancouver, BC, Canada, [9]The Hosp. for Sick Children, Toronto, ON, Canada, [10]BC Cancer, Vancouver, BC, Canada

Background: Hereditary cancer syndromes (HCS) account for 5-10% of all cancers. HCSs such as hereditary breast and ovarian cancer syndrome (HBOC) or Lynch syndrome (LS) can increase one's lifetime risk of cancer by over 80%. HCS patients therefore require lifelong follow-up care, including screening and appointments with a wide range of specialists. However, it is unknown what patients' care experiences are after receiving a molecular diagnosis of HCS, a major factor influencing clinical utility of genetic testing.
Aim: To understand the care experiences and needs of HCS patients following receipt of a molecular diagnosis of HCS.
Methods: HCS patients who received a positive genetic test result for HBOC or LS were purposely sampled from clinics in Ontario (ON), Newfoundland and Labrador (NL), and British Columbia (BC), reflecting variation across genetic testing services, screening and follow-up programs. Semi-structured interviews were completed with interpretive description used for analysis.
Results: Qualitative interviews were conducted with 73 patients (51 females, 21 males, 1 gender-diverse; age range 25-80 yrs) diagnosed with HBOC (n= 39) or LS (n= 34). Several key themes emerged including navigation, advocacy, and access. HCS patients expressed difficulties in navigating recommended follow-up care, and often mentioned a lack of guidance and knowledge from their healthcare professionals. Several patients highlighted the need for adequate communication about screening practices following risk-reducing surgeries and the importance of trust in care. Patients often had to self-advocate for referrals and screening appointments and noted differences in access to these services across jurisdictions. Importantly, access to genetic services, specialists and eligibility for prevention programs were described as sometimes limited for HCS patients. Participants commented that their circle of care (i.e., number and diversity of specialists involved in care) was likely smaller than those with a previous cancer diagnosis and noted impacts on access to support services.
Conclusions: This is the first study to compare HCS patients' experiences with care after receipt of genetic testing, a determinant clinical utility. We found that HCS patients face numerous healthcare challenges. Receiving adequate guidance from healthcare professionals, getting timely access to screening, professional and informational supports and trust in care were most widely discussed among participants. Identifying the needs and challenges for patients with HCS can optimize care experiences and ultimately improve patient outcomes and the value of genetic testing.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Session I

PB2391 How to conduct equitable genetics research to include underrepresented populations: A systematic review of best practices.

Authors:

V. Aguda[1], S. Grewal[1], K. Cheung[2], V. Jha[3], E. Reble[1], M. Clausen[1], D. Lightfoot[1], A. Lofters[4], N. Persaud[1], M. Aronson[5], A. Pollett[5], H. Feilotter[6], J. Lerner-Ellis[5], Y. Bombard[3,1]; [1]Unity Hlth.Toronto, Toronto, ON, Canada, [2]Boston Univ., Boston, MA, [3]Univ. of Toronto, Toronto, ON, Canada, [4]Women's Coll. Hosp., Toronto, ON, Canada, [5]Mount Sinai Hosp., Toronto, ON, Canada, [6]Queen's Univ., Kingston, ON, Canada

Introduction: The genetics community at large has recognized that current data sources and analytical approaches are inherently biased toward individuals of European ancestry. Underrepresentation of genomes from diverse ancestries results in uncertainty about the interpretation of benign versus disease-causing genetic variation leading to a higher rate of inconclusive test results in those populations and risk prediction models that do not transfer across genetic ancestries. Without a genetic diagnosis, access to personalized treatment and management options, as well as cascade testing for relatives, is restricted, often in populations who are already under-referred to genetic services. As genetic information is increasingly translated into clinical care, promoting equitable practices is paramount to prevent genetic technologies from further exacerbating existing health disparities. While the lack of diversity in genetic research has been well characterized, there are limited guidelines for researchers on best practices for conducting equitable genetic research. In this study, we aim to systematically review and synthesize the literature on best practices for conducting equitable genetic research. **Methods:** A systematic review is being conducted. The review has been registered in PROSPERO (CRD42023384475). Medline, EMBASE, and Cochrane Databases were searched for articles that described strategies, guidelines, contextual factors, and best practices for conducting equitable research in genetics. Grey literature was also searched. Data extraction is being conducted by two reviewers. Quality assessment is being conducted using established critical appraisal tools from the Joanna Briggs Institute. Mixed methods are being used for a descriptive synthesis. **Results:** The searches returned 5366 articles. Title and abstract screening are ongoing. To date, 286 studies have proceeded to full-text screening. Preliminary themes from articles in full-text screening include increasing awareness and access to research opportunities, aligning research goals and questions with the priorities of communities, providing culturally-sensitive research materials, and the importance of co-design and collaboration rather than tokenistic participation. Full-text screening, data extraction, critical appraisal, and synthesis are expected to be completed by October 2023. **Conclusions:** This evidence synthesis is the foundation for the development of best practices for how genetic research with diverse patient populations can be done in a manner that is culturally safe, equitable, and ultimately benefits all involved.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Session II

PB2392 † Identifying patients at-risk for breast/ovarian cancer within Kaiser Permanente Mid-Atlantic States - the evaluation of genetics referrals and testing from a FHS-7 best practice advisory alert

Authors:

C. Jonas, P. Suwannarat, S. Basra, E. Watson, C. M. Boyd, A. Deneal, D. Canedo, S. Gustafson, K. Hsieh, S. Kim; Mid-Atlantic Permanente Med. Group, Rockville, MD

Background: Ovarian cancer patients are often diagnosed at a late stage, and the disease is associated with high mortality. According to the CDC, 10% of ovarian cancer cases are due to a hereditary genetic variant. In 2021, Kaiser Permanente Mid-Atlantic States (KPMAS) created an electronic health record based Best Practice Advisory (BPA) alert to identify patients who may be at higher risk for breast/ovarian cancer. This BPA displays based on a documented family history of certain cancers. Patients scoring more than 1 on the Family History Screening 7 (FHS-7) are offered a Genetics referral for genetic counseling. Results: During Sept. 1, 2021-Sept. 30, 2022, there were 5,466 OB/GYN visits with the BPA acknowledged, resulting in 2,769 (51%) Genetics referrals and 1,480 (53%) completed virtual genetic counseling visits. Genetic testing was ordered in only 24% of genetic counseling visits (n=348), but testing was completed 86% of the time (n=301/348; 86%)—indicating that if patients in this cascade are offered genetic testing, a large majority complete it. These results prompted the study team to examine the other 76% (N=1,131) of visits in which genetic testing orders were not placed. Results showed: 50% (n=561) of patients did not meet NCCN criteria for testing, 25% (n=278) met NCCN criteria but were told to encourage their affected relative to test first, 24% (n=275) met NCCN criteria but declined testing, and 1% (n=17) had already tested. Among the group told to encourage their affected relative to test, none had evidence of returning the relative's genetic testing results. Conclusions: Our results show a key drop off at the point of ordering genetic testing. Specifically, patients who met testing criteria were instructed to encourage an affected relative to test first and were not offered testing themselves (n=278; 19% of 1,480 total completed visits). As a result, none of these eligible patients completed genetic testing. While testing an affected family member is preferred, it is not always possible. Failing to test patients who meet criteria increases the chance that at-risk individuals are not identified and is a missed opportunity to offer genetic information. Recommending contacting affected family members that might be more appropriate for initial genetic testing is a long-accepted practice that may need to be revisited.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Session III

PB2393 Impact of SeqFirst in hospitalized neonates

Authors:

T. Wenger[1], K. MacDuffie[1], A. Keefe[1], M. Sikes[1], L. Kruidenier[2], O. Sommers[1], H. Gildersleeve[1], K. Buckingham[1], J. Chong[1], D. Veenstra[3], K. Dipple[4], D. Miller[1], D. Copenheaver[5], J. Juusola[6], K. Retterer[7], K. McWalter[8], A. Snook[8], P. Kruszka[8], J. H. Yu[1], M. Bamshad[1]; [1]Univ. of Washington, Seattle, WA, [2]Seattle Children's Hosp., Seattle, WA, [3]Univ of Washington, Seattle, WA, [4]Seattle Children s Hosp. and Univ. of Wash, Seattle, WA, [5]Gene Dx, Gaithersburg, MD, [6]GeneDx, Clarksburg, MD, [7]Geisinger, Danville, PA, [8]GeneDx, Bethesda, MD

The SeqFirst neo project aims to improve access equity to genomic testing of hospitalized neonates by offering rapid whole genome sequencing (rWGS) to infants not fully explained by prematurity, infection or trauma. Our group previously reported that enrolled infants (n=126) had a diagnostic yield of 50% at the time of rWGS. In order to better understand the impact of the SeqFirst approach versus conventional care, we reviewed the electronic health record for the first 12 months of life for 1) infants enrolled in SeqFirst (n=126) and 2) controls - consisting of all infants admitted to the NICU during the study period who were medically eligible for SeqFirst but instead received conventional care (n=113). Additionally, we explored the impact on families via parental questionnaires at enrollment, after rWGS, 6 and 12 months after enrollment. (Response rates per questionnaire 74.8-86.4%) Semi-structured 1:1 interviews were also conducted with neonatologists who cared for infants enrolled in SeqFirst. Infants in the control group who received conventional phenotype-driven care had a diagnostic rate of 8.8% in the NICU. After 12 months of follow up, only one additional genetic diagnosis had been made in the outpatient setting (9.7%). Precise genetic diagnosis in living infants led to similar rate of change in management for infants in SeqFirst group (89%) and those who received conventional care (87.5%). Among infants who died during the study period, those who were enrolled in SeqFirst were more likely to have a precise genetic diagnosis compared to those who received conventional care (67% vs. 22%). Parental questionnaires suggested low decisional regret and a positive impact on families for infants who received a precise genetic diagnosis as well as those who had nondiagnostic rWGS. Semistructured interviews with eight neonatologists who had patients enrolled in SeqFirst described the positive impacts they observed compared to conventional care, including expanded testing and quick access to results which helped guide medical management. Neonatologists additionally reported that the experience with SeqFirst had lingering impacts on their practice, broadening their awareness of the value of genetic testing and boosting their willingness

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Session I

PB2394 Improving access to exome sequencing in a medically underserved population through the Texome Project: A summary of the first 60 cases.

Authors:

B. Vuocolo[1,2], R. German[1,2], C. Bacino[1,3,4], C. Murali[3,1], S. Lalani[1,3], S. Baskin[5], E. Roeder[5], C. Schmid[5], S. McLean[5], R. Littlejohn[5], O. Juarez[5], M. Stuebben[5], M. Nutter[5], B. Mitchell[5], D. El Achi[5], L. Vossaert[6,4], N. Owen[1,4], C. M. Eng[1,4], P. Liu[1,4], Z. Liu[1,2], D. Mao[1,2], S. Yamamoto[1,2], S. Kim[1,2], S. Pasupuleti[1,2], H. Bellen[1,2,7], M. Wangler[1,2]; [1]Baylor Coll. of Med., Houston, TX, [2]Jan and Dan Duncan Neurological Res. Inst., Houston, TX, [3]Texas Children's Hosp., Houston, TX, [4]Baylor Genetics Lab., Houston, TX, [5]Baylor Coll. of Med., San Antonio, TX, [6]Baylor Coll. Med., Houston, TX, [7]Howard Hughes Med. Inst., Houston, TX

Exome sequencing has been proven to be a useful clinical tool in ending diagnostic odysseys, but health insurance and other systemic barriers prohibit individuals in underserved communities from equitably accessing this genetic testing. The Texome project is an NIH-funded research study that reduces barriers to genetic evaluation by offering exome sequencing to patients with undiagnosed, rare diseases who have financial barriers to obtaining exome sequencing clinically. Participants receive a genetic evaluation, CAP/CLIA exome sequencing, and return of results free of charge. To date, there have been 132 applications reviewed by the Texome clinical team, and 104 have been accepted into the study, with an acceptance rate of 78.8%. Diverse regions of Texas are represented in our cohort, and a majority of patients are publicly insured (53.1%) and identify as Hispanic/Latino (68.1%). To date, 60 individuals have undergone exome sequencing, with an etiologic diagnosis confirmed in 18 (30%) of cases; 17 were solved by proband-only ES with Sanger confirmation of available parents, and one additional case was solved after reflexing to trio ES and discovering a recently-described de novo RHOB variant. Of these solved cases, 5 (27.8%) are confirmed or presumed to be inherited (3 autosomal recessive and 2 dominant). There is an even distribution of solved cases between the adult and pediatric patients. Variants in ITGA2B, RHOB, ZIC2, CTCF, ERCC4, B4GALT7, and SHANK3 were identified in adult patients of Texome, ending decades-long diagnostic odysseys and contributing to the genotypic and phenotypic understanding of these disorders. For cases that remain unsolved, we are utilizing advanced bioinformatics pipelines to identify novel variants or genes that may cause disease. Findings from the Texome project support the need for expanding access to genetic testing and services in underserved populations and highlight the utility of proband exome sequencing with Sanger confirmation as a first-tier test in medically underserved populations.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Session II

PB2395 Increased Rate of Return of Secondary Findings: Impact of the ACMG v3.1 Gene Lists.

Authors:

M. Kamen, M. Similuk, S. Byers, J. Yan, B. Seifert, R. Ghosh, C. Jodarski, R. Gore, K. Lewis, N. Reynolds-Lallement, B. Beers, E. Damskey, H. Orellena, J. Johnston, L. Biesecker, M. Walkiewicz; NIH, Bethesda, MD

Background: In June of 2022, the American College of Medical Genetics and Genomics (ACMG) released the ACMG Secondary Findings (SF) v3.1 list. The SF v3.1 policy statement announced 5 new genes to be included in the minimum SF gene list. This was a minor update to the 2021 ACMG SF v3.0 list. There are limited empirical estimates of the extent to which these additions will increase the SF molecular diagnostic rate, particularly considering the relatively high allele frequency (1.0%-2.5%) of the pathogenic c.424G>A *TTR* variant among individuals of West African ancestry. We sought to address this gap by reviewing the study cohort recruited to the National Institute of Allergy and Infectious Diseases (NIAID) Centralized Sequencing Program (CSP) to determine the impact on the rate of return for the ACMG SF v3.1 vs v3.0.
Methods: We analyzed the genomes of a cohort of families with heterogenous clinical phenotypes enrolled in the CSP. We applied ACMG's Standards and Guidelines for the Interpretation of Sequence Variants to evaluate pathogenicity. We returned pathogenic and likely pathogenic variants to study participants per SF guidelines. Specifically, we looked at all coding variants with a minor allele frequency of ≤1%. Notable exceptions were variants with autosomal recessive modes of inheritance in *BTD* and *HFE* with a minor allele frequency extended to 5%, the pathogenic *GAA* variant, NM_000152.5 c.-32-13T>G as well as the pathogenic *TTR* variant c.424G>A.
Results: We applied the 2021 SF v3.0 guidelines to the genomic analysis of 1,091 families and detected 35 total SF (3.2% SF rate; 95% CI [2.2%,4.4%]). Subsequently, we applied the 2022 SF v3.1 updated guidelines to the analysis of 748 additional families and detected 30 total SF (4.0% SF rate; 95% CI 2.7% to 5.6%). Among the 5 new genes included in the SF v3.1 update, we identified the high frequency TTR c.424G>A variant in 9 families, (1.2%; 95% CI [0.55%,2.27%]) and this made up all new SF not reflected in the v3.0 list. While SF v3.0 vs v3.1 guidelines increased the rate of SF return from approximately 3.2% to 4.0%, this difference was not statistically significant (95% CI [-0.37%,0.56%], Chi-squared = 0.008, P = 0.93).
Discussion: These data suggest providers may expect an increase in the SF rate of return, largely driven by the frequency of the pathogenic c.424G>A *TTR* variant. The associated hereditary transthyretin amyloidosis has reduced penetrance and is mitigated by FDA-approved treatments, making it an attractive target for opportunistic screening and personalized treatment. The small sample size limits further analysis but provides preliminary evidence to support further characterization of SF yield and outcomes.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Session III

PB2396 Insurance Record-Based Analysis of the Burden of Disease in Patients with Galactosemia in the US

Authors:

D. Gallo[1], N. Longo[2], J. L. Fridovich-Keil[3], R. Chang[4], J. B. Spencer[5], B. Moy[1], A. G. Ellis[6]; [1]Jaguar Gene Therapy, LLC, Lake Forest, IL, [2]Dept. of Pediatrics, Univ. of Utah Spencer Fox Eccles Sch. of Med., Salt Lake City, UT, [3]Dept. of Human Genetics, Emory Univ. Sch. of Med., Atlanta, GA, [4]Children's Hosp. of Orange County, Div. of Metabolic Disorders, Orange, CA, [5]Dept. of Gynecology and Obstetrics, Emory Univ. Sch. of Med., Atlanta, GA, [6]Stratevi, LLC, Boston, MA

Objective: Classic galactosemia is a rare genetic disease that can cause severe lifelong complications. Data on the lifetime clinical trajectory of disease are limited. This study helps characterize the real-world burden of disease among US galactosemia patients. **Methods:** Data from MarketScan (January 2006 - September 2022) were analyzed. Patients were identified using International Classification of Diseases, 9th Revision and 10th Revision codes for galactosemia (271.1 and E74.21, respectively). Pediatric and adult galactosemia patients were matched with up to 2 individuals from the general population (controls) based on year of birth, sex, region, insurance type, and years of continuous enrollment. The prevalence (n, %) of 17 conditions among the galactosemia patients was compared with demographically matched controls. The list of conditions was generated based on a prior analysis of a US claims database (Symphony Health) and input from clinical experts. The prevalence of conditions by age groups and by sex was examined. **Results:** 1,957 galactosemia patients and 3,431 matched controls were included. The average age was 23, and 51% of patients were male. The prevalence of all conditions was greater in the galactosemia cohort compared with matched controls and was significantly higher for 16 conditions. Of note, developmental delay, bone health disorders, neurological complications, and intellectual disabilities were each at least twice as high for the galactosemia cohort vs. controls. Furthermore, conditions not previously associated with galactosemia were observed to be higher in the galactosemia cohort, including liver illness. The prevalence of specific conditions differed between younger and older age cohorts, indicating different manifestations of disease across ages. For example, dietary counseling and complications and liver illness were common among people <1 year of age and significantly higher for patients with galactosemia. In people 40-69 years of age, blood related disorders, kidney related disorders, liver illness, malaise and fatigue, and neurological complications were common and significantly higher in the galactosemia cohort. Among females, larger differences in gonadal complications and bone health were observed starting in adolescence and into adulthood. **Conclusion:** The prevalence of 16 selected comorbid conditions is significantly higher among patients with galactosemia than for a general population of matched controls. The prevalence of conditions varied across age groups, providing insights into the lifetime burden of disease and raising awareness of previously unexpected complications in adults.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Session I

PB2397 Measuring and Maximizing Outcomes in Population Genomic Screening: a High-Touch Recontact Pilot

Authors:

K. Tangney, D. Erwin, D. Brockman, S. Zarkesh Esfahani, N. M. Johnson, J. Lee, A. Khan, A. Collins, A. Nguyen, R. Garcia, C. L. Neben, S. Topper; Color Hlth., Burlingame, CA

Population genomic screening can provide crucial health information to millions at risk for hereditary conditions, but its impact depends on follow-up care utilization for which there is limited longitudinal data. At Color Health, we provide population screening and genetic counseling for common hereditary conditions through employee benefits programs, research studies, and healthcare providers. To assess how people navigate clinical recommendations after receiving genetic results and counseling, we recontacted 200 Color Health participants by phone who received a positive genetic result 1-3 years prior for familial hypercholesterolemia (FH), hereditary breast and ovarian cancer (HBOC), or Lynch syndrome (LS). We collected participant-reported actions since receiving results, and connected them to local resources or an in-house care advocate if further care navigation was indicated or requested.

Of 200 selected participants, 37% (n=74) were re-contacted. The average age of participants was 50 years (range 22-84), 47% (n=35) self-identified as female, 70% (n=52) were non-Hispanic White, and all were English-speaking. A total of 34 participants (46%) had a HBOC pathogenic variant (PV), 27 (36%) had a LS PV, and 13 (18%) had a FH PV. Of 11 HBOC females eligible for screening or prevention, 36% (n=4) reported initiating breast MRI, 55% (n=6) reported risk-reducing mastectomy, and 55% (n=6) reported bilateral salpingo-oophorectomy. Of 13 eligible HBOC males, 54% (n=7) reported initiating PSA, and 46% (n=6) initiated male breast cancer screening. Initiation of screening and prevention was higher with the 27 LS participants - 70% (n=19) initiated colonoscopy screening and 47% (n=7 females) reported risk reducing hysterectomy - but lower with the 13 FH participants - 31% (n=4) initiated or altered their lipid treatment.

Within two weeks of initial receipt, many participants shared results with a provider (65%; n=49) and at-risk relatives (73%; n=54). During recontact, 3 additional participants chose to share results. A total of 10 participants sought care navigation: 7 were connected to local resources by genetic counselors and 3 had help navigating health insurance or connection to a local primary care provider by care advocates.

This pilot showed that many population screening participants (57%; n=42) reported initiating a recommended screening, prevention or treatment after receiving a high-risk genetic result and post-test counseling. Accessible population screening combined with genetic counseling and care advocacy can be effective in identifying and empowering high-risk individuals to take appropriate next steps in their healthcare.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Session II

PB2398 Medically actionable findings from the Texome Project: Implementation of genomic medicine in underserved groups in Texas.

Authors:

R. German[1,2], B. Vuocolo[1,2], C. Bacino[1,3,4], C. N. Murali[3,1], S. Lalani[1,3], S. Baskin[5,1], E. Roeder[5,1], C. Schmid[5,1], S. McLean[5,1], R. Littlejohn[5,1], O. Juarez[5,1], M. Nutter[5,1], D. El Achi[5,1], B. Mitchell[5,1], M. Stuebben[5,1], A. Grace[1,5], L. Vossaert[1,4], N. Owen[1,4], C. M. Eng[1,4], P. Liu[1,4], Z. Liu[1,2], D. Mao[1,2], S. Kim[1,2], S. Pasupuleti[1,2], S. Yamamoto[1,6,7,2], H. Bellen[1,2,7,8], M. Wangler[1,2,7]; [1]Dept. of Molecular and Human Genetics, Baylor Coll. of Med., Houston, TX, [2]Jan and Dan Duncan Neurological Res. Inst., Texas Children's Hosp., Houston, TX, [3]Texas Children's Hosp., Houston, TX, [4]Baylor Genetics Lab., Houston, TX, [5]Dept. of Pediatrics, Baylor Coll. of Med., San Antonio, TX, [6]Dept. of NeuroSci., Baylor Coll. of Med., Houston, TX, [7]Program in Dev.al Biology, Baylor Coll. of Med., Houston, TX, [8]Howard Hughes Med. Inst., Houston, TX

The Texome Project is a genomic medicine implementation program for communities that lack access to genetics services in Texas. The Texome Project provides whole exome sequencing (WES), AI-optimized bioinformatics analysis, and model organisms to diagnose and study genetic disease in medically underserved groups. To date, 60 CAP/CLIA WES results have been disclosed to participants since beginning enrollment in December 2021. We present 5 cases where WES led to a change in medical management that would not have been indicated without access to genomic medicine. Case 1) A 7-year-old male was evaluated for global developmental delay, hypotonia, strabismus, atrial septal defect, and seizures. WES found a pathogenic variant in *ATP1A3* associated with alternating hemiplegia of childhood 2. The patient was started on flunarizine, a calcium channel blocker, and has not reported any further hemiplegic episodes. Case 2) A 40-year-old female with bilateral carotid paragangliomas, intellectual disability, and dysmorphic features was enrolled in the Texome Project. A variant in *DNMT3A* was uncovered on WES. This result represents a possible diagnosis of Heyn-Sproul-Jackson syndrome and was linked to the patient's paragangliomas and resulted in a recommendation to begin annual thyroid exams to screen for recurrent malignancies. Case 3) A 7-year-old male with autism, developmental delay, and a family history of hypocalciuric hypercalcemia was enrolled in the Texome Project. WES reported a pathogenic variant in *CASR* indicating a molecular diagnosis of hypocalciuric hypercalcemia type I. This finding resulted in a referral to follow with nephrology for monitoring. Case 4) A 9-year-old female with intellectual disability, autism, mild dysmorphic features, and epilepsy. She received WES, which reported a pathogenic variant in *GNB1* associated with Intellectual developmental disorder 42. Given this finding, the patient was recommended to seek physical therapy, occupational therapy, and a sleep study to better manage their condition. Case 5) A 5-year-old female presented with spasticity, cerebral palsy, toe walking, and recurrent urinary tract infections. After enrollment into the Texome Project, WES uncovered a pathogenic variant in *ADAR* consistent with Aicardi-Goutières syndrome 6. Tofacitinib, a JAK inhibitor, was recommended and administered by the patient's rheumatologist. The family has reported a positive response to treatment, with lessening spasticity. These cases demonstrate the clinical utility of whole exome sequencing in medically underserved groups to end diagnostic odysseys and provide management recommendations.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Session III

PB2399 Narrative explorations of rare and undiagnosed genetic diseases: Elucidating lived experiences of the diagnostic odyssey.

Authors:

I. Simeone, C. Dvorak; Hayward Genetics Ctr., Tulane Univ. Sch. of Med., New Orleans, LA

Background: It is estimated that approximately 72% of rare diseases have a genetic etiology, and many patients facing such diseases are misdiagnosed or remain undiagnosed for several years. Previous studies have characterized the negative psychosocial, financial, and emotional burdens on these patients and their families, suggesting an unmet need for psychological support among this population.
Purpose: Considering the role of family dynamics in shaping the lived experience of these patients, there exists a paucity of scholarly work on the sociological dimensions of rare and undiagnosed genetic diseases. Patients facing rare, complex, and difficult-to-diagnose genetic diseases often undertake a protracted and unpredictable "diagnostic odyssey" in seeking answers. What happens when a genetic disease does not fit within any existing paradigms? How can patients (and their caregivers) facing ultra-rare genetic diseases look to narrative and storytelling as a source of support?
Methods: To address these questions, our study was anchored by close readings of two illness memoirs, *The Family Gene: A Mission to Turn My Deadly Inheritance into a Hopeful Future* by Joselin Linder and *This Boy We Made: A Memoir of Motherhood, Genetics, and Facing the Unknown* by Taylor Harris. We aimed to call attention to the unique voices of patients who may feel overlooked and burdened by the complexity of their medical histories and elusive nature of their genetics.
Results: Illness memoirs written by writers grappling with rare and undiagnosed genetic diseases reveal ubiquitous themes of incomprehensibility, medicalization of family and kinship, and liminality; figurative language such as metaphor and anthropomorphism allows for these writers to effectively adopt a "style" for their illness, better allowing them to grapple with their lived experiences. Beyond the utility of the illness memoir for patients to generate their own "language of illness," such narratives—particularly those that may enrich a reader's understanding of how best to support patients despite having limited/no answers—can be a foundational tool in the educational experience of medical students.
Conclusion: Existing at the margins of established disease taxonomies, rare and undiagnosed genetic diseases are defined foremost by the "inexpressibility topos" inherent in navigating an illness sphere for which there exists no precedent, language, or greater "illness community" to lean on. Future research may focus on characterizing the beneficial effects of a "narrative intervention" (writing, reading, or verbally telling stories) across all participants in the therapeutic relationship.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Session I

PB2400 Neurodevelopmental Comorbidities in Newborn Screening Conditions.

Authors:

Z. Talebizadeh, G. Tona, Y. Unnikumaran, A. Brower; American Coll. of Med. Genetics and Genomics, Bethesda, MD

Newborn screening (NBS) is a large-scale public health program in the US that screens 3.8 million newborns for up to 81 genetic conditions each year. Many of these conditions have comorbidities, including autism, neurodevelopmental disorders (NDD), and intellectual/developmental disabilities (IDD). These comorbidities can have a significant impact on health outcomes across the lifespan. Most screened conditions are inborn errors of metabolism (IEM). PKU, the first condition identified by NBS, is an inherited metabolic disorder that can cause developmental delays and IDD if not treated. The Newborn Screening Translational Research Network (NBSTRN) is a program funded by the National Institute of Child Health and Human Development since 2008. NBSTRN is charged with developing, maintaining, and enhancing tools, resources, and expertise supporting NBS research. One of the tasks led by NBSTRN is to provide direction for developing question/answer sets used in the Longitudinal Pediatric Data Resource (LPDR) to create consensus-based, and standardized common data elements (CDEs) for NBS conditions. There is a growing interest in the NBS community to assess neurodevelopmental trajectories through long-term follow-up studies. This could be streamlined by employing uniform CDEs. To address this unmet need, we conducted a landscape analysis. Our objective was to identify CDEs that can be used to study neurodevelopmental trajectories for NBS conditions having NDDs as comorbidity. We used a mixed methods approach (data mining, focus group, and literature reviews) to conduct a feasibility assessment, including the identification of needs and priorities. Data mining included examining health information from a completed, 10-year natural history study of IEMs, called Inborn Errors of Metabolism Collaborative (IBEMC). 44 IEM conditions are currently included in the Recommended Uniform Screening Panel. Deidentified data from the IBEMC projects are available for secondary data analysis in LPDR. The IBEMC data dictionary was reviewed to identify NDD-related CDEs. PubMed search and MedGen review were conducted to estimate the prevalence of NDD-related comorbidities in NBS conditions. Furthermore, existing CDE repositories were evaluated to identify a list of NDD-related CDEs. Branching logic was developed to assist with prioritizing the adoption of these CDEs in NBS research studies. Our findings can inform future efforts toward advancing research infrastructure for this established public health program. The renewed awareness of the risk of NDDs after a positive NBS and diagnosis could lead to improving treatment guidelines for mental health conditions.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Session II

PB2401 Novel applications of the *All of Us* Research Program platform in a course-based undergraduate research experience.

Authors:

K. McDougal; Towson Univ., Towson, MD

Course-based Undergraduate Research Experiences (CUREs) are opportunities for students to design and conduct authentic research projects within the framework of an academic course. CUREs expose a larger number of students to research than traditional research experiences and can be beneficial for individuals from populations underrepresented in research spaces. The *All of Us* Research Program is a large, unique data resource accessible to researchers at all stages of their careers through a centralized cloud-based platform. This creates an interesting and valuable opportunity to expose students to research experiences in biomedical genomics using expansive data from diverse populations. A CURE using the *All of Us* Researcher Workbench has been run in three different iterations. The first was a Registered Tier design using electronic health record, survey, wearable, and demographic data types. Then, once genomic data was released, a Controlled Tier design was implemented using limited genotype array data combined with the data types used in the previous Registered Tier version. Finally, the most recent iteration was a Controlled Tier design where students performed genome-wide association studies using whole genome sequencing data in concert with the electronic health record, survey, and demographic data types. In all iterations students worked in pairs to design a research proposal on the topic of their choice, performed the research in the *All of Us* Researcher Workbench, and reported their findings in a poster presentation. This CURE is an example for those at other institutions who wish to utilize the *All of Us* platform to enhance student exposure to authentic research experiences in biomedical genomics with a precision medicine focus.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Session III

PB2402 Offering return of results at the time of enrollment improves uptake of genomic screening among participants of a diverse biobank

Authors:

E. Soper[1], S. Suckiel[2], E. Kenny[1], N. S. Abul-Husn[3]; [1]Icahn Sch. of Med. at Mount Sinai, New York, NY, [2]Icahn Sch. of Med. at Mount Sinai, New York City, NY, [3]23andMe, Sunnyvale, CA

Background: Genomic screening for actionable variants can uncover disease risk, allowing intervention with the aim of improving health outcomes. Biobank-based return of results (RoR) offers an opportunity to implement genomic screening at scale. However, the ideal timing for offering RoR to participants remains unknown. We compared RoR uptake among participants of a diverse biobank in New York City who could consent to RoR at enrollment or at a later time.
Methods: The study population consisted of 127 adult participants of the Bio*Me* Biobank enrolled between 8/2/2016 - 2/7/2020 with an unconfirmed research finding (*TTR* V142I pathogenic variant). The Bio*Me* research protocol was amended 10/1/2018 to allow RoR. Participants enrolled after that date (N=56) could opt in/out of RoR at consent. Participants enrolled prior to this (N=71) were not offered RoR at consent, but could subsequently update their consent to opt in/out of RoR. We outreached to participants to collect a new sample for clinical confirmation of the research finding. Individuals who had consented to RoR at enrollment were informed of the presence of a research finding and asked to provide a new sample if they wished to proceed with clinical confirmation. Those enrolled prior to the protocol amendment could not be informed about a research finding, but were offered the opportunity to re-consent for RoR with concurrent sample collection for clinical confirmation. Using Fisher's exact test, we compared the uptake of RoR, defined as receiving a clinically confirmed result, among both groups overall, and among the subset of individuals from each group who we successfully recontacted.
Results: 52 of 56 participants who were consented for RoR at enrollment, and 70 of 71 who were not, were eligible for outreach (i.e., living and had not opted out of RoR). Overall, RoR uptake was 2.9x higher among those opting in to RoR at enrollment (33%) compared to those not previously consented for RoR (11%, p=0.006). In the subset of individuals we successfully recontacted, RoR uptake was 2.4x higher among those opting in to RoR at enrollment (44%) compared to those not previously consented for RoR (18%, p=0.016).
Conclusions: A significantly higher proportion of participants received results when offered the option of RoR at the time of enrollment into Bio*Me*. Crucially, we were able to inform participants of a potential genomic finding at outreach if they had already consented to receive this information. These findings highlight the importance of engaging participants about RoR at the time of biobank enrollment to maximize the reach of genomic screening.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Session I

PB2403 Participants and stakeholders views on feedback of genetics research findings of the H3Africa kidney disease research network, Ghana**.**

Authors:

G. Larbie Gafa[1], V. Boima[2,1], I. Rockson[1], S. T. Chatio[3], A. Ghansah[4], B. L. Salako[5], A. Ojo[6], P. Tindana[7], D. Adu[1,2]; [1]H3-Africa Kidney Disease Res. Network, UGMS, Accra, Ghana, [2]Dept. of Med. and Therapeutics, Univ. of Ghana Med. Sch., Coll. of Hlth.Sci., Univ. of Ghana, Accra, Ghana, [3]Navrongo Hlth.Res. Ctr., Ghana Hlth.Service, Navrongo, Ghana, [4]Noguchi Mem. Inst. for Med. Res., Univ. of Ghana, Legon, Ghana, [5]Coll. of Med., Univ. of Ibadan, Ibadan, Nigeria, [6]Kansas Univ. Med. Ctr., Kansas, KS, [7]5. Dept. of Hlth.Policy, Planning and Management, Sch. of Publ. Hlth., Coll. of Hlth.Sci., Univ. of Ghana, Legon, Ghana

Genetics and genomics research raise important ethical issues, particularly those related to obtaining valid consent, engaging with relevant communities in the implementation of research and feedback of research findings. There is evidence that variants of the Apolipoprotein L1 gene in Africans and people of African descent increase the risk of developing chronic kidney disease (CKD). A study carried out by the H3Africa Kidney Disease Research Network showed that 28.2% of the study participants carried 2 *APOL1* high-risk variants. It is also known that relatives of subjects with CKD may be at increased risk of developing CKD. Should participants and their relatives be informed about their risk of developing CKD? What findings should be given and who should be involved in the feedback process? The aim of this study was to seek the views of research participants, families and stakeholders on feedback of genetic research findings. This research employed an exploratory qualitative study approach for data collection. The study explored the views of genetic/genomic researchers, participants, families and members of research ethics committees on what count as good ethical practice in deciding what, who and how to return genetic research findings of the kidney disease research in Ghana through in-depth interviews, focus group discussions and deliberative workshops. Data was coded and structured with NVivo software and thematically analysed. There was a consensus that relevant individual and aggregate genetics results should be fed back to participants and communities. Most participants preferred to receive their personal results from a doctor or a research scientist. There is an ethical imperative to return validated clinically relevant individual genetic research results to the kidney disease research participants and families and aggregate results to communities. We recommend the H3Africa Kidney Disease Research Network to educate participants and families on the concept of risk variants, train Genetic Counsellors and explore innovative strategies to support the feedback process.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Session II

PB2404 Poor foundation in genetics and related disciplines a clog in the progress of bioinformatics and computational biology in Nigeria

Authors:

I. Uzoma[1], J. Iloekwe[2], I. A. Taiwo[3]; [1]Molecular-Haematology Lab., Dept. of Med. Lab. Sci., Univ. of Nigeria Nsukka, Enugu, Nigeria, [2]Sch. of Sci. and Technology Pan-Atlantic Univ., Lekki, Lagos, Nigeria, [3]Human Genetics Res. Group, Genetics Res. Unit, Dept. of Cell Biology and Genetics, Faculty of Sci., Univ. of Lagos, Akoka., Lagos, Nigeria

The advent of genomics is associated with a rapid rise in biological data needing the use of mathematics, statistics, and computer algorithms to handle and examine the huge biological information stored in databases. Despite the seemingly obvious relevance of bioinformatics and computational biology in the current age of genomics, growth in the field of computer-aided management and study of biological data has been sluggish in Nigeria. Some scientists have deliberated on the state of bioinformatics and computational biology in Nigeria; some have addressed possible factors militating against advancement and progress in this area, but none has addressed the fundamental issue which might be the major limiting factor responsible for the slow progress poor foundation in genetics and related disciplines because of rigidity in our curricula at the pre-tertiary and tertiary institutions. Empirical indications point to the need to overhaul our academic curricula at all stages in Nigeria and other underdeveloped countries with a view to aligning our educational systems with the emerging truths of genomics, bioinformatics, and computational biology.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Session III

PB2405 Preliminary Findings from a Pilot Study of Disability Identity Among Parents of Pediatric Participants in the Undiagnosed Diseases Network

Authors:

K. Mintz, S. S. Lucas-Griffin, H. Tabor; Stanford Univ. Sch. of Med., Stanford, CA

Rare and undiagnosed genetic conditions often contribute to individuals experiencing a range of functional limitations and disabilities. Yet few, if any, studies have explored experiences of disability in this population. We present preliminary findings from a qualitative study of experiences of disability identity among parents of children enrolled in the Undiagnosed Diseases Network (UDN). Twenty parents of children aged 3 to 17 enrolled in the UDN participated in the study. Each parent completed two qualitative, semi-structured interviews. We suggest that the experiences of parents in the study exemplify three ways in which clinical genetics and genomics researchers might partner with families in navigating living with significant disabilities. First, researchers may consider expanding the scope of research questions to emphasize improving a patient's quality of life, even if curing a genetic condition is unlikely. Second, researchers might help individuals and families better confront the difficulties they face applying for and receiving services for disability-related needs, for example by providing a letter documenting their disabilities. Such assistance is critical since many government programs require individuals to have a widely recognized diagnosis before receiving services. Third, many individuals interviewed found the labels associated with disability and chronic illness used in research stigmatizing. This suggests that clinical genetics and genetics research communities should carefully examine and potentially revise the language used to categorize patients and research participants. Taken together, these preliminary findings provide opportunities for genomic research studies to critically reflect on the appropriate role of and approaches to reciprocity between investigators and participants in genetics and genomics research.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Session I

PB2406 Prenatal gene editing for neurodevelopmental disease: ethical and scientific considerations.

Authors:

R. Major[1], E. Juengst[2]; [1]Univ. of North Carolina, Chapel Hill, NC, [2]Univ. of North Carolina at Chapel Hill, Chapel Hill, NC

Neurodevelopmental disorders are notoriously difficult to identify and treat early enough to make a meaningful difference in symptom alleviation. However, as prenatal diagnostic technologies become more powerful and prevalent, there may be opportunities to target diseases before they manifest. Ultrasounds, genetic screenings, exon sequencing, and whole genome sequencing could be used to detect the vast majority of neurogenetic diseases while a fetus is still in-utero. Gene editing trials are quickly making their way into the clinic and have already shown great promise in the treatment of blood and eye diseases. Although much debate exists around heritable gene editing, few have considered the ethical implications and promise of prenatal gene editing. Prenatal gene editing is particularly contentious because there are two patients to consider: the fetus and the pregnant person. This presentation will explore the ethical and scientific considerations of prenatal gene editing by reviewing current ethical perspectives and guidelines on analogous examples of fetal medicine and assessing these considerations in the context of neurodevelopmental disorders for which prenatal gene editing may be considered. The goal of this presentation is to provide a roadmap of sorts for the field to navigate the ethical and scientific landscapes of prenatal gene editing for neurodevelopmental disease.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Session II

PB2407 Prevalence of ACMG secondary findings: Implications for population health and health disparities.

Authors:

M. Macneal[1], M. Higashi[1], T. Brandt[1], J. Ge[1], W. Walker[2], D. Veenstra[3], R. Essner[2]; [1]GeneDx, Stamford, CT, [2]AdventHlth., Orlando, FL, [3]Univ of Washington, Seattle, WA

Background: Exome sequencing (ES) and genome sequencing (GS) are widely used to diagnose genetic disorders. An additional benefit of ES/GS is the detection of secondary findings (SFs) unrelated to the primary diagnosis but with medical actionability (PMID: 36143288). The American College of Medical Genetics and Genomics (ACMG) has published a guideline for reporting secondary findings (PMID: 34012068), with a recent update (ACMG SF v3.1) listing 78 genes (PMID: 35802134), including the Centers for Disease Control and Prevention's (CDC) Tier 1 conditions for primary screening: Lynch syndrome, hereditary breast and ovarian cancer syndrome, and familial hypercholesterolemia (FH) (PMID: 25577298). While population genomic screening for these disorders is likely to be cost-effective in U.S. adults younger than 40 years (PMID: 37155986), studies have shown lower implementation rates among racial and ethnic minority groups, rural communities, uninsured or underinsured people, and those with lower education and income (PMID: 35482015).
Objective: Establishing the frequency and prevalence of medically actionable SFs across racial and ethnic groups will strengthen health system efforts to design effective prevention and health improvement programs.
Methods: Deidentified ES and GS SF data (N=104066) were analyzed from the GeneDx diagnostic testing database dating back approximately ten years. This database is enriched for individuals presenting with typical ES/GS testing indications (e.g. pediatric neurodevelopmental disorders, congenital anomalies). ACMG SF frequencies were stratified by imputed ancestry.
Results: The overall prevalence of SFs was 2.35% (N=2567), with the East Asian population at a rate of 3.29%, which is statistically significant when compared to other imputed races. Statistically significant prevalence differences by imputed race were found in the following genes: *TTN* (Black and South Asian), *TTR* (Black), *HFE* (White), *APOB* (East Asian), *MYH7* (East Asian).
Limitations: Not able to distinguish founder effects among white patients (e.g. Ashkenazi Jewish, Finnish)
Conclusions: The frequency of *APOB* mutations (CDC Tier 1) are significantly higher in East Asian families, suggesting an elevated risk for FH. Higher frequencies in several groups confirms some known associations, but reveals possible novel increased disease prevalence. As self-reported race is not always available or accurate, imputed ancestry presents an opportunity to enrich sample sizes for research and population health programs. Narrowing our analysis to only SFs reduces the likelihood of confounding with pediatric illness and primary reason for testing.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Session III

PB2408 Primary care patient and provider attitudes about population-based genetic testing and a chatbot to support informed decision-making.

Authors:

S. Andrews[1], R. Moultrie[1], O. Kutsa[1], K. Williams[1], K. Carpenter[2], A. Parker[2], J. Hunter[1], B. Biesecker[1]; [1]RTI Intl., Research Triangle Park, NC, [2]Univ. of Florida Coll. of Med., Jacksonville, FL

The three CDC Tier 1 conditions (Hereditary Breast and Ovarian Cancer, Lynch Syndrome, and Familial Hypercholesterolemia) are estimated to affect 2 million people in the US; however, an estimated 90% are unaware of their genetic risk. Population-based genetic screening is a promising approach to increase identification. Genetic counselors play an essential role in genomic medicine; however, leveraging their expertise for population-level screening is not feasible or financially sustainable. Digital tools such as chatbots are a promising, scalable approach to provide pretest genetic counseling for population-level genetic screening. From May to July 2022, we conducted semi-structured interviews with patients (N=20) and primary care providers (PCPs, N=9) from primary care clinics at an academic medical center that serves a large, diverse metropolitan city. Interviews assessed attitudes about predictive genetic testing, openness to population-level genetic screening, and reactions to a chatbot as a tool to facilitate patient informed decision-making. Rapid qualitative analysis methodology was used to analyze interview data. Most patients were receptive to population screening but had concerns about privacy and discrimination based on results, possible anxiety about results, and fear of invasive procedures (e.g., blood draws). To support decision-making about genetic testing, patients expressed the need to know: potential harms; how results will be stored; procedures involved (e.g., saliva vs blood draw); and how to interpret results. PCPs expressed the need to inform patients that: early disease identification through genetic testing can improve morbidity and mortality; results could have implications for family members; and testing is completely voluntary. Moreover, PCPs expressed the need for clear follow-up procedures and concerns about insurance coverage among patients with a positive result. Finally, PCPs stressed the importance of education to ensure understanding that a negative result does not mean they will never have heart disease or cancer. Both patients and PCPs reacted positively to the chatbot as a tool to educate patients about genetic testing. They reported the tool was easy to navigate and had sufficient information to make an informed decision about genetic testing, though using a chatbot might be difficult for elderly people who are less comfortable with technology. This work provides important information about receptivity of patients and PCPs to population screening and the potential for chatbots as a tool to facilitate decision-making around the implementation of large volume, population-level genetic screening.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Session I

PB2409 Process steps and resource requirements associated with innovative genetic testing: A time and motion study in the clinical sequencing evidence-generating research (CSER) consortium

Authors:

L. Desrosiers[1,2], H. Smith[3,1], J. Bensen[4], J. Chau[5], C. Finnila[6], L. Hindorff[7], H. Hoban[8], N. Kelly[9], F. Lynch[10], B. Rolf[11], N. Sahin Hodoglugil[8], M. Wasserstein[12], T. Yip[13], K. Hassmiller Lich[4]; [1]Baylor Coll. of Med., Houston, TX, [2]Texas Children's Hosp., Houston, TX, [3]Harvard Med. Sch. and Harvard Pilgrim Hlth.Care Inst., Boston, MA, [4]Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, [5]NIH/NIAID, Bethesda, MD, [6]HudsonAlpha Inst. for Biotechnology, Huntsville, AL, [7]NIH/NHGRI, Bethesda, MD, [8]Univ. of California San Francisco, San Francisco, CA, [9]Children's Hosp. at Montefiore and the Albert Einstein Coll. of Med., Bronx, NY, [10]Kaiser Permanente Northwest, Portland, OR, [11]Univ. of Washington, Seattle, WA, [12]Albert Einstein Coll. of Med., Bronx, NY, [13]UCSF, San Francisco, CA

Background As the material cost of genomic sequencing (GS) (e.g., exome and genome) declines, use of GS in clinical care has expanded. In addition to material costs, clinical GS requires personnel time and care delivery infrastructure, the costs of which are understudied. The NHGRI Clinical Sequencing Evidence-Generating Research (CSER) consortium consisted of 6 projects investigating implementation of GS innovations (GSI) in care of diverse populations, thus providing a unique setting to explore resource requirements for GSI implementation.
Methods Detailed process flow diagrams were generated for each project and 11 process steps common across the projects were defined (cPS). In 5 of 6 projects, at least 20 patients were followed and data on personnel role and time were collected for each project's detailed process steps (dPS). We calculated the mean time for each dPS and summed means at the cPS to compare across projects.
Results Data are collected for 15 prenatal, 113 pediatric, and 24 adult patients (n=152), of which 107 (70%) were medically underserved or underrepresented (MUS-UR) and 48 (32%) had complex GS results. Across projects, GSI averaged 202 min (Range of Averages (ROA): 83-341 min). Genetic counselors (GCs) spent the most time (mean 131 min; ROA: 57-241 min) per patient, followed by research coordinators (mean 38 min; ROA: 0-77 min) and MDs (mean 25 min; ROA: 0-86 min).
cPS8 "Review, prepare, and return results to patients" took the most time (mean 69 min; ROA: 35-115 min) followed by cPS5 "Pre-test clinic visit" (mean 50 min; ROA: 11-86 min). cPS10 "Hand-off to next step" required the least time (mean 4 min; ROA: 1-7 min).
For MUS-UR patients, genomic innovations averaged 204 min (ROA: 81-341 min) compared with 208 min (ROA: 82-364 min) for patients not designated as MUS-UR. GC time averaged 132 min (ROA: 56-246 min) for MUS-UR and 132 min (ROA: 47-262 min) for not MUS-UR. GSI for patients with complex results averaged 244 min (ROA: 88-372 min) compared with 174 min (ROA: 68-305 min) for non-complex (NC) results. There was a 52 min increase in average time required for cPS8 in complex (Avg: 101 min) versus NC results (Avg: 49 min).
Conclusion In addition to direct costs associated with GSI, implementation also requires resources in the form of staff and provider time. A large portion of this time spent by GCs. Our analysis likely underestimates the time required in clinical (vs. research) settings, and we did not capture all steps within each project (e.g., laboratory processes) due to practical limitations. Implementation of GSI will require decision makers to consider personnel resources, particularly for patients with complex results.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Session II

PB2410 Providing trainees hands-on experience in clinical genetics research through a case report writing workshop increases trainee engagement in clinical genetics.

Authors:

J. Owens[1], R. Hopkin[1], V. Viswanath[2], L. Alvey[3], A. Shillington[4]; [1]Cincinnati Children's Hosp., Cincinnati, OH, [2]Univ. of Cincinnati Sch. of Med., Cincinnati, OH, [3]Univ. of Cincinnati Coll. of Med., Cincinnati, OH, [4]Cincinnati Children's Hosp., Cincinnatti, OH

Nearly all clinical geneticists have interesting patients and multiple small projects they would like to publish if they had time. Meanwhile, the number of trainees pursuing careers in clinical genetics has struggled to match the recent increased demand for genetics services. To increase medical student interest and engagement with clinical genetics, we created a “Report Writing Workshop” from our local genetics student interest group to provide an opportunity for medical students to obtain hands-on experience in writing and publishing small clinical projects.Twenty 1st and 2nd year medical students volunteered to write ten case reports. The workshop aimed to break down the writing process into small, approachable tasks and allow participants to make autonomous decisions allowing ownership over their work. 16 potential case reports were collected from genetics attendings and residents, who had intended to write the cases themselves in the future. Medical students were asked to choose 2-3 projects that were most interesting to them. Participants placed themselves in groups or were assigned to groups based on mutual interest in cases. The workshop included five ~90 minute meetings that respectively focused on an introduction to report development, literature review, chart review, writing an introduction, and writing a clinical history. The first 10 minutes of meetings summarized the topic of the meeting, with the remainder of the time provided for teams to work on data gathering and writing. In developing the discussion for each paper, groups worked independently to provide answers to targeted take home points that the report added to the field. These answers were combined and organized together for publication. Participants provided the core content of the manuscript to a genetics fellow, who reviewed and edited their work to reflect the terminology and content expected for academic publication. All 10 case reports will be submitted to medical genetics journals, and all 20 students intend to present their work at ASHG in 2023.This scalable approach allowed for a large cohort of medical students to be mentored by a small group of genetics fellows overseen by a single attending. We attribute success of the workshop to the quality cases that merited publication, the group dynamics that added a social component to maintain momentum, and breaking down the process into specific small steps for each section. In our experience, it benefited the attendings by moving their projects forward, improved the fellows ability to provide medical education and mentorship opportunities, and increased medical student interest in clinical genetics.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Session III

PB2411 † Psychological and behavioral outcomes of returning all clinically relevant secondary findings from genomic sequencing: Preliminary results from the Incidental Genomics RCT

Authors:

Y. Bombard[1,2], C. Mighton[2,1], S. Shickh[2,1], M. Clausen[2], R. Kodida[2], E. Reble[2], J. Sam[2], S. Grewal[2], S. Panchal[3,4], M. Aronson[5,4], S. R. Armel[6,4], T. Graham[7,4], N. Forster[8], J-M. Capo-Chichi[9,10], E. Greenfeld[11,10], A. Noor[11,10], I. Cohn[12], C. F. Morel[8,13], C. Elser[3,13], A. Eisen[7], J. C. Carroll[14,15], E. Glogowski[16], K. Schrader[17,18], J. Lerner-Ellis[11,10], R. H. Kim[6,13], K. E. Thorpe[19]; [1]Inst. of Hlth.Policy, Management & Evaluation, Univ. of Toronto, Toronto, ON, Canada, [2]Genomics Hlth.Services Res., St. Michaels Hosp., Unity Hlth.Toronto, Toronto, ON, Canada, [3]The Marvelle Koffler Breast Ctr., Mount Sinai Hosp., Sinai Hlth., Toronto, ON, Canada, [4]Dept. of Molecular Genetics, Univ. of Toronto, Toronto, ON, Canada, [5]Zane Cohen Ctr. for Digestive Diseases, Mount Sinai Hosp., Sinai Hlth., Toronto, ON, Canada, [6]Bhalwani Familial Cancer Clinic, Princess Margaret Cancer Ctr., Toronto, ON, Canada, [7]Odette Cancer Ctr., Sunnybrook Hlth.Sci. Ctr., Toronto, ON, Canada, [8]Fred A. Litwin Family Ctr. in Genetic Med., Univ. Hlth.Network, Toronto, ON, Canada, [9]Genome Diagnostics Div., Toronto Gen. Hosp., Univ. Hlth.Network, Toronto, ON, Canada, [10]Lab. Med. and Pathobiology, Univ. of Toronto, Toronto, ON, Canada, [11]Pathology & Lab. Med., Mount Sinai Hosp., Sinai Hlth., Toronto, ON, Canada, [12]Div. of Clinical Pharmacology & Toxicology, Hosp. for Sick Children, Toronto, ON, Canada, [13]Dept. of Med., Univ. of Toronto, Toronto, ON, Canada, [14]Dept. of Family & Community Med., Univ. of Toronto, Toronto, ON, Canada, [15]Granovsky Gluskin Family Med. Ctr., Mount Sinai Hosp., Sinai Hlth., Toronto, ON, Canada, [16]Sanofi, North Bergen, NJ, [17]British Columbia Cancer Agency, Vancouver, BC, Canada, [18]Dept. of Med. Genetics, The Univ. of British Columbia, Vancouver, BC, Canada, [19]Dalla Lana Sch. of Publ. Hlth., Univ. of Toronto, Toronto, ON, Canada

Introduction: Genomic sequencing (GS) is an effective diagnostic test, but questions remain about which types of clinically relevant secondary findings (SFs) should be returned. Guidelines call for further evidence on health outcomes of returning SFs. Other studies have returned limited SFs and without controls; evidence is lacking on outcomes of simultaneously returning all SFs. We conducted an RCT (Incidental Genomics trial, NCT03597165) to evaluate outcomes and costs of returning all clinically relevant SFs. Here, we report preliminary findings on psychological and behavioral outcomes. **Methods:** Adults with cancer were recruited and randomized. Participants in both arms had GS with cancer results. Intervention arm participants could choose to learn multiple types of SFs: medically actionable, Mendelian, early-onset neurodegenerative, carrier status, common disease risk variants, and pharmacogenomic variants. The primary outcome was distress, measured by the Hospital Anxiety and Depression Scale (HADS) 2 weeks after return of results. Secondary outcomes included patient-reported medical (e.g., appointments, screening) and lifestyle behaviors (e.g., diet, exercise), 2 weeks, 6 weeks, 6 months, and 1-year post-return of results. HADS at 2-weeks was compared between arms with an ANCOVA model to adjust for baseline scores. Proportions of participants reporting medical and lifestyle behaviors were compared at all timepoints with chi-square tests or Fisher's exact test. **Results:** Participants (n=287) were 87.1% female, 57.5% White/European, and average 57.2 years old. The adjusted mean difference in HADS anxiety in the intervention arm compared to the control arm was -0.19 (95% CI -1.1 to 0.68, p=0.7), implying less increase. The CI excludes the minimal clinically important difference (MCID: 2.5). The adjusted mean difference in HADS depression in the intervention arm compared to the control arm was 0.02 (95% CI -0.67, 0.71, p=0.96), with the CI also excluding the MCID (2.5). A difference in health behavior uptake was not observed at 6 months. There was higher uptake of medical behaviors (e.g., appointments) in the intervention arm at 6 weeks (14.8% vs. 3.1%, p=0.002). There was higher uptake of lifestyle behaviors in the intervention (e.g., diet changes) arm at 2 weeks (7.4% vs. 0%, p=0.002) and 12 months (4.4% vs. 0%, p=0.03). **Discussion:** We did not find strong statistical evidence that returning SFs increased distress relative to GS with primary findings only, however we found evidence that SFs led to higher uptake of medical and lifestyle-related behaviors. These findings suggest returning SFs is safe and effective for influencing health behaviors.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Session I

PB2412 Racial Disparity in the Diagnosis of Mitochondrial Disease

Authors:

D. McGinn[1], C. Muraresku[2], K. Valverde[3], M. Falk[2], R. Ganetzky[2]; [1]Children s Hosp. of Philadelphia and Perelman Sch. of Med. at the Univ. of Pennsylvania, Philadelphia, PA, [2]Children's Hosp. of Philadelphia, Philadelphia, PA, [3]Univ. of Pennsylvania, Philadelphia, PA

Background: Primary Mitochondrial Diseases (PMD) caused by inherited disruption of oxidative phosphorylation are collectively the most common inborn error of metabolism with broad variability including age at presentation and severity. A 2020 North American study of individuals with PMD reported >85% were white, significantly discrepant from the 2020 US Census where 61% self-reported as white. Here we present results of an investigation of potential racial disparities in ascertainment and disease severity in individuals with PMD to determine if patients of color are underdiagnosed. **Methods:** We retrospectively reviewed electronic health records on 337 patients with molecularly confirmed PMD under an IRB-approved study. Only probands with sufficient available demographic, diagnostic setting (inpatient/outpatient), and symptom data were included. Symptom severity was assessed using a multisystem aggregate severity score index [SSI], where lower scores indicate more severe disease. Potential confounders including socioeconomic status, distance to the hospital, and insurance coverage were also collected. **Results:** 207 patients met inclusion criteria: 53.6% were female; Mean age was 18.1 years; 74.9% were white, 11.1% Hispanic, 9.7% Asian, and 7.2% Black; Mean age at PMD diagnosis was 9.72 years; 75.8% were diagnosed as an outpatient; average SSI was 13.55. Notably, Black (mean SSI=12.0, p=0.013) and Hispanic (mean SSI=12.5, p=0.016) patients had more severe symptoms compared with white patients with non-Hispanic backgrounds (mean SSI=14.2). Notably, Asian, Black, and Hispanic patients were significantly more likely to be diagnosed as inpatients (40%, 40%, and 35%, respectively) compared with white PMD patients (20%). Self-identified patients of color were more likely to be diagnosed with PMD in the inpatient setting. Additionally, Black and Hispanic patients tended to have more severe symptoms as compared with those from other backgrounds. **Conclusions:** This data supports the premise that patients of color tend to require a higher diagnostic threshold to achieve a diagnosis of PMD; previous reports demonstrating a higher prevalence of patients self-reported as white may be influenced by ascertainment bias; and people of color are likely under-diagnosed within the PMD community, representing a critical but previously unrecognized health disparity. Given a major barrier to diagnosis is obtaining a referral to a mitochondrial medicine clinic, overcoming this disparity requires wide-scale education of healthcare providers regarding the pan-racial occurrence of PMD.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Session II

PB2413 Rapid whole-genome sequencing in children in acute care in a universal health care system: Physicians underestimate parental expectations and concerns about the impact on their child's care

Authors:

C. Azuelos[1,2], C. Varin-Tremblay[1,2], J. L. Michaud[1,2], A-M. Laberge[1,2]; [1]CHU Sainte-Justine, Montreal, QC, Canada, [2]Université de Montréal, Montreal, QC, Canada

Purpose: The use of rapid whole genome sequencing (rGS) in the investigation of rare disease is used increasingly as a diagnostic tool for children in acute care. Parental and physician experiences of rGS in a universal health care system need to be better understood to ensure its appropriate use. This study aims to assess parental and physician perspectives about rGS and the impact of results on clinical care.
Methods: A 2-year prospective provincial multi-center study is ongoing in the four pediatric academic centers in Quebec, Canada. Children are eligible if they are aged 0-18 years and hospitalized for an acute condition with a high likelihood of genetic etiology. In addition to rGS, parents and physicians complete questionnaires at recruitment and one-month after disclosure of results. Parental questionnaires include sociodemographic data, understanding of rGS, expectations, and concerns about rGS results. Physician questionnaires include their assessment of parental understanding and concerns and their own expectations of test results.
Results: By April 2023, parents and physicians had completed questionnaires at recruitment for 39 children. They had similar assessments of the likelihood of identifying a genetic cause with rGS (p=0.89). Parents understood rGS well (mean knowledge score 4.38/5). Parental expectations about the impact of rGS results for their child are significantly higher than physicians' expectations for getting a clearer prognosis (4.08 vs 2.85; p=0.0001), guiding management (4.42 vs 2.26; p<0.0001), influencing the level of care (4.21 vs 1.33; p<0.0001), identifying an actionable secondary finding (3.22 vs 1.13; p<0.0001) and enabling genetic counseling (3.84 vs 2.90; p=0.0018). Physicians also underestimate parents' level of worry about rGS results (1.73 vs 2.50; p=0.03).
Conclusion: Parents understand rGS and have high expectations about the impacts of test results. Physicians tend to underestimate parental expectations and concerns. Data collection is ongoing and further analyses will include comparison with parental and physician perspectives post-disclosure of rGS results. These findings will inform the implementation of rGS in clinical care in our province, to reduce the discrepancies between physician and parent expectations.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Session III

PB2414 Results of survey about concerns and motivation of the Latvian general population and researchers towards citizen-science research projects in the field of genomics.

Authors:

V. Rovite[1], A. Svandere[1], S. Mezinska[2], K. Kaleja[1], R. Peculis[1], N. Kante[1]; [1]Latvian BioMed. Res. and Study centre, Riga, Latvia, [2]Univ. of Latvia, Faculty of Med., Riga, Latvia

Introduction. Citizen-science is research carried out by citizens or amateur scientists in cooperation with scientists or within the framework of specific scientific projects, based on scientifically developed methods. Examples of citizen-science include bird watching and counting, collecting environmental data, performing astronomical observations and reporting these data to scientists. Citizen-science makes science accessible to the public and promotes public trust. The study aimed to evaluate the concerns and motivation of the Latvian general population and researchers towards citizen-science research projects in the field of genomics.Methods. We developed a specific survey that consisted of seven citizen-science research project examples (vignettes) that included various levels of participant involvement, use of data and relation to genetic research. After each research project vignette description survey participants need to answer questions about their motivation and concerns to participate. The same vignettes were used for the researcher survey, but the researchers were asked to answer questions on their potential motivation or concerns using such citizen-science research data for their research studies. The survey data was analysed according to specific vignettes where we analysed what was the most frequent motivation and concerns in specific types of studies for potential research participants from the general population. We also looked at how the concerns and motivation differed for the same individual for different types of studies and what were the main principal differences in opinions of researchers and citizens. Results and conclusions. For researchers the greatest motivation to use citizen-science were possibility to obtain more data with lower costs, but concerns were regarding the quality and traceability of data collection. For general public the most relevant motivation to participate in citizen-science activities was personal benefits or potential help to society in future, but concerns were the lack of time to contribute to these activities and data protection issues.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Session I

PB2415 † Scientists' views on whether genome editing warrants unique governance guidelines.

Authors:

R. Cadigan[1], M. Waltz[1], M. Flatt[2], A. Davis[1], J. Conley[1], G. Henderson[1], G. Inamine[1], E. Juengst[1]; [1]Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, [2]Cuyahoga Community Coll., Highland Hills, OH

Advances in genome editing technology such as CRISPR have revolutionized biomedical research. As human clinical trials get underway, early results have been both promising and cautionary. They also reinvigorate questions as to how best to regulate such research to address ethical concerns while also encouraging important biomedical advances. Governance efforts, like the International Commission on the Clinical Use of Human Germline Genome Editing and a World Health Organization Expert Advisory Committee, have been established to provide guidance on conducting genome editing research and its translation to the clinic. To gain insight into how these international governance efforts align with the views of practicing scientists, we conducted interviews with over 60 international scientists involved in genome editing to explore their perceptions of the ethical issues associated with human genome editing research and how such research should be governed. Our ongoing analyses of these interviews have yielded noteworthy and sometimes contradictory perspectives on the question of whether genome editing research warrants unique governance guidelines. While scientists may agree with descriptions of genome editing technology as unparalleled and exceptional with unique ethical implications, many simultaneously reject calls for special forms of governance, arguing that genome editing is no different than other emerging technologies and can be adequately governed by existing regulatory mechanisms. In this presentation, we will analyze this contradiction by exploring the ways in which scientists accept or reject the idea of genome editing as an exceptional technology and how this relates to their thoughts on governance. Debates on whether existing regulations are sufficient, especially within the context of heritable editing and so-called ‘rogue’ editing, will be presented. We will conclude with a discussion of the implications that these scientists’ rhetorical contradictions may have for future genome editing governance efforts at national and global levels.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Session II

PB2416 "Should I let them know I have this?": Experiences with genetic discrimination amongst patients with hereditary cancer syndromes.

Authors:

R. Gopalakrishnan[1], J. Sam[1], C. Butkowsky[1,2], E. Reble[1], M. Clausen[1], S. Rajeziesfahani[3], B. Sparkes[3], V. Aguda[1], M. Aronson[4], D. Bishop[5], L. Dawson[3], A. Eisen[6], T. Graham[6], J. Green[3], C. Mighton[1,2], J. Pauling[7], C. Pavao[8], P. Pechlivanoglou[2,9], C. Remocker[8], S. Savas[3], S. Sun[10], T. Tiano[7], A. Tilley[5], K. Thorpe[1,2], K. Schrader[10], H. Etchegary[3], Y. Bombard[1,2]; [1]St. Michael's Hosp., Toronto, ON, Canada, [2]Univ. of Toronto, Toronto, ON, Canada, [3]Mem. Univ. of Newfoundland, St. John's, NL, Canada, [4]Zane Cohen Ctr., Sinai Hlth.System, Toronto, ON, Canada, [5]Patient Partner, St. John's, NL, Canada, [6]Sunnybrook Hlth.Sci. Ctr., Toronto, ON, Canada, [7]Patient Partner, Toronto, ON, Canada, [8]Patient Partner, Vancouver, BC, Canada, [9]The Hosp. for Sick Children, Toronto, ON, Canada, [10]BC Cancer, Vancouver, BC, Canada

Introduction: Hereditary cancer syndromes (HCS) represent approximately 1 in 10 of all cancer patients. Among HCS, Hereditary Breast and Ovarian Cancer Syndrome (HBOC) and Lynch Syndrome (LS) are the most prevalent. HCS patients are genetically predisposed to developing cancer, and require complex care which may include genetic testing, screening and prophylactic surgery. In addition to medical burdens, many patients are concerned about stigmatization based on their HCS diagnosis; this study aims to describe experiences with genetic discrimination that patients may face on their care journey.
Methods: Semi-structured qualitative interviews were conducted with HCS patients residing in Ontario, British Columbia and Newfoundland & Labrador, Canada. All interviewees had a confirmed molecular diagnosis of HBOC or LS. Interpretive description was used to analyze the data.
Results: Across Ontario (n=26), British Columbia (n=23) and Newfoundland & Labrador (n=24), 73 patients with HBOC (n=39) and LS (n=34) were interviewed. The sample consisted of 51 females, 21 males and 1 gender diverse individual. Overarchingly, patients worried about whether to share their HCS genetic diagnosis with others due to fears of being judged, stigmatized or discriminated against. Genetic discrimination regarding insurance coverage was of particular concern; patients discussed experiences where they were denied coverage, received lesser coverage or paid higher fees upon disclosing their HCS status. These experiences extended to various insurance types, including life, health and disability insurance. Patients noted that insurance companies had roundabout ways of soliciting family health and genetic testing history without explicitly asking if an individual was positive for a certain cancer-associated gene. Beyond the insurance space, patients were also wary of genetic discrimination by employers, family and friends as they felt sharing their HCS diagnosis may limit personal and professional opportunities. For example, patients did not want employers to question their job performance nor did they want family members to comment on their appearance after being branded as “sick.” Lastly, many patients said they were unaware of the Genetic Non-Discrimination Act and expressed interest in learning more about the Act.
Conclusion: Genetic discrimination is a concern for many HCS patients, especially in terms of impacts on insurance coverage. Despite non-discrimination legislation, patients are uncertain of their rights and wary about sharing their diagnosis with others. This work emphasizes the need to address genetic discrimination experienced by HCS patients.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Session I

PB2418 Sickle cell trait insights from the *All of Us* research program.

Authors:

H. Ramirez[1], F. Wossenseged[1], D. Seddighi[1], A. J. Buscetta[1], Y. Wang[2], P. Joseph[2], V. L. Bonham[1]; [1]NIH/NHGRI, Bethesda, MD, [2]NIH/NIAAA, Bethesda, MD

Sickle cell trait (SCT), HbAS, is one of the most common hemoglobin mutations in the world, with an estimated prevalence of 300 million carriers globally. In the United States, more than 3 million individuals have SCT. Since 2006, all newborns in the United States have been screened for SCT in all 50 states, but adults' knowledge of their carrier status and documentation in their electronic health record is unknown. The hematology research community has called for increased large scale epidemiological and genomic research on SCT, including large population-based studies. In addition, there is limited research on clinical complications that may be associated with SCT carrier status. To address this gap, electronic health record (EHR) and whole genome sequencing (WGS) data from the NIH *All of Us* (AoU) Research Program was used to identify SCT carriers. Carrier status was determined using ICD-9 and ICD-10 codes (282.5 and D57.3, respectively) in the EHR and by genomic analysis ("AT" genotype of rs334). As of February 2023, the total cohort within the *AoU* database included 409,420 participants, of which 254,700 participants (62.2%) have EHR data and 245,400 (59.9%) have WSG data available. Of those with EHR data, a cohort of 1,061 (0.42%) participants were identified as SCT carriers via ICD diagnosis. We determined that 750 (70.6%) participants within the EHR cohort (n=1061) also have WGS data. Through genomic analysis, 595 (79.3%) participants were confirmed as sickle cell carriers with heterozygous genotype of rs334. Furthermore, of the 245,400 participants with WGS data within the *AoU* research program, 5,484 (2.2%) SCT carriers with heterozygous genotype of rs334 were identified. Further investigation of this cohort will be done to determine how many participants also have EHR data available in the *AoU* database. We will report on the association of nine clinical outcomes and carrier status (carrier vs. non-carrier). We will utilize a matched control cohort without known history of SCT or other hemoglobinopathies. The nine clinical outcomes of interest were identified from a 2018 systematic review, Clinical Outcomes Associated with Sickle Cell Trait (PMID: 30383109). The discrepancies between EHR-reported SCT carriers and those identified and verified via genomic analysis in the *AoU* population informs the need for adult screening for SCT, comprehensive education, and guidelines for communication of rare health risk for newly identified carriers.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Session II

PB2419 System Requirements to Support Primary Care Pediatricians in the Return of Genome Informed Risk Assessment

Authors:

J. Connolly[1], D. Karavite[1], S. Terek[2], M. Harr[3], H. Hakonarson[1], R. Grundmeier[4]; [1]Children's Hosp. of Philadelphia, Philadelphia, PA, [2]The Children s Hosp. of Philadelphia, Philadelphia, PA, [3]Children S Hosp. of Philadelphia, Philadelphia, PA, [4]Childrens Hosp. of Philadelphia, Philadelphia, PA

Purpose: Genome informed risk assessment (GIRA) - informed by polygenic risk scores, family history, and patient history - is a new testing modality that has the potential to improve healthcare by identifying individuals at risk for common conditions. However, returning GIRA results in clinical practice requires assessment of the needs and capabilities of providers and healthcare systems. As part of an electronic Medical Records and Genomics (eMERGE) network initiative, we conducted a systems requirement assessment to inform the design of a comprehensive return of GIRA results and clinical decision support system for pediatric primary care providers (PCPs).
Methods: PCPs from ten practices affiliated with Children's Hospital of Philadelphia were eligible to participate. We administered a seven-phase interview comprised of surveys and semi-structured interview questions to elicit PCP responses across a spectrum of sociotechnical requirements for a system to manage GIRA results. Interviews were conducted with 20 primary care providers (PCPs) from ten practices within a single health system.
Results: Twenty participants completed the interview process (19 doctors, 1 nurse practitioner). Results demonstrated that PCPs are interested in being involved in the process of returning genetic testing results to patients, but require a comprehensive system that delivers actionable recommendations, supports patient education, streamlines communication, and facilitates multi-disciplinary collaboration. Further, analyses highlighted the need to integrate risk reports with clinical decision support tools that can accommodate different clinical scenarios and provider workflow preferences.
Conclusion: These findings underscore the point that the return of genetic testing results and decision support for PCPs extends beyond the electronic health record, and include systems and processes for information delivery, patient education, communication, and collaboration.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Session III

PB2420 Systematic use of NGS as significant impact factor for diagnosis of rare diseases in national health care system

Authors:

M. Vinkšel[1], B. Peterlin[1], K. Writzl[2], A. Maver[1]; [1]Univ. clinical centre Ljubljana, Ljubljana, Slovenia, [2]Clinical Inst. of Med. Genetics UMCL, Ljubljana, Slovenia

Rare diseases are numerous, heterogeneous and diverse and are in 80% of genetic origin, mostly chronic diseases, and many cause early death. Due to the low prevalence of individual rare disease, lack of knowledge, scarcity of expertise and life-treating nature, rare diseases to emerge as a public health priority worldwide. Novel technologies such as NGS approaches has transformed approach to the diagnosis, which is now more comprehensive, timely and focused on a patient. In Slovenia, NGS was implemented for a systematic use in a national health care system in 2012 for pediatric and adult patients, who presented with a hypothesis of any genetic disease. An overall diagnostic yield of 34% was achieved. Diagnostic yield across diverse organ systems was 23%-56,7%. For cases with a clinical diagnostic hypothesis for monogenic disease genetic etiology was established in 47%. For unclear diagnostic hypothesis or for diagnostic hypothesis with a possible monogenic etiology approximately one third of the referrals were resolved. Among reported 1517 genes 945 genes were clinically actionable. Therefore, NGS approaches should be consider as first tier diagnostictool in the national health systems due to its importance for ending diagnostic odyssey, provision of optimal medical care and prevention.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Session I

PB2421 The Clinician-reported Genetic testing Utility InDEX (C-GUIDE): Development and validation for prenatal care

Authors:

R. Hayeems[1,2], S. Luca[1], B. Xiao[1], C. Armour[3,4], C. Boswell-Patterson[3], K. Boycott[3], D. Dyment[3], J. Lazier[3], L. Chad[1], T. Kolar[5], W. J. Ungar[1,2]; [1]The Hosp. for Sick Children, Toronto, ON, Canada, [2]Univ. of Toronto, Toronto, ON, Canada, [3]Children's Hosp. of Eastern Ontario, Ottawa, ON, Canada, [4]Univ. of Ottawa, Ottawa, ON, Canada, [5]British Columbia Children's Hosp., Vancouver, BC, Canada

Background: Integrating new genetic tests into clinical practice requires evidence of clinical utility. Our team previously developed and validated a measure of clinical utility, the Clinician-reported Genetic testing Utility InDEX (C-GUIDE), in a rare disease pediatric population. However, the utility of genetic testing for clinical decision-making and patient care is patient population-specific. In prenatal care, genetic test results may trigger decisions related to pregnancy interruption, monitoring, delivery, or neonatal management. Therefore, it is important to define and measure clinical utility for this specific population.
Aim: To develop and validate a measure of clinical utility for genetic testing in prenatal care.
Methods: First, we modified the wording of C-GUIDE to suit the prenatal context. Face and content validity were assessed by 19 geneticists and genetic counsellors who routinely order prenatal genetic testing. Feedback on the wording and relevance of each item was obtained through audio-recorded interviews and a survey in which participants were asked to accept, modify or reject items. Feedback was synthesized to inform revisions. Next, to test the construct validity of C-GUIDE Prenatal, 4 geneticist raters completed C-GUIDE on a retrospective sample of cases that received prenatal genetic testing. They also completed a global assessment of utility using a single anchor item. Construct validity was assessed using a generalized estimating equations model.
Results: Most C-GUIDE items were considered relevant. Clarification regarding the timepoint at which utility was to be assessed (i.e., pre- vs. post-natal) and the index patient in question (i.e., fetus or pregnant person) was required. Recommended items were added. Of the 101 retrospective cases reviewed in the validation study, genetic results were positive and informative in 32 cases, negative and informative in 38 and negative and uninformative in 31. On average, a 1-point increase in the global item score was associated with an increase of 1.2 in the C-GUIDE score ($p=0.04$). Compared to uninformative results, informative positive and informative negative results were associated with an increase of 11.1 ($p<0.001$) and 5.5 ($p<0.001$) in the C-GUIDE score, respectively.
Conclusions: The significant positive associations between C-GUIDE total and the global item score and between C-GUIDE total and result type, in the hypothesized direction, provide evidence of construct validity. This tool may contribute to a body of evidence pertaining to the clinical utility of prenatal genetic testing, in turn informing adoption and coverage decisions.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Session II

PB2422 The Estonian Biobank's MyGenome Portal: A comprehensive platform for return of results to over 200,000 biobank participants.

Authors:

N. Pervjakova[1], L. Leitsalu[1], K. Metsalu[1], K. Kruusmaa[1], P. Pärkson[1], K. Teder[1], K. Kaasla[1], L. Sissas[1], K. Fischer[2], V. Kukushkina[1], P. Kleemann[1], A. Reigo[1], H. Alavere[1], K. Krebs[1], S. Reisberg[3], G. Hudjashov[1], V. Pankratov[1], M. Nelis[1], L. Sirje[1], K. Maal[1], K. Läll[1], D. Sokurova[1], R. Mägi[1], S. Kasela[1], L. Luitva[1], A. Metspalu[1], M. Metspalu[1], L. Milani[1]; [1]Inst. of Genomics, Univ. of Tartu, Tartu, Estonia, [2]Inst. of Mathematics and Statistics, Univ. of Tartu, Tartu, Estonia, [3]Inst. of computer Sci., Univ. of Tartu, Tartu, Estonia

The MyGenome Portal (MGP) of the Estonian Biobank is a research platform with multiple objectives: provide individual results to over 200,000 biobank participants; ensure transparency regarding the use of participants' data; improve public health and genomic literacy; serve as a platform for research; and support precision medicine initiatives. The MGP is structured into three sections: personalized results, educational content, and studies. The portal incorporates dynamic consent that allows participants to choose specific categories for the return of results and which research projects they wish to join.

The personalized results section currently provides participants with reports on type 2 diabetes (T2D) and coronary artery disease (CAD) as well as pharmacogenomics, caffeine metabolism, and ancestry at different resolutions. For T2D and CAD, the reports incorporate 10-year and lifetime cumulative risk calculations based on polygenic risk scores, lifestyle factors, and prevalent diseases. The pharmacogenomics report covers 26 medications with high-evidence recommendations. The educational section complements the personalized results by providing respective background information. The interactive components and visual aids of the reports together with the educational section empower participants to gain a comprehensive understanding of the topics covered, facilitating informed decision-making regarding their health.

The dedicated research section serves as a digital solution for data collection. Among the first projects is an investigation into the perceived impact of genetic information reported through the portal. The research section will expand to include questionnaires, randomized clinical studies, and other research initiatives. The development of the MGP involved consulting with expert groups and a qualitative study to assess user experience. Participants expressed satisfaction with the user-friendly interface and overall concept of the portal. They also expressed interest in receiving personalized reports and recommendations for a healthy lifestyle. Next, a preliminary launch is planned with 10,000 participants, with the full launch scheduled by the end of 2023 (October 2023).

Beyond facilitating the communication of personalized risk information, the MGP acts as a testbed for precision medicine initiatives. The study group within the MGP enables the creation of various sub-populations that represent the general population, allowing estimation of the impact on public health. The MGP empowers individuals, and promotes research, contributing to the ongoing transformation of healthcare.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Session III

PB2423 The Genetics Navigator: Protocol for a Mixed Methods Randomized Controlled Trial Evaluating a Digital Platform to Deliver Genomic Services across Pediatric and Adult Populations

Authors:

G. D'Amours[1], M. Clausen[1], S. Luca[2], E. Reble[1], R. Kodida[1], D. Assamad[2], M. Aronson[3,4], F. Bernier[5], M. Brudno[6,7], J. C. Carroll[8,9], L. Chad[10,11], R. Cohn[10], G. Costain[10,11], I. Dhalla[12,13], H. Faghfoury[14], J. Friedman[15], S. Hewson[10], T. Jamieson[16], R. Jobling[10], A-M. Laberge[17,18], J. Lerner-Ellis[19,20,21], E. Liston[3,10], M. Mamdani[1], C. Marshall[21,22], M. Osmond[23], Q. Pham[12,24], F. Rudzicz[25,26], E. Seto[12,24], S. Shastri-Estrada[27,28], C. Shuman[3], J. Silver[3,14], M. Smith[29], K. Thorpe[1,30], W. J. Ungar[2,12], R. Hayeems[2,12], Y. Bombard[1,12]; [1]Li Ka Shing Knowledge Inst., St. Michael's Hosp., Unity Hlth.Toronto, Toronto, ON, Canada, [2]Child Hlth.Evaluative Sci., The Hosp. for Sick Children, Toronto, ON, Canada, [3]Dept. of Molecular Genetics, Univ. of Toronto, Toronto, ON, Canada, [4]Zane Cohen Ctr., Sinai Hlth.System, Toronto, ON, Canada, [5]Dept. of Med. Genetics, Alberta Children's Hosp., Calgary, AB, Canada, [6]Ctr. for Computational Med., The Hosp. for Sick Children, Toronto, ON, Canada, [7]Dept. of Computer Sci., Univ. of Toronto, Toronto, ON, Canada, [8]Dept. of Family Med., Sinai Hlth.System, Toronto, ON, Canada, [9]Dept. of Family and Community Med., Univ. of Toronto, Toronto, ON, Canada, [10]Div. of Clinical and Metabolic Genetics, The Hosp. for Sick Children, Toronto, ON, Canada, [11]Dept. of Pediatrics, Univ. of Toronto, Toronto, ON, Canada, [12]Inst. of Hlth.Policy, Management and Evaluation, Univ. of Toronto, Toronto, ON, Canada, [13]Care Experience Inst., Unity Hlth.Toronto, Toronto, ON, Canada, [14]Fred A. Litwin Family Ctr. in Genetic Med., Univ. Hlth.Network & Mount Sinai System, Toronto, ON, Canada, [15]Dept. of Med. Genetics, Univ. of British Columbia, Vancouver, BC, Canada, [16]Dept. of Med., Univ. of Toronto, Toronto, ON, Canada, [17]Div. of Med. Genetics, Dept. of Pediatrics, CHU Sainte-Justine, Montréal, QC, Canada, [18]Dept. of Pediatrics, Université de Montréal, Montréal, QC, Canada, [19]Lunenfeld Tanenbaum Res. Inst., Sinai Hlth.System, Toronto, ON, Canada, [20]Pathology and Lab. Med., Sinai Hlth.System, Toronto, ON, Canada, [21]Dept. of Lab. Med. and Pathobiology, Univ. of Toronto, Toronto, ON, Canada, [22]Dept. of Pediatric Lab. Med., The Hosp. for Sick Children, Toronto, ON, Canada, [23]Childrens Hosp. of Eastern Ontario Res. Inst., Univ. of Ottawa, Ottawa, ON, Canada, [24]Ctr. for Digital Therapeutics, Toronto Gen. Hosp. Res. Inst., Univ. Hlth.Network, Toronto, ON, Canada, [25]Vector Inst. for Artificial Intelligence, Toronto, ON, Canada, [26]Faculty of Computer Sci., Dalhousie Univ., Halifax, NS, Canada, [27]Genetics Navigator Advisory Board, Toronto, ON, Canada, [28]Dept. of Occupational Sci. and Occupational Therapy, Temerty Faculty of Med., Toronto, ON, Canada, [29]Patient Partner, Canadian Organization for Rare Disorders, Toronto, ON, Canada, [30]Dalla Lana Sch. of Publ. Hlth., Univ. of Toronto, Toronto, ON, Canada

Background. Genetic testing is increasingly used in mainstream care across medical disciplines, leading to increased demand for genetics services. Combined with workforce shortages, this has resulted in growing waitlists and pressure on care delivery. E-health tools are one scalable and patient-centred solution; however tools developed to date are limited in scale and none integrate all components of genetics service delivery. The Genetics Navigator is an interactive patient-centred digital platform that collects medical and family history, provides pre- and post-test genetic counseling, and returns results. The study objective is to measure the effectiveness of the Genetics Navigator across pediatric and adult populations when integrated with usual care by a genetics clinician compared to usual care alone. **Methods and analysis.** One hundred and thirty-two participants (66 patients and 66 parents of patients eligible for genetic testing) will be recruited from three Ontario genetics clinics. Participants will be randomly assigned to the intervention or usual care. Those randomized to the intervention will use the Genetics Navigator in the pre-test period before meeting with their genetics clinician, and either before, during, or after results disclosure by their genetics clinician. Participants randomized to usual care will meet with their genetics clinician in the pre-test period and for results disclosure. The primary outcome is patient/parent distress 2 weeks after results disclosure. Secondary outcomes include knowledge, decisional conflict, anxiety/depression, empowerment, quality of life, personal utility, satisfaction, acceptability, health and digital literacy, and mental health resource use. Quantitative outcomes will be measured with validated surveys and compared between study groups for patients and for parents using inferential statistics and regression models. A sub-set of participants will be interviewed to explore user experience, using interpretive description to analyze the data. We will also conduct a cost-effectiveness analysis to assess the incremental cost of the Genetics Navigator compared to usual care per QALY gained from the perspective of the public payer for the patient and parent groups. **Conclusion.** Rigorously-evaluated patient platforms that digitize genetic services are needed. This study will provide robust evidence on the effectiveness of the Genetics

Navigator, a patient-centred e-health navigation platform to support end-to-end genetics service delivery across pediatric and adult populations.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Session II

PB2425 The Genomic Medicine Action Plan (GMAP) is a scalable, EHR-integrated utility-messaging tool for patients and their non-genetics providers receiving clinical genomic screening results.

Authors:

R. Wildin, D. Gerrard, P. vanWoert, D. Leonard; Univ. of Vermont Hlth.Network, Burlington, VT

Genetic test laboratory reports do not communicate clinical use guidance consumable by all patients and their non-genetics providers, creating a barrier to efficient and effective application for preventive health. In our genomic population health screening test implementation, our clinical laboratory creates and includes a GMAP messaging document along with reference lab reports when placing them in the electronic health record (EHR) and mailing a printed copy for patients. The GMAP: - Contains clear and concise messages written for patients and their primary care providers - Describes specific action recommendations and key education messages simultaneously for each of zero to many reported pathogenic or likely pathogenic variants (variants of uncertain significance [VUS] are not reported). - Is evidence-based and implements pre-defined care pathways for health risk results - Is integrated in one EHR lab result with a searchable text version of the associated test reports and their PDF versions - Reiterates program-related opportunities for free genetic counseling, cascade testing, for low-cost partner testing, and encourages sharing with healthcare professionals and relatives. The GMAP also: - Is revised and reissued when a variant classification is changed by the testing laboratory, indexed for patient-record global text search in the EHR (since 2022), and built in Epic for each patient from reusable formatted text blocks stored as Epic smart phrases - Provides telephone and email contacts for help from genetics professionals - Is empirically associated with a very low uptake of offered free genetic counseling - Is associated with increased self-reported genomic practice competency among primary re providers ordering the test. In addition to the core GMAP template, examples of real cases posing special challenges to clear and concise messaging will be presented, along with adaptations made from 2019 through 2023. Challenging examples including complex genotypes (e.g. in CYP21A2), possibly mosaic variants, deleted genes representing hematologic neoplasm, and reclassification of variants from VUS (not reported) to likely pathogenic or pathogenic or vice versa. The real-world examples illustrate our scalable approach to maximize utility of genomic screening results and minimize patient and non-genetics provider confusion. The GMAP model provides a reference solution for optimizing the critical information transfer of actionable information for clinical and personal utility. This may benefit other programs implementing large gene-count adult population health screening programs in a clinical, EHR-reported environment.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Session III

PB2426 The G-Test: Genetic Testing Eligibility Screening Tool for Prostate Cancer

Authors:

E. Chen, Y. Moo, J. Hoo, D. Chen; Arbelos Genomics, Chino, CA

Precision medicine is vital in effective treatment and improved cancer patient outcomes. Multi-gene panel cancer testing has significantly transformed the clinical methodology for testing at-risk patients and their families with hereditary cancer. Among all cancer types, prostate cancer has been a prevalent malignancy affecting many men worldwide. Previous studies demonstrate around 15% of individuals diagnosed with advanced prostate cancer carry a germline pathogenic variant; the National Comprehensive Cancer Network (NCCN)Version: 1.2023 recommends prostate cancer patients undergo germline genetic testing when the result can potentially impact treatment strategies. Over the past decade, however, most prostate cancer patients have not received germline genetic testing due to various factors. One major challenge is that physicians face difficulties comprehending the risk assessment principle. We aimed to increase the accessibility of prostate cancer genetic tests for patients by simplifying the patient eligibility determination process. By studying the latest NCCN guidelines on genetics and molecular analysis principles, we developed an algorithmic screening tool by posing questions specific to cancer stage, patient family history, and tumor characteristics. Questionnaire answers are processed and stored in a backend database with patient testing qualifications accessible to physicians through web interface queries. To test our model, we performed a retrospective study on 30 model patients' records, twenty of which are eligible for cancer genetic testing and the remaining ten are not as determined by NCCN guidelines. We have five primary care physicians review the records of 30 model patients and let them determine each eligibility for genetic testing using the NCCN guidelines, then the designed screening tool. The results show an increase in identifying eligible patients from 11.2 out of 20 cases to 16.3 out of 20 cases. We have then introduced the screening tool in outpatient urology settings and will continue to evaluate the change in the genetic test ordering rate. Simplifying the patient eligibility determination process for physicians allows broader access to genetic tests, enabling a larger pool of qualified patients to undergo testing and potentially benefit from personalized treatment approaches.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Session I

PB2427 The impact of recurrent 16p12.2 deletions on health in general population and personalized management of undiagnosed adult CNV carriers in the Estonian Biobank.

Authors:

M. Nõukas[1], M. Palover[1], A. Reigo[1], H. Alavere[1], L. Leitsalu[1], M. Jõeloo[1], I. Xenarios[2,3], A. Reymond[4], N. Tõnisson[1,5], K. Männik[2]; [1]Estonian Biobank, Inst. of Genomics, Univ. of Tartu, Tartu, Estonia, [2]Hlth.2030 Genome Ctr., Fondation Campus Biotech, Geneva, Switzerland, [3]Lausanne Univ. Hosp., Geneva, Switzerland, [4]Ctr. for Integrative Genomics, Univ. of Lausanne, Lausanne, Switzerland, [5]Dept. of Clinical Genetics in Tallinn, Tartu Univ. Hosp., Tartu, Estonia

The role of recurrent 16p12.2 deletions has previously been described in children with neurodevelopmental and psychiatric problems, but extensive phenotypic heterogeneity is also known. Here, we further plan to expand the knowledge on how: i) the recurrent 16p12.2 deletions affect health in the adult population and ii) to disclose these CNV findings to individuals at high genetic risk for complex health problems.

We detected 106 carriers (ages 22-83) of 16p12.2 deletions in the Estonian Biobank (EstBB; n=195,904) and mapped their disease traits using linked electronic health registries (EHRs). We further collect in-depth phenotype information from 16p12.2 deletion carriers through recall visits, return genetic risk findings, provide counseling tailored to their presented health conditions and evaluate the impact of disclosed genetic finding on their health and social support.

According to EHRs, depression (35%; 37/106) and anxiety disorders (22%; 17/106) were the most common neuropsychiatric diagnosis in adults with 16p12.2 deletion. Notably, mental and behavioural disorders due to use of alcohol (14%; 15/106) were frequently detected. Only 1 out of 106 carriers of recurrent 16p12.2 deletion was aware of their genetic diagnosis. Amongst the carriers of 16p12.2 deletion recontacted so far, 24 EstBB participants have consented for disclosure.

The study adds a population view to the knowledge of the understudied relationship between the recurrent 16p12.2 deletions, neuropsychiatric and other phenotypic traits in adults. Although not yet considered actionable in current guidelines, recall visits may relieve the anxiety in symptomatic CNV carriers.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Session II

PB2428 The International Consortium on Newborn Sequencing (ICoNS)

Authors:

N. Encina[1,2,3], A. Brower[4], A. Ferlini[5], H. Peay[6], L. Downie[7], D. Bailey[8], S. Kingsmore[9], W. Chung[10], R. Scott[11], M. Wasserstein[12], D. Bick[13], R. Green[14]; [1]The Intl. Consortium on Newborn Sequencing (ICoNS), Boston, MA, [2]Harvard T.H. Chan Sch. of Publ. Hlth., Boston, MA, [3]Ariadne Labs, Boston, MA, [4]American Coll. of Med. Genetics and Genomics, Dakota Dunes, SD, [5]Univ. OF FERRARA, Ferrara, Italy, [6]RTI interNatl., Durham, NC, [7]Victorian Clinical Genetics Services, Melbourne, Australia, [8]RTI Intl., Research Triangle Park, NC, [9]Rady Children s Inst. for Genomic Med., Coronado, CA, [10]Columbia Univ., New York, NY, [11]Genomics England, London, No, United Kingdom, [12]Albert Einstein Coll. of Med., Bronx, NY, [13]Genomics England Ltd., Gig Harbor, WA, [14]Harvard Med. Sch., Boston, MA

CONSORTIUM NAME: International Consortium on Newborn Sequencing (ICoNS)
BACKGROUND: Newborn sequencing (NBSeq) has the potential to offer a lifetime of personalized health care and disease prevention that is specific to each individual genome. When fully realized, NBSeq will mark a disruptive transition into personalized medicine and public health. A number of research projects and commercial offerings are already underway to explore the implementation of newborn sequencing in clinical settings.
OBJECTIVES: Given the proliferation of clinical implementation efforts, we sought to create the first organization specifically dedicated to communicating and sharing progress and best practices in the implementation of NBSeq. Initial topics for consideration included gene selection for initial panels, variant interpretation, patient communication and planning for clinical follow up.
METHODS: Principal investigators from 8 separate groups that were already conducting or imminently planning to conduct NBSeq research began regular meetings and agreed to establish the International Consortium on Newborn Sequencing (ICoNS). ICoNS was created to harmonize activities leading to evidence-based best practices for implementing newborn sequencing.
RESULTS: We have established ICoNS as an alliance of active investigators in NBSeq research with an annual interdisciplinary conference. An initial conference featuring 19 presentations was held in October 2022 in Boston, MA with over 300 attendees from 35 countries, and the next conference is planned for October 5-6, 2023 in London. Discussions of organization and governance are underway.
CONCLUSION: ICoNS presents an opportunity to gather multi-disciplinary governmental, academic and industry stakeholders and experts from around the world in order to accelerate and harmonize research progress and real-world implementation in NBSeq.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Session III

PB2429 The landscape of clinical sequencing in diverse populations: lessons learned from the Clinical Sequencing Evidence-generating Research (CSER) consortium.

Authors:

Z. Bly[1], L. Aviles-Santa[2], L. Biesecker[3], G. Cooper[4], S. Fullerton[5], B. Gelb[6], C. Horowitz[6], G. Jarvik[7], D. Kaufman[8], M. Norton[9], S. Plon[10], B. Powell[11], C. Sillari[1], A. Slavotinek[12], P. Tarczy-Hornoch[13], B. Wilfond[14], L. Hindorff[1], B. Korf[15]; [1]NIH/NHGRI, Bethesda, MD, [2]Natl. Inst. of Minority Hlth.and Hlth.Disparities, Bethesda, MD, [3]NIH NHGRI, Bethesda, MD, [4]Hudson Alpha Inst. for Biotechnology, Huntsville, AL, [5]Univ. of Washington, Seattle, WA, [6]Icahn Sch. of Med. at Mount Sinai, New York, NY, [7]Univ Washington Med Ctr., Seattle, WA, [8]NIH - NHGRI, Chevy Chase, MD, [9]Univ California San Francisco, San Francisco, CA, [10]Baylor Coll. Med., Houston, TX, [11]Univ North Carolina at Chapel Hill, Chapel Hill, NC, [12]Cincinnati Children`s Hosp. Med. Ctr., Cincinnati, OH, [13]Univ of Washington, Seattle, WA, [14]Seattle Children's Res. Inst., Seattle, WA, [15]Univ Alabama at Birmingham, Birmingham, AL

Background: Clinical Sequencing Evidence-Generating Research (CSER) was a national consortium funded by the National Institutes of Health committed to researching and supporting genomic sequencing in a clinical context. Since Phase 1 of CSER began in 2009, over 500 articles have been published by CSER investigators. The second phase of CSER (2017-2023) focused on generating evidence for clinical utility of genome sequencing. Seven clinical sites and a Coordinating Center, comprising 19 sub-sites across 11 states in the United States, worked collaboratively to recruit over 5,200 participants (~75% from diverse populations). Publications were authored by sites and/or by 8 cross-CSER working groups with members from across these different sites. **Methods:** In 2020, CSER investigators developed a schema to categorize the publications with respect to the clinical workflow and the corresponding roles of patients, providers, and laboratories. CSER papers were categorized into 1 of 7 "boxes" - e.g. Assessment and Referral, Sequencing and Analysis, or Return of Results and Counseling -- based on a list of over 40 relevant research questions. CSER papers published from 2017-2023 were categorized by three individuals, with each paper considered by two individuals, and discrepancies resolved by consensus. **Results:** Of the 209 papers evaluated 31% fell into the Sequencing and Analysis category and another 31% fell into the Return of Results and Counseling category. The categories with the fewest papers were case level papers in the Medical Problem (5%) or Interpretation in Light of Clinical Question (2%) boxes. The research questions most frequently addressed by CSER-related work are 1) how can reports and resources be communicated to participants of diverse backgrounds (n=48), 2) how can clinical sequencing data contribute to better understanding variant function (n=48), and 3) what is the impact of a diagnosis on subsequent care (n=47)? This reflects the multidisciplinary scope and goals of CSER as a program focused on clinical sequencing in diverse populations as well as ethical, legal and social (ELSI) matters. **Conclusions:** The categories and themes in the schema analysis highlight CSER's dedication to improving clinical sequencing at all stages of the patient, provider and laboratory experience, improving outcomes for diverse populations, and engaging with relevant stakeholders. The data generated also allow for identification of potential gaps in the consortium's research and sharing of lessons learned with the broader community.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Session I

PB2430 The PRAGMatIQ study: Implementation of rapid genome sequencing for hospitalized children in a publicly funded health care system

Authors:

C. Varin-Tremblay[1], G. D'amours[1], J. M. Gauthier[1], S. Gravel[2], S. Lévesque[2], N. Anastasio[2], J-F. Soucy[1], F. F. Hamdan[1], A-M. Laberge[1], J. L. Michaud[1]; [1]CHU Sainte-Justine, Montreal, QC, Canada, [2]CHU Sherbrooke, Sherbrooke, QC, Canada

Background: The use of rapid genome sequencing (rGS) in ill children with a suspicion of a monogenic condition has shown favorable results with regards to diagnostic yield, clinical utility, and reduced cost of care across multiple studies. Data from the use of this technology in publicly funded health care systems are scarce.
Objectives: Pediatric RApid GenoMics in Quebec (PRAGMatIQ) is a multi-center study that aims to assess, in real-world clinical practice, the medical, behavioral and economic impacts of rGS in inpatient children in the province of Quebec.
Design & Methods: Children admitted to an inpatient unit, including pediatric wards and pediatric and neonatal intensive care units, in one of the four pediatric tertiary hospitals of Quebec with an unexplained clinical presentation and a suspicion of a monogenic condition are eligible to participate in the study. Over a period of three years, 750 trios will be enrolled by medical geneticists and will receive rGS. Medical chart review and questionnaires for patients and clinicians will be used to assess the medical, behavioral and economic impacts of rGS.
Results: A pilot study involving 82 families was completed in one of the centers between January 2019 and December 2020. The PRAGMatIQ study, which started in November 2022, has enrolled 83 patients and their parents as of May 2023. Preliminary results show that rGS is beneficial for ill children across all pediatric units, age range and clinical presentations. So far, rGS has led to a molecular diagnosis in 40.2% of patients and impacted the clinical management in 31% of patients. There is evidence that negative results can also be informative to clinicians and have an impact on patient management. The median turnaround time of 20 days from patient enrollment to report of results is exposing the challenges of Quebec's health care system in implementing rGS. Additional data will become available in the upcoming months.
Conclusions: This province-wide study demonstrates how rGS can be implemented in a publicly funded healthcare system and can benefit ill children and their families. Variations in turnaround times highlight real-world challenges. Different strategies are being considered to reduce these delays.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Session II

PB2431 The role of multidisciplinary genomic teams in precision medicine: an implementation science guided systematic review

Authors:

A. Ma[1,2], R. O'Shea[2], L. Wedd[3,1], C. Wong[1], R. Jamieson[2,1,3], N. Rankin[4,2]; [1]Sydney Children's Hosp. Network, Sydney, Australia, [2]Univ. of Sydney, Sydney, Australia, [3]Children's Med. Res. Inst., Sydney, Australia, [4]Univ. of Melbourne, Melbourne, Australia

Introduction

Genomics and precision medicine hold promise for enhancing healthcare, especially in the era of advanced therapeutics. However, there is a great need for an implementation science based approach, that takes into account the health services, practitioner and patient needs and barriers, before the benefits of precision healthcare can be accessed by all.

Due to the increasing complexity of genomic data interpretation, and need for close collaboration with clinical, laboratory, and research expertise, genomics often requires a multidisciplinary approach. This systematic review aims to establish the evidence for effectiveness of the genomic multidisciplinary team, and the implementation components of this model that can inform precision care.

Methods

Databases were searched for papers examining the genomic multidisciplinary approach for diagnosis and management, and implementation outcomes regarding effectiveness, adoption, efficiency, safety, and acceptability (Proctor et al.). A narrative synthesis of these outcomes and themes were mapped against the Genomic Medicine Integrative Research (GMIR Horowitz et al.) framework to inform practice, and identify evidence-practice gaps in the literature.

Results

A total of 1039 studies were screened, and 17 papers met selection criteria. All studies pointed towards the effectiveness of the genomic MDT approach, with 10-78% diagnostic yield depending on clinical context, and an increased yield of 6-25% attributed to the MDT approach itself, due to improved expert clinical-laboratory collaboration. The genomic MDT was also found to be highly efficient, especially in interpretation of variants of uncertain significance, timeliness for a rapid result, made a significant impact on management, and was acceptable for adoption by a wide variety of subspecialists. Only one study utilized an implementation science based approach to the genomic MDT.

Conclusions

The genomic MDT approach appears to be highly effective and efficient, facilitating higher diagnostic rates and improved patient management. However, key gaps remain in health systems readiness for this collaborative model, and there is a lack of implementation science based research especially addressing the adaptability, scalability and sustainability of this model.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Session III

PB2432 † The Sickle Cell Women and Girls (S.W.A.G) Project of Accra, Ghana: An Evidence-Based Study at the Ghana Institute of Clinical Genetics

Authors:

S. THOMAS[1], A. Asare[2], R. Attah[3], L. Agyiri[2], J. Ayete-Nyampong[4], E. Anku[2], J. A. Osae-Larbi[5], C. R. Sudfeld[1], A. Benneh-Akwasi Kuma[2], E. V. Asare[2]; [1]Harvard Sch. of Publ. Hlth., Boston, MA, MA, [2]Ghana Inst. of Clinical Genetics, Accra, Ghana, [3]Univ. of Ghana, Accra, Ghana, [4]Korle Bu Teaching Hosp., Accra, Ghana, [5]West African Genetic Med. Ctr., Accra, Ghana

ABSTRACT The Sickle Women and Girls (S.W.A.G.) Project is a NIH-funded research initiative, with collaborators between the United States and Ghana. In this study, we are investigating the reproductive health experiences among women and girls (13-25 years of age) with sickle cell disease (SCD) in the Greater Accra Region. Women and girls of reproductive age living with SCD is highly prevalent in throughout Ghana, with one in three Ghanaians having either sickle cell trait or SCD, such as: HbAS, HbSS, or HbSC. Due to reproductive-related conditions, examining SCD, at the intersection of maternal-child health requires critical attention. This study utilized a qualitative research design, consisting of 42 semi-structured interviews at the Ghana Institute of Clinical Genetics. We sought to investigate how women and girls with SCD transition through adolescents, involving social and cultural stigmatizations, menstruation health, perceptions of contraception, and perceptions of testing partners for sickle cell trait. Data was collected by a team of six research assistants (who spoke both English and Twi local language). Data was analyzed using NVIVO 12 qualitative software. Results show two thematic patterns, including: 1) increased social and emotional isolation due to disease stigmatization; 2) and female adolescents' desiring increased knowledge and support for transitioning from pediatrics to adult care for SCD treatment and family planning. Overall, our study provides further evidence for intervention that calls researchers to bridge the gap between the clinical aspects of SCD, while also uplifting the social, emotional, and mental health needs of SCD across the African Diaspora.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Session I

PB2433 The social construction of genomics and blindness in Nigeria: The Eyes of Africa (EOA) study.

Authors:

O. Olawoye[1], K. Salami[2], T. Realini[3], M. Hauser[4], A. Ashaye[1]; [1]Dept. of Ophthalmology,Coll. of Med. Univ. of Ibadan, Nigeria, Ibadan, Nigeria, [2]Dept. of Sociology Univ. of Ibadan, Ibadan, Nigeria, [3]Dept. of Ophthalmology, West Virginia Univ., West Virginia, VA, [4]Duke Univ Med Ctr, Durham, NC

Background: Genomics is an emerging field that has great potential to improve our understanding of ocular diseases and public health practice. **Purpose**: This study explored the social construction of genomics and blindness in two different communities in Nigeria. **Methodology**: This qualitative study utilized in-depth interviews (IDIs) to generate data from both members of the Nigeria Association of the Blind (NAB) and members of the Gbedu community in Oyo State of Nigeria. It utilized a multiple descriptive interview design to depict and understand research participants' views about genomics and blindness. The design also enhanced an in-depth understanding of participants' perception and interpretation of genomics and blindness in their context uncovering the meaning they give to their perception of a social phenomenon. Eighteen sighted and 12 blind participants were purposively interviewed in the community. Audio recordings of the IDI sessions were transcribed and labelled appropriately. Data were reviewed and edited to ensure proper interpretation and construction of accurate meaning and then processed using the NVIVO version 12 Pro. Both deductive and inductive approaches were employed. **Results:** Our study revealed some level of knowledge regarding the hereditary basis of some ocular diseases. However, this was tempered by many fallacious beliefs such as maternal versus paternal contributions to inheritance, and supernatural/religious underpinnings among both the blind and sighted in the community. Both groups of participants preferred saliva-based sample collection over blood-based. Although the participants (blind and sighted) were willing to participate in genomics testing for ocular diseases, they expressed concern for the individual and community benefits in both the short and long term. Participants expected a return of individual results if they participated in a genomic study. There was also a disconnect between the perceived causes of blindness (processed food, use of spectacles in early life, supernatural fallacies) and evidence-based causes of blindness. **Conclusion**: Knowledge limitations, fallacies, and participant expectations influenced the attitude of participants towards genomics in ocular sciences; therefore, this should be addressed with appropriate community engagement programs. It is important that participants' opinions about anonymity are respected to engender greater support and participation in genomics research. The development of actionable genomics procedures in ocular diseases should be sensitive to the needs and preferences of the study participants.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Session III

PB2435 Towards Representative Genomic Research: The Children's Rare Disease Cohorts Experience

Authors:

Z. Frazier, E. Brown, S. Rockowitz, T. Lee, B. Zhang, A. Sveden, N. Chamberlin, K. Dies, A. Poduri, P. Sliz, M. Chopra; Boston Children's Hosp., Boston, MA

Background: Due to racial, cultural and linguistic marginalization, some populations experience disproportionate barriers to genetic testing in both clinical and research settings. It is difficult to track such disparities due to non-inclusive self-reported race and ethnicity categories within the electronic health record (EHR). Inclusion and access for all populations is critical to achieve health equity and to capture the full spectrum of rare genetic disease. Objective: We aimed to create revised race and ethnicity categories. Additionally, we identified racial and ethnic under-representation amongst three cohorts: (1) the general Boston Children's Hospital patient population (general BCH) (n=3,067,921), (2) the BCH patient population that underwent clinical genomic testing (clinical sequencing) (n=1,791), and (3) Children's Rare Disease Cohort research initiative (CRDC) participants (n=3,627). Design and Methods: Race and ethnicity data was collected from the EHRs of the general BCH, clinical sequencing, and CRDC cohorts. We constructed a single comprehensive set of race and ethnicity categories. EHR-based race and ethnicity variables were mapped within each cohort to the revised categories. Then, the numbers of patients within each revised race and ethnicity category were compared across cohorts.
Results: There was a significantly lower percentage of Black or African American /African, non-Hispanic/non-Latine individuals in the CRDC cohort compared with the general BCH cohort (4.60% vs 9.72%, $p < 0.0001$), but there was no statistically significant difference between the CRDC and the clinical sequencing cohorts (4.60% vs 5.82%, $p = 0.1794$). There was a significantly lower percentage of multi-racial, Hispanic/Latine individuals in the CRDC cohort compared to the clinical sequencing cohort (7.36% vs 11.52%, $p < 0.0001$). White, non-Hispanic/non-Latine individuals were over-represented in the CRDC compared to the general BCH cohort (71.88% vs 66.80%, $p < 0.0001$) and the clinical sequencing cohort (71.88% vs 65.96%, $p = 0.0002$).
Conclusion: We highlight underrepresentation of certain racial and ethnic populations in sequencing cohorts compared to the general hospital population. We propose a range of measures to address these disparities, to strive for equitable future precision medicine-based clinical care and for the benefit of the whole rare disease community.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Session I

PB2436 Transforming Genomics Research through Community Engagement and Return of Results: A Case Study from French Polynesia

Authors:

H. Babiker[1], K. Wasik[1], S. LeBaron von Baeyer[1], K. Fox[2], P. Tristan[3], T. Mairai[1], V. Wheeler[4]; [1]Variant Bio, Inc., Seattle, WA, [2]Indigenous Futures Inst., Univ. of California San Diego, La Jolla, CA, [3]Lille Catholic Univ., Lille, France, [4]Australian Natl. Univ., Canberra, Australia

Many communities around the world are wary of genomics research, and for good reasons. All too often, researchers have failed to address local priorities and power inequities when designing and executing these studies. What is more, findings are rarely communicated to participating communities. Here, we highlight an approach that seeks to address colonial and ongoing legacies of extractive research by performing in-depth community engagement, respecting local cultural protocols, and sharing meaningful benefits as well as population-level genomic and health results first and foremost with participating communities. The study was conducted in French Polynesia where ~1100 individuals participated in a sample and data collection campaign spanning genomics, transcriptomics, metabolomics, and extensive biomedical health evaluation. In the course of the study we made multiple findings with significant implications for public health in French Polynesia. For instance we discovered that more than a quarter of the adult population of French Polynesia is estimated to have gout (compared to 0.9% in metropolitan France and 3.9% in the US). We will discuss how appropriate community engagement enabled us to conduct the study and gather data, and how these results are currently shaping the discourse around treatment of gout and public health across the Polynesian archipelago.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Session II

PB2437 Understanding ableism in abstracts for genetic and genomic autism research

Authors:

N. Sugden[1], J. Kee[2], C. Gunter[1]; [1]Natl. Human Genome Res. Inst., Bethesda, MD, [2]Thomas Jefferson Univ., Philadelphia, MD

In recent years, genetic and genomic autism research has come under increasing scrutiny, moving to the center of debates about autism acceptance and the future of research and care. Psychiatric and neurodevelopmental genetic research is especially ethically fraught because of its historical association with eugenics, leading to current-day concerns over its use to guide reproductive decision-making via family histories and prenatal testing or screening for a number of neurodevelopmental conditions. Many critiques of genetic and genomic autism research hinge on the idea that such research is often ableist: that it values abled over disabled and nonautistic over autistic life, and assumes that autistic children and adults require interventions in order to meet societal expectations rather than vice versa. The language of research is an indicator of its evolving priorities and ethics. Research documentation, including abstracts in grant applications, are subject to both formal requirements and explicit and implicit norms. Prospective and ongoing research is a space of uncertainty, promise, and persuasion, where prospective grantees feel the requirement to "sell their science." For these reasons, the language of research cannot be taken to straightforwardly correspond to individual researchers' beliefs and behavior. Nevertheless, the presence of widespread or egregious ableist language may indicate structural and institutionalized ableism - and evaluating the language of research can prompt researchers and research institutions to reflect on the way they discuss and pursue their work and lead to changes that promote greater inclusion and equity. We conducted a mixed-methods analysis of 541 abstracts associated with NIH-funded genetic and genomic autism research since 1980, to reveal the prevalence of ableist language and discourse about autistic people and to identify when they are more likely to be used. In our pilot study of 100 abstracts, 66% contained ableist language or discourse, including ideas about deficits and risk as well as the use of the outdated term "mental retardation." Analysis of the full sample is revealing more detailed patterns over time and between types of research and funding.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Session III

PB2438 Understanding parental/caregiver perspectives during pediatric genome-wide sequencing: A scoping literature review

Authors:

P. Murali[1], J-H. Yu[2]; [1]Univ. of Washington, Seattle, WA, [2]Univ. of Washington Sch. of Med., Seattle, WA

Background: Genome-wide sequencing (GWS) has changed the landscape of pediatric genetics, enabling accurate risk assessment and targeted treatments. However, the needs and perspectives, of parents/caregivers of pediatric patients preceding, during, and following GWS remain poorly understood. An examination of these needs is critical to providing informed care that incorporates parent/caregiver values, addresses their concerns, and fosters collaborative decision-making in the context of pediatric GWS. Therefore, we conducted a scoping literature review to characterize parent and caregiver perspectives and needs in pediatric GWS. Methods: Based on scoping review methods by Arksey and O'Malley, a literature search was conducted across PubMed, Cinahl, Embase, Psycinfo, and Web of Science for English-language primary research articles reporting on the experiences, perspectives, and needs of parents/caregivers of pediatric patients who have undergone GWS. Results: An initial search yielded 574 articles of which 49 met inclusion criteria. Preliminary findings focused on impact of diagnosis, test features, support needs and engagement, and ethical considerations. Regarding diagnostic impact, parents expressed positive attitude towards results of testing regardless of positive, negative, or uncertain findings. An understanding of etiology provided closure and a sense of relief. Regarding test features, parents expressed a need for clear and understandable information following treatment and intervention following GWS. Some expressed disappointment when GWS results did not result in treatment changes. Parents emphasized the importance of support from healthcare professionals and peer networks to cope with the challenges associated with GWS. Finally, parents faced ethical dilemmas related to incidental findings as well as potential implications for other family members. Discussion: This scoping review highlights the multifaceted nature of these experiences for patients based on the type of result, the medical specialty in which GWS was conducted and the severity of disease in the impacted child, among others. These experiences inform parent/caregiver needs as they navigate the process of GWS diagnostic testing that should be better understood by clinicians, genetics counselors, and policymakers to ensure appropriate support and guidance. Furthermore, these knowledge gaps emphasize the need for future research to explore parental perspectives in diverse populations, the long-term impact of GWS on families, and the effectiveness of interventions aimed at supporting parents and caregivers throughout the diagnostic journey.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Session I

PB2439 Usability of a pre-test genetic counseling video for familial hypercholesterolemia as part of RESEQUENCE-GC.

Authors:

E. Brown[1], C. Pendleton[1], C. Applegate[1], C. Tichnell[2], B. Murray[1], B. Rivers[1], R. McClellan[1], A. Barth[1], L. Yanek[2], C. Taylor[1], J. Day[1], C. James[3]; [1]Johns Hopkins Univ., Baltimore, MD, [2]Johns Hopkins Univ, Baltimore, MD, [3]Johns Hopkins Hosp, Baltimore, MD

Familial hypercholesterolemia (FH) is an autosomal semi-dominant condition characterized by high LDL-C levels and an increased risk for coronary artery disease (CAD). If untreated, risk of early onset CAD is significantly elevated. Therefore, the Centers for Disease Control and Prevention (CDC) classifies FH as a Tier 1 condition for genomic application. Multiple cardiac guidelines also recommend patients receive genetic counseling in addition to genetic testing. Despite these recommendations, access to genetic counselors may be limited due to increased need for interpretation of genetic testing results in the setting of ongoing workforce shortage. RESQUENCE-GC is an NHGRI-funded trial to evaluate the utility of a pre-test genetic counseling video instead of the traditional genetic counseling visit providing pre-test education for patients with a personal or family history of hereditary cardiac conditions including FH. Here we describe the development and assessment of the FH videos. Two 8-minute educational animated videos were developed using principles from the Cognitive Theory of Multimedia Learning for the FH panel and targeted familial variant testing. Videos were created using Maxon Cinema4D, Adobe Creative Cloud, and TechSmith Camtasia Studio. Flesch-Kincaid Grade Level was 8.9 and 9.4 for the videos. For evaluation, the videos were sent to both the general population recruited via a usability testing company (UserHappy) and members of a patient support group (Family Heart Foundation). Participants were emailed a questionnaire that included 8 true/false pre-video knowledge questions, one of the two videos, free response feedback, engagement questions [User Engagement Scale; scores for each question range from 1-5, where 5 indicates highest agreement] and the same 8 true/false questions for post-video assessment. In total, 78 individuals participated from the patient support group (N=24) and the general population (N=54). Participant knowledge increased for both videos. Participants answered 84% of questions correctly pre-test which increased to 89% post-test (mean change score 0.043±0.13, p=0.002). In terms of user engagement, on average for both videos individuals indicated they found the video worthwhile (4.3±0.8) and would feel comfortable making a genetic testing decision post-video (4.3±0.8). These data suggest a pre-test genetic counseling video visit may be a suitable alternative for cases where a traditional pre-test genetic counseling visit is challenging. This option may help increase accessibility to genetic counselors and decrease wait-times with an end result of more people gaining access to genetics services.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Session II

PB2440 Using computational phenotype trajectories to inform a holistic model of care for rare diseases

Authors:

T. Groza[1], K. Stevens[1], D. Gration[2], K. Togher[2], S. O'Neil[3], J. Rogers[4], S. Jamuar[5,6], W. Lim[6,7,8], G. Baynam[1]; [1]Rare Care Ctr., Perth Children's Hosp., Nedlands, Australia, [2]Western Australian Register of Dev.al Anomalies, Subiaco, Australia, [3]Child and Adolescent Hlth.Services, Nedlands, Australia, [4]Linear, Nedlands, Australia, [5]KK Women s and Children s Hosp., Singapore, Singapore, [6]SingHlth.Duke-NUS Inst. of Precision Med., Singapore, Singapore, [7]Genome Inst., Singapore, Singapore, [8]SingHlth.Duke-NUS Genomic Med. Ctr., Singapore, Singapore

Rare diseases (RD) are a worldwide health care challenge, collectively affecting at least 1 in 16 people. Patients and families living with suspected RDs often spend more than five years on a diagnostic odyssey. While advances in genomic sequencing coupled with the adoption of standardised terminologies and an increasing willingness to share data internationally have enabled progress in this field, for many patients the effort will ultimately be futile; more than 50% of RD patients do not receive a molecular diagnosis. The management of the phenotypic presentation of the condition thus becomes critical for the patient and family well-being. Devising care pathways, for both diagnosed and undiagnosed patients - covering prognosis, surveillance, coordinated clinical care or family planning - is essential to support emotional relief and appropriate access to resources. This is the primary aim of the Rare Care Centre, which has developed a transformative, community co designed cross-sector Model of Care, to provide holistic care and support in an integrated and responsive service.

The Model of Care is underpinned by computational models of disease that capture the various dependency relationships governing phenotypes and produce phenotype trajectories that map a more accurate view over prognosis and functional consequences. For example, an unmanaged *hydrocephalus* leads to *cortical damage*, which could lead to *blindness* and hence impact *education*. Having this relationship externalised and interpreted in a given temporal context can lead to an appropriately informed care plan - e.g., by performing cerebrospinal fluid shunting - known to improve the hydrocephalus - as well as planning for specific education programs.

Lowering the granularity of the currently available computationally-tractable knowledge and shifting the focus to phenotypes opens new avenues for studying specific genotype-phenotype relationships, including the underpinning biological processes that are disrupted. More importantly, coding phenotype trajectories and their insights into the specific time points that trigger management events enable the Rare Care team to proactively intervene and deliver a more personalised care plan, for both diagnosed and undiagnosed patients.

Our framework for RD management - currently developed for lysosomal storage disorders - covers a deeper understanding of the relationships between phenotypes and an encoding of actionable knowledge spanning across the multidimensional needs that impact not only the patients' physical and mental health, but every aspect of their lives, including education, finances, social activities and employment.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Session III

PB2441 Using e-Technology Among Populations of Advanced Age to Investigate Genetic Mechanisms of Exceptional Longevity in the SuperAgers Family Study

Authors:

A. Montgomery[1], S. Milman[2], S. Shokouhi[1], N. Barzilai[3], P. Jain[1], A. Shuldiner[4]; [1]Vibrent Hlth., Fairfax, VA, [2]Albert Einstein Coll. of Med., Bronx, NY, [3]Einstein Coll. of Med., Bronx, NY, [4]Regeneron, Tarrytown, NY

Background/Objective: Although 95% of older adults in the US have at least one chronic condition, individuals who live past the age of 95 demonstrate delayed onset of age-associated disease. Previous studies show that healthy aging is strongly influenced by genetics; however, the genetic basis for exceptional longevity and healthspan remains primarily unknown. The SuperAgers Family study aims to investigate genetic, phenotypic, and biological mechanisms related to healthy aging and exceptional longevity. The current pilot phase will determine feasibility and identify best practices in using e-technology to conduct nationwide, remote health research among individuals of advanced age.
Methods: SuperAgers aged 95 or older, their offspring, and offspring's spouses are invited to join the study via e-mail or mail. All participants use a web-based platform via computer or mobile device to enroll, confirm age, and complete study surveys. Manual linkage of SuperAgers and offspring profiles initiates receipt of a saliva DNA kit by mail. Participants collect samples at home and return them by mail. Research staff provide support as needed by phone or via secure computer-assisted telephone interview. Qualitative process evaluation was conducted to evaluate the study workflow.
Results: Preliminary data from SuperAgers (n = 118) and offspring/spouses (n = 110) were included in process evaluation. Collectively, participants had an average age of 84 years (± 15 SD), 66.23 % were female, and 96.05 % were non-Hispanic White. During eligibility screening, 3.1% of participants misidentified themselves as either SuperAger, offspring, or spouse, leading to incorrect group assignments. Further, 35.9% of participants entered incorrect SuperAger identifying information causing failure of the family member linkage process. Workflow adjustments were made to prevent recurrence of identified errors.
Discussion: Process evaluation of the SuperAgers Family Study pilot revealed participant entry errors occurring during eligibility screening and family member linkage processes. As a result, workflow safeguards were successfully implemented to prevent repeat errors. These findings may inform development of best practices for conducting digitally delivered research studies in aging populations.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Session I

PB2442 Values-Based Policymaking in Prenatal Human Gene Editing

Authors:

R. Isasi[1], K. Riggan[2], Y. Bombard[3], N. Bonhomme[4], K. Meagher[5], M. Michie[6], K. Musunuru[7], K. Ormond[8], S. Rubeck[9], R. Feys[1], M. Allyse[2]; [1]Univ. of Miami Miller Sch. of Med., Miami, FL, [2]Mayo Clinic, Rochester, MN, [3]Univ of Toronto, Toronto, ON, Canada, [4]Genetic Alliance, Washington, DC, [5]BioMed. Ethics Res. Program, Mayo Clinic, Rochester, MN, [6]Case Western Reserve Univ. Sch. of Med., Cleveland, OH, [7]Univ. of Pennsylvania, Philadelphia, PA, [8]Stanford Univ and ETH-Zurich, Zurich, Switzerland, [9]Dept. of Bioethics, Case Western Reserve Univ. Sch. of Med., Cleveland, OH

Human gene editing technologies in both the research and clinical contexts continue to progress rapidly. Recent experiments in translation in reproductive settings have emphasized the need for more values based and patient-centered governance mechanisms for translation of prenatal gene editing. Several national and international bodies have called for such governance; however, they have been technocratically framed around moratoria or other forms of suppression. We propose a values-based governance approach that moves beyond traditional technocratic considerations of safety and efficacy and considers collective normative deliberation about the ethics of prenatal intervention. Importantly, we call for special consideration to emerging translational justice challenges. Based on international policy comparative study combining soft (e.g., professional guidelines) and hard laws (e.g., statutes), we will analyze normative frameworks to assess underlying normative underpinnings of proposed policy approaches, professional feasibility of proposals, and values and priorities around the potential translation of prenatal gene editing in humans. We further outline a framework for incorporating translational justice considerations within governance structures for gene editing in both research and clinical settings.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Session II

PB2443 Voluntary workplace genetic testing in a large healthcare system: Does testing prompt employee health behavior changes or use of health services?

Authors:

E. Charnysh[1], A. Vogle[1], W. Uhlmann[2], A. Prince[3], R. Ferber[4], S. McCain[4], K. Sanghavi[1], W. Feero[5], C. Lee[1], J. Roberts[6]; [1]The Jackson Lab. for Genomic Med., Farmington, CT, [2]Univ. of Michigan Sch. of Med., Ann Arbor, MI, [3]Univ. of Iowa, Iowa City, IA, [4]Univ. of Michigan, Ann Arbor, MI, [5]Maine-Dartmouth Family Med. Residency, Augusta, ME, [6]Univ. of Michigan Sch. of Publ. Hlth., Ann Arbor, MI

Some employers are offering voluntary genetic testing through workplace wellness programs, known as workplace genetic testing (wGT). Little is known about the perspectives of employees offered wGT or its potential impact on their health behaviors and use of health services. From Dec. 2021 to May 2022, we conducted a web survey (n=776) of employees at a large healthcare system who were offered wGT through a third-party vendor that included genetic tests for hereditary cancer and cardiovascular disease risk as well as medication response (pharmacogenomics - PGx). This study reports on findings from the 418 respondents (53.9%) who reported receiving results (female, 88.3%; white, 83.0%; college or graduate degree, 80.6%). Across these 418 respondents, 211 unique positive results were reported among 159/418 (38.0%) individuals. Specifically, 125/418 (29.9%) of these respondents reported receiving at least one PGx result, 48/418 (11.5%) reported at least one result indicating an increased risk of cancer, and 38/418 (9.1%) reported at least one result indicating an increased risk of heart disease. Notedly, 39/418 (9.3%) reported receiving more than one positive finding among the three categories. Regarding health behavior changes following wGT, 29/159 (18.2%) individuals with positive (pos) results reported making changes to their diet, and 27/159 (17.0%) reported making changes to their physical activity/exercise. This was comparable to those with negative/unknown (neg/unk) results: 41/259 (15.8%) reported diet changes and 43/259 (16.6%) reported physical activity changes. Respondents with both result types reported using health services or making plans for healthcare utilization as a result of learning their test results, including making an appointment with a provider (pos all: n=32/159 [20.1%]; neg/unk all: n=49/259 [18.9%]), cancer screening (pos cancer: n=10/48 [20.8%]; neg/unk for cancer: n=101/370 [27.3%]), surgeries/cancer prevention procedures (pos cancer: n=6/48 [12.5%]; neg/unk for cancer: n=55/370 [14.9%]), heart disease screening (pos cardio: n=10/38 [26.3%]; neg/unk for cardio: n=97/380 [25.5%]), and making changes to use of medications (pos PGx: n=21/125 [16.8%]; neg/unk for PGx: n=44/293 [15.0%]). Findings suggest undergoing wGT prompted a range of health behavior changes, regardless of test results, in a subset of employees, as well as the use of health services related to the prevention of cancer and heart disease. Future planned analyses within this dataset will seek to identify factors influencing behavioral responses to wGT. This is the first study to report on behavior and healthcare service utilization outcomes for wGT.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Session III

PB2444 What are patient perspectives on privacy and trust in digital genomic tools? A qualitative study

Authors:

V. Jha[1], S. Saeedi[2], M. Clausen[3], D. Assamad[4], S. Grewal[5], D. Hirjikaka[5], W. Lee[4], S. Luca[4], A. Shaw[6], R. Hayeems[4], Y. Bombard[7], on behalf of the Genetics Navigator Study Team; [1]Univ. of Toronto, Toronto, ON, Canada, [2]McGill Univ., Montreal, QC, Canada, [3]St. Michael s Hosp., Toronto, ON, Canada, [4]The Hosp. for Sick Children, Toronto, ON, Canada, [5]St. Michael's Hosp., Toronto, ON, Canada, [6]Unity Hlth., St. Michaels Hosp., Toronto, ON, Canada, [7]Univ of Toronto, Toronto, ON, Canada

Background: Digital tools have emerged as a promising solution to increase efficiency and capacity of genomic services. Patient facing digital tools can improve access, decrease wait times, and boost patient engagement. However, along with digital healthcare comes a radical change in the storage of personal health information (PHI) which can raise concern about privacy and security risks. As patient facing digital genomic tools continue to develop, it is important to understand and incorporate patients' perspectives on privacy and security.**Aim:** To assess patient perspectives on privacy and security in digital genomic tools.Methods: A qualitative study was conducted using semi-structured interviews with patients or parents of patients who underwent genetic testing. Participants were either recruited from the Canadian Organization for Rare Disorders or had previously participated in a Toronto-based cancer genetics study. Interviews focused on features that would make a digital genomics platform feel secure and trustworthy. Interview transcripts were analyzed using thematic analysis and interpretive description.Results: Thirty participants who previously received genetic testing for themselves (n = 17) or their child (n = 13) were interviewed (n=20 females, n=15 above 50 years old). Overall, participants were willing to store and access genomics personal health information (PHI) in a patient-facing digital platform. The main benefits of digital genomics services identified by participants were the ability to access and control their own PHI. Participants emphasized that a digital genomics platform should prioritize patient control of information to give patients more agency and increase efficiency in their care. Participants expressed that the benefits, such as patient empowerment and personalized care, outweighed the perceived risks, such as potential data leaks. Participants described factors that increased their risk tolerance, including the digitization of other sensitive matters such as banking information. In order to minimize risks, participants emphasized transparency about what security measures are in place and who has access to their PHI. Participants described this information as a prerequisite, which should be easily found on a consent form or a page on a digital platform.Conclusions: Patients are willing to access their genomic information digitally as long as security measures are clearly explained and patients are able to access and control their own information. These findings inform the design of digital genomic platforms to enhance patients' sense of security, which is critical for the uptake and usage of any platform.

Session Title: Genetic Counseling, ELSI, Education, and Health Services Research Poster Session I

PB2445 Where we stand on gene therapy: The current capabilities, the public opinion, and the policy.

Authors:

M. Svoboda[1], M. Dinulos[2]; [1]Geisel Sch. of Med. at Dartmouth, Hanover, NH, [2]Dartmouth Hitchcock Med Ctr, Lebanon, NH

The incessant advancements in genomic sequencing, nucleic acid editing, and gene delivery have recently produced the first approved gene therapies that are now available to the public. While these new technologies promise unprecedented opportunities in both biomedical research and clinical practice, they also present novel ethical dilemmas with practical implications for patient care, making us reconsider issues like decision-making delegation, the boundary between a “treatment” and an “enhancement,” and even species identity. The further we move the limits of our knowledge, the more frequently we encounter new limits that are of ethical nature instead: the question of “What can we do?” becomes “What should we do?”

Using the findings presented by studies in genetics available to date, we describe the extent of the technological advancements in gene therapies on the basic, translational, and clinical level. While there are currently several approved both in- and ex-vivo gene therapies, many more are undergoing clinical trials and yet even more are in the translational pipelines or basic research, promising an explosion of novel capabilities to tackle diseases with known genetic basis.

We then explore the public opinion of gene therapies and discover that in general, while somatic cell editing for patients with known serious pathologies is supported, embryonic and germline cell editing is met with mixed and mostly unfavorable views. The support further declines when interventions described as “enhancements,” as opposed to “treatments,” are considered.

Lastly, we turn to reconcile the technological advancements and public opinions with the current policy regarding gene therapies with specific focus on the most contentious aspect thereof, heritable germline editing, and we find the existing policy insufficient to govern these practices. To facilitate a fruitful discussion in anticipation of future scientific advances, we suggest a specific set of guidelines, under which heritable genome editing in humans may be theoretically permissible. We conclude that these standards have not yet been, and perhaps never may be, met and further discourse is needed before such therapies are approved.

Session Title: Evolutionary and Population Genetics Poster Session I

PB3001 A draft diploid reference pangenome construction from a distinct Arab population

Authors:

N. Nassir[1], M. Almarri[1], M. Kumail[1], N. Mohamed[1], m. AlObathani[1], B. Jamalalail[1], S. Hanif[1], h. Elsokary[1], D. Satsangi[1], H. Al Suwaidi[1], M. Hamdi[2], A. Abu Tayoun[1], A. Albanna[1], H. Khansaheb[2], S. Duplessis[1], A. Alsheikh-Ali[1], M. Uddin[3]; [1]Mohammed bin rashid Univ. of medicine and Hlth. Sci.s, Dubai, United Arab Emirates, [2]Dubai Hlth.Authority, Dubai, United Arab Emirates, [3]Mohammed Bin Rashid Univ. of Med., Dubai, United Arab Emirates

Current reference genomes do not adequately represent the ethnic diversity of humans, with near absent representation of Arab populations. The recent pangenome initiative identified a massive intra and inter-population genetic diversity from *de novo* genome assemblies. However, the pangenome consortium lacks representation from the Arab population. In this study, we sequenced 25 deeply phenotyped healthy adults from the United Arab Emirates (UAE) using Pacific Bioscience (70X High fidelity reads with N50=18.70 kb, median Q=32.8) and Oxford Nanopore (54X ultralong reads; N50=59.57 kb, median Q=16.4) yielding high-quality *de novo* assemblies covering 97.88% of CHM13 (QV66.78). Preliminary analysis from a trio suggests our assemblies are highly contiguous (NG50=146.01 Mb) compared to 40 Mb for current pangenomes, spanning 3.03, 2.97Gb for maternal and paternal haplotypes, respectively. The assembly exhibits low switch (0.015204, 0.00008) and Hamming (0.020121, 0.000055) error rates, likely reflecting high LOH stretches due to population consanguinity. Initial mapping to GRCh38 revealed a significant number of population-specific variants not reported in other ethnicities. Furthermore, we identified 4.65 Mb of new sequence (outside of hg38) from the trio, which may contain novel genes and regulatory elements. This represents the most contiguous diploid human genome to date, revealing extensive unreported structural variation and enabling high-quality haplotyping of clinically relevant alleles in this underrepresented population. Our study provides a foundational resource for future genetic research and precision medicine initiatives in the UAE and other populations with similar genetic backgrounds.

Session Title: Evolutionary and Population Genetics Poster Session II

PB3002 A genetic assessment of the likelihood of patrilocal mating in Neanderthals

Authors:

N. McAllister; Univ. of Michigan, Ann Arbor, MI

There is no current consensus on common mating practices in Neanderthals. Mating practices provide important information about the social organization of a species, making comparisons of Neanderthal and ancient *Homo sapiens* practices an important piece in understanding the differences in social behavior between us. The issue of patrilocality, or the movement of females into the male partner's household, has captured the interest of paleoanthropologists and resulted in a topic of debate. Evidence for patrilocality comes from a sample of 13 closely related Neanderthals at the remote eastern boundary of their range. It is unclear if the conclusion of patrilocality is generalizable to other times/regions of Neanderthal occupation. We study this question by inferring pairwise coalescent times in a set of 16 Neanderthal individuals from across the temporal and spatial range of their occupation. We account for the low (< 2x) coverage in these individuals by integrating over uncertainty in the genotype calls. As autosomes and sex chromosomes spend different proportions of coalescent history in males and females, differences in coalescent times between autosomes and sex chromosomes reflect different population histories of males and females. Under patrilocality, we assume that the migration rates of males are lower than the migration rates of females, resulting in differences in the time to most recent common ancestor of X-chromosomes and autosomes, even after accounting for differences in the chromosomes' effective population size. Our results can help indicate whether patrilocality was a common practice among Neanderthals outside of the remote eastern boundary region and whether or not the mating practices of Neanderthals are comparable to ancient *Homo sapiens*.

Session Title: Evolutionary and Population Genetics Poster Session III

PB3003 A likelihood-based framework for demographic inference from genealogical trees

Authors:

C. Chiang[1], J. L. Cahoon[1], B. L. Dinh[1], D. Ortega-Del Vecchyo[2], C. D. Huber[3], M. D. Edge[1], N. Mancuso[1], C. Fan[1]; [1]Univ. of Southern California, Los Angeles, CA, [2]Natl. Autonomous Univ. of Mexico, Queretaro, Mexico, [3]Penn State Univ., University Park, PA

Methods to infer demographic history from genetic data typically rely on extracting low-dimensional features from the underlying gene genealogies. As a result, the likelihood functions in these methods incompletely capture the observed information and are often degenerate or divergent, leading to poor optimization or unstable inference. Here we propose a novel framework called the genealogical likelihood (gLike) that addresses these limitations by deriving the full likelihood of a genealogical tree under any hypothesized demographic history. It is motivated by the fact that a genealogical tree does not specify the populations of its ancestral lineages, and a large number of possible trajectories of their population memberships may exist that are compatible with the observed coalescences throughout the history. Employing an innovative graph-based structure, gLike is able to summarize the relationships between all such trajectories and efficiently compute the exact marginal probability under a parameterized demographic model. Through extensive simulations of multiple admixtures as well as admixed demographic models from stdpopsim, we show that when the true genealogy is known, gLike has excellent accuracy (mean estimation error < 5%) in estimating dozens of demographic parameters, such as ancestral population sizes, admixture timing, and admixture proportions. Moreover, when using genealogical trees estimated from genetic data with tsinfer+tsdate, we show that gLike outperforms conventional demographic inference methods that leverage only the allele frequency spectrum. Furthermore, our framework traces ancestral histories by analyzing solely a sample from the admixed population, removing the need to have sampled the ancestral populations that may no longer exist. To illustrate the utility of gLike on real data, we performed demographic inference using estimated genealogical trees based on 1,000 individuals each from the Latino American and Native Hawaiian populations of the Multiethnic Cohort, yielding parameter estimates that align with established historical knowledge of their past demographic history. For example, we found the Native Hawaiians to be more recently admixed than the Latinos, with a slightly smaller initial population size. Overall, our proposed gLike framework harnesses the complete genealogical information to offer exceptional sensitivity and accuracy in inferring complex demographies.

Session Title: Evolutionary and Population Genetics Poster Session I

PB3004 A multivariate fixation index for measuring genomic differentiation of human populations

Authors:

I. Jordan, G. P. Yerradoddi, S. Sharma; Georgia Inst. of Technology, Atlanta, GA

Univariate variance partitioning methods (e.g. Wright's fixation index Fst) are used to measure genetic differentiation among populations, whereas multivariate methods (e.g. principal component analysis) are used to classify individuals' population origins. Univariate genetic differentiation methods average information across loci, refining the precision of single locus estimates, whereas multivariate classification methods accumulate information across loci, resulting in high precision for population origin assignments. We have developed a multivariate fixation index for measuring genomic differentiation of human populations, which combines the variance partitioning and classification approaches to yield a more accurate picture of the apportionment of human genomic diversity. Our multivariate method leverages the correlation structure of human genomic variation, accumulating information across loci, to estimate the within and between population genomic variance components based on multilocus data. In other words, the multivariate fixation index allows us to characterize the relative similarity for random pairs of genomes drawn from the same versus different populations, rather than the relative similarity for random pairs of loci within versus between populations as measured by the univariate fixation index. We applied our multivariate approach to whole genome sequence data from the 1000 Genomes Project, partitioning variation among populations within regions and among regions. The multivariate fixation index reveals human populations to be substantially more genomically differentiated than can be seen using univariate methods, consistent with the fact that pairs of individuals from the same population tend to be far more closely related than pairs of individuals from different populations.

Session Title: Evolutionary and Population Genetics Poster Session II

PB3005 A new model-based method controlling for gene flow using ancient genomes obtains less biased estimates of the selection coefficient

Authors:

X. Li[1], J. Novembre[2]; [1]Univ. of Chicago, Chicago, IL, [2]Univ Chicago, Chicago, IL

Studying genetic signatures of natural selection using ancient DNA may help identify functionally important variants. However, the complex demographic history of humans can hide selection signals or lead to spurious discovery. In particular, gene flow from diverged populations can cause rapid shifts in allele frequencies that may be mistaken as selection signatures.

To address this limitation, we developed a new method to infer selection while controlling for the effects of gene flow. We model the allele frequency change of a target population as a function of genetic influx from other source populations, drift, and selection. First, we infer the posterior distribution of migration proportion parameters using SNPs from putatively neutral regions. Second, conditioning on the migration proportions, we estimate the selection coefficients for variants in interest. At each step, to obtain the posterior distribution efficiently, we use variational inference.

To test this method, we applied our method to simulated data. In these simulations, compared to a method that does not account for migration, our approach obtains a more accurate and unbiased selection coefficient. For instance, in the case of 30% pulse migration, an initial frequency of 0.01 in the target population, and a selection coefficient of 0.001, the approach without considering gene flow infers on average a 3-fold larger selection coefficient estimate ($s=0.0034$ vs. $s=0.0012$). In addition, when we simulate neutrally evolving alleles using a realistic demographic model, the approach that does not account for gene flow obtains a 11.4% false positive rate while our method obtains a well-controlled 3.6% false positive rate.

We applied our method to publicly available data from 481 individuals from the Carpathian Basin, including samples ranging from 5000 BCE to 900 CE (Mallick et al., 2023). Our method fits two major events of gene flow: 1) from Steppe-like ancestry during the Bronze Age and 2) from Xiongnu-like ancestry in the Medieval period. Correcting for these gene flow events, we identify selection signals associated with light skin pigmentation and immune response. For instance, our method recovers a well-studied selection signal at rs1426654 (posterior mean of $s= 0.024$ with 95% credible interval [0.00078, 0.047]) while a method that does not account for gene flow does not have enough power (posterior mean of $s= 0.021$ with 95% credible interval [-0.0090, 0.050]).

While gene flow among populations has been shown to be pervasive across human history, our new method can be a powerful way to infer selection using ancient DNA without biases from gene flow.

Session Title: Evolutionary and Population Genetics Poster Session III

PB3006 A novel approach, LASeQTL, shows local ancestry has widespread effects on regulation of gene expression and improves functional annotation of GWAS signals

Authors:

H-H. Chen[1,2], W. Zhu[3], L. Petty[2], H. Polikowsky[3], J. McCormick[4], S. Fisher-Hoch[4], H. Highland[5], K. North[5], E. Gamazon[2], J. Below[2]; [1]Academia Sinica, Taipei, Taiwan, [2]Vanderbilt Univ. Med. Ctr., Nashville, TN, [3]Vanderbilt Univ., Nashville, TN, [4]Univ. of Texas Sch. of Publ. Hlth., Brownsville, TX, [5]Univ. North Carolina at Chapel Hill, Chapel Hill, NC

Expression quantitative trait loci (eQTL) are powerful genetic predictors of molecular phenotypes and are the basis of many functional characterizations of genetic associations with disease. Yet, to date most eQTL studies have been performed in populations with predominantly European ancestries. Here, we leverage genetic and RNA sequencing data from 645 whole blood RNA samples from the Cameron County Hispanic Cohort to conduct eQTL mapping on haplotype-specific expression while considering local ancestry. To accomplish this, we developed a novel approach to map local ancestry-specific (LAS) eQTL, LASeQTL. Global genetic ancestry analysis identified an average of 49.0% Amerindian (AMR), 45.7% European (EUR), and 4.6% West African ancestry in our samples. Here, we focus on EUR and AMR specific eQTLs. In total, we identified over 3 million cis-eQTLs for 8,523 genes (FDR-adjusted p-value<0.05), of which 4,724 genes have at least one LASeQTL. Interestingly, in 1,411 of these genes, 37,544 LASeQTLs exhibited opposing effect directions on expression between the EUR and AMR ancestral haplotypes. Highlighting the genome-wide differences in genetic regulation by ancestral haplotypes, the median genetic correlation of LAS gene expression was 0.50 between EUR and AMR in the CCHC among heritable genes (heritability>0.05), and 0.53 for highly heritable genes (heritability>0.4); in contrast, a high concordance was observed between CCHC local European specific gene expression and GTEx (European >84.6%), median genetic correlation was 0.59 for heritable genes and 0.91 for highly heritable genes. To demonstrate the importance of our approach, we further performed colocalization analysis with our identified ancestry-specific eQTLs and the Hispanic/Latino type 2 diabetes GWAS from DIAMANTE (n=33,808); eight GWAS signals colocalized with our LASeQTLs whereas only one signal was colocalized with the GTEx eQTLs. Such differences may explain the differential predictive value of functionally oriented models of disease risk in admixed populations and the inconsistent GWAS and TWAS findings across diverse populations. In conclusion, we represent the first profile of a local ancestry-specific regulatory landscape of gene expression in Hispanic/Latinos, and demonstrate its importance on functional interpretation of previous genetic studies' findings, continuing to emphasize the importance of diversity in genetic research.

Session Title: Evolutionary and Population Genetics Poster Session I

PB3007 A Pangenome Reference of 36 Chinese populations

Authors:

S. Xu; Fudan Univ., Shanghai, China

Human genomics is witnessing an ongoing paradigm shift from a single reference sequence to a pangenome form, but populations of Asian ancestry are underrepresented. Here, we present the first effort of the Chinese Pangenome Consortium (CPC) with a collection of 116 high-quality and haplotype-phased *de novo* assemblies based on 58 core samples representing 36 minority Chinese ethnic groups. With an average 30.65× high-fidelity long-read sequence coverage, an average contiguity N50 of more than 35.63 Mb, and an average total size of 3.01 Gb, the CPC core assemblies add 189 million base pairs of euchromatic polymorphic sequences and 1,367 protein-coding gene duplications to GRCh38. We identified 15.9 million small variants and 78 thousand structural variants, of which 5.9 million small variants and 34 thousand structural variants were not reported in a recently released pangenome reference[1]. The CPC data demonstrate a remarkable increase in the discovery of novel and missing sequences when individuals are included from underrepresented minority ethnic groups. The missing reference sequences were enriched with archaic-derived alleles and genes that confer essential functions related to keratinization, UV response, DNA repair, immunological responses, and lifespan, implying great potential for tracing missing links in human evolution and recovering missing heritability in complex disease mapping.

Session Title: Evolutionary and Population Genetics Poster Session II

PB3008 A reference panel to improve genotype imputation for Native Hawaiians

Authors:

B. Dinh[1], X. Sheng[1], E. Tang[1], F. Chen[1], E. Young[2], K. Taparra[3], L. R. Wilkens[4], L. Le Marchand[4], I. Hall[5], N. Stitziel[2], C. A. Haiman[1], C. W. K. Chiang[1]; [1]Univ. of Southern California, Los Angeles, CA, [2]Washington Univ. in St. Louis, St Louis, MO, [3]Stanford Univ., Stanford, CA, [4]Univ. of Hawai'i, Honolulu, HI, [5]Yale Univ., New Haven, CT

Imputation is now a fundamental component to conduct powerful association studies. In recent years, public imputation reference panels have increased their size, representation, and accessibility. Recent findings have shown that the state-of-the-art reference panel from the TOPMed consortium now imputes comparably for minority populations such as African Americans and Latinos as it does for individuals of European ancestries, an improvement over previous reference panels. Nevertheless, the accuracy and effectiveness of imputation remain low among global populations whose ancestries are not adequately represented in TOPMed. We created an imputation reference panel comprised of 10,721 whole-genome sequenced individuals from the Multiethnic Cohort (MEC), with multi-ancestry representations from five ethnic groups: African Americans (1,270), Native Hawaiians (1,065), Japanese Americans (3,106), Latinos (4,141), and non-Hispanic whites (1,139). In total, we constructed the reference panel based on 82M SNPs and 5M indels passing quality control measures. Despite being approximately 10% the size of TOPMed, we find that our reference panel outperforms TOPMed for populations not well-represented such as Native Hawaiians (NH) and Japanese Americans (JA). We imputed 3,276 NH and 4,176 JA individuals genotyped on the MEGA array and find that both populations showed improved imputation accuracy with the MEC panel. For example, for alleles with 1-5% frequency, the mean imputation Rsq for NH and JA are 0.921 and 0.908, respectively, compared to 0.886 and 0.751 when imputed on the TOPMed imputation server. Comparing to genotypes from 100 held-out sequenced individuals, we also confirmed that NH and JA individuals benefited from the MEC reference panel, with an increase of 0.03 and 0.13 in squared-Pearson's correlation coefficient (r^2), respectively. These improvements are stronger for NH more enriched with Polynesian ancestries (0.05 increase in r^2 compared to TOPMed), suggesting that our reference panel may also benefit other Pacific Island populations enriched with Polynesian ancestries in general. As this improvement due to increased representation does not require sequencing a prohibitive number of individuals, our study highlights the urgent need to generate genomic resources that will enable better association studies for underserved populations across the globe.

Session Title: Evolutionary and Population Genetics Poster Session III

PB3009 A scan of selective signals in whole-genome sequencing data from Canary Islanders

Authors:

J. Lorenzo Salazar[1], R. Gonzalez-Montelongo[1], L. Rubio-Rodriguez[1], A. Muñoz-Barrera[1], D. Jáspez[1], A. Díaz-de Usera[1], V. Garcia Olivares[1], A. Iñigo Campos[1], B. Guillen-Guio[2,3,4], I. Marcelino-Rodríguez[2,5], A. Corrales[2,6], D. Comas[7], C. Flores[1,2,6]; [1]Genomics Div., Inst. of Technology and Renewable Energy (ITER), Granadilla de Abona, Spain, [2]Res. Unit, Hosp. Univ.rio Ntra. Sra. de Candelaria, Santa Cruz de Tenerife, Spain, [3]Dept. of Population Hlth.Sci., Univ. of Leicester, Leicester, United Kingdom, [4]NIHR Leicester BioMed. Res. Ctr., Leicester, United Kingdom, [5]Publ. Hlth.and Preventive Med. Area, Univ. de La Laguna, Santa Cruz de Tenerife, Spain, [6]CIBER de Enfermedades Respiratorias (CIBERES), Inst. de Salud Carlos III, Madrid, Spain, [7]Departament de Ciències Experimentals i de la Salut, Inst. de Biología Evolutiva (CSIC-UPF), Univ.t Pompeu Fabra, Barcelona, Spain

Rationale. The genetic uniqueness of the contemporary Canary Islands (Spain) population is explained by historic admixture during and after the conquest of the islands in the XV century. Ancestry influences from Iberians (IBS) and North Africans (NAF) constitute the main genetic components that are recognized in the population. Studies with SNP array data have revealed a few loci that are enriched in one of the genetic ancestries in the current inhabitants and some of these loci showed evidence of selection.
Objective. To use whole genome sequencing (WGS) to systematically scan for regions evidencing putative natural selection in the genome of the contemporary Canarian population.
Methods. WGS at ~30x from 100 unrelated Canary Islanders (CAN) and 100 NAF (Illumina HiSeq2500, HiSeq4000, and NovaSeq6000, paired-end 150 bp), and 100 IBS (The 1000 Genomes project) have been processed. Small germline genetic variation was characterized using BWA-GATKv4 (hg19) based pipelines. Putative signals of selection were identified from genetic neutrality and diversity tests (D-Tajima, PBS, iSAFE) at different window sizes. Values at the extremes of the distributions were prioritized, functionally annotated, and used in the enrichment analyses by combining GREAT, Open Targets Genetics, and Ensembl.
Results. Approximately 25 million variants were identified in CAN, of which ~18% were not present in dbSNP (avsnp150) and 43% were not described in TOPMed freeze3. A focus on the top 0.1% of the D-Tajima distribution revealed more than 1,300 putative regions of selection. This includes the highly polymorphic region encompassing HLA (*TSBP1*), as well as other well-known regions under natural selection (*LCT*). The enriched terms in the prioritized genetic regions include autoimmune, ophthalmologic, and oncologic diseases, among others.
Conclusion. Our results support the existence of a genuine gene pool in the current Canary Islanders. Further studies to evaluate the relationship between the putative signals of selection, disease links, and the admixture sources are underway.
Funding. Ministerio de Ciencia e Innovación (RTC-2017-6471-1), cofinanciado por el Fondo Europeo de Desarrollo Regional “Una manera de hacer Europa” de la Unión Europea; Fundación CajaCanarias y Fundación Bancaria “La Caixa” (2018PATRI20); Acuerdo ITER OA17/008; Cabildo de Tenerife (CGIEU0000219140); Ministerio de Educación, Cultura y Deportes (FPU16/01435); Convenio Marco de Cooperación Consejería de Educación-Cabildo Insular de Tenerife 2021-2025 (CGIAC0000014697).

Session Title: Evolutionary and Population Genetics Poster Session I

PB3010 A significant difference in the prevalence of genotypes and alleles in SNPs known to be associated with strength, endurance and injury propensity in the local population compared to the prevalence in global databases

Authors:

D. Bercovich[1,2], H. Salem[3]; [1]Tel Hai Academic Coll., Tel Hai, Israel, [2]Galile Genetic Analysis - GGA Ctr., Kazerin, Israel, [3]Ariel Univ., Ariel, Israel

There are number of SNPs that are known in the literature where different genotypes in the same SNPs may influence people's optimal athletic ability and determine whether they are suitable for strength type sports or more suitable for endurance sports type and their susceptibility to injuries after sports activities (such as muscle injury or the ability to recover and the susceptibility to inflammation in the Achilles tendon). In this study, we examined the prevalence of these SNP genotypes and alleles frequencies in about 3600 individuals from the Israeli population from all ethnic groups and compared the prevalence of these genotypes and alleles to those known in databases (such as dbSNP). A large number of differences were found in the prevalence of genotypes known to affect these traits, such as SNP rs1042713, in which the genotype affecting suitable for endurance, the AA genotype, is found worldwide at a prevalence of 20% and in the Israeli population only 3%. Also, in SNP rs12594956, the AA genotype that affects this trait is found worldwide at a prevalence of 50% and in Israel only by 31% . in the rs679620 SNP, the CC genotype, which is known to predict susceptibility to Achilles tendon inflammation, is found worldwide at a prevalence of 39% and in the Israeli population only 25%. The work will present differences in SNP prevalence related to strength and endurance tendency to muscle injuries and inflammation of the Achilles tendon.

Session Title: Evolutionary and Population Genetics Poster Session II

PB3011 A symphony of genes: Exploring the diverse landscape of human copy number variable Y-chromosomal ampliconic genes.

Authors:

M. Loftus[1], P. Hallast[2], P. Ebert[3], F. Yilmaz[2], The Human Genome Structural Variation Consortium, T. Marschall[3], C. Lee[2], M. K. Konkel[1]; [1]Clemson Univ., Clemson, SC, [2]The Jackson Lab. for Genomic Med., Farmington, CT, [3]Heinrich Heine Univ., Düsseldorf, Germany

The human Y chromosome is essential to sex determination and fertility. Most of the Y chromosome is comprised of the male-specific region (MSY) that does not recombine with the X chromosome. Across the MSY there are seven ampliconic regions, which collectively contain nine multi-copy ampliconic genes (*BPY2*, *CDY*, *DAZ*, *HSFY*, *PRY*, *RBMY1*, *XKRY*, *VCY*, and *TSPY*). These ampliconic genes perform pivotal roles in male-specific functions, particularly spermatogenesis, and their aberrant expression has been associated with a multitude of diseases, including a variety of cancers. Unfortunately, studies for some of these genes have faced hindrances due to the incomplete assembly of their genomic regions (e.g., the *TSPY* array). The initial assembly of the first human Y chromosome sequence was generated nearly two decades ago, yielding a high-quality, yet incomplete sequence (approximately 53.8% unresolved in GRCh38 Y). Recently, a significant milestone was achieved by the Telomere-to-Telomere (T2T) Consortium, who reported the first complete *de novo* assembly of a human Y chromosome from end-to-end (Rhie et al. - manuscript under review, preprint available on bioRxiv). Combining PacBio HiFi and Oxford Nanopore Ultra-Long read sequencing technologies, we have assembled 43 human Y chromosomes with half of our Y chromosome assemblies encompassing African lineages. From the analysis of these 44 (43 + T2T Y) new assemblies, in conjunction with an exon-centric approach, we find considerable genetic diversity, including copy number variation, of the ampliconic genes. Two ampliconic genes, *RBMY1* and *TSPY* are particularly copy number variable (*RBMY1*: 5 - 11 copies; *TSPY*: 23 - 46 copies). Utilizing chimpanzee as an outgroup, we determined which gene copies are the most ancestral and examined the potential functional diversity of these genes. Lastly, we provide examples of how recombination events (e.g., gene conversion and non-allelic homologous recombination) might maintain the function of ampliconic gene arrays through the purging of low divergence (less than 2%) haplogroup-specific pseudogenes. Our findings not only enhance our understanding of more than 180,000 years of human Y chromosome ampliconic gene evolution, but could lead to the better diagnosis and treatment of associated diseases.

Session Title: Evolutionary and Population Genetics Poster Session III

PB3012 † A timeline of human evolution: Leveraging GWAS and comparative genomic data to contextualize human-evolved diseases and morphological traits

Authors:

E. Kun[1], M. Sohail[2], V. Narasimhan[1]; [1]The Univ. of Texas at Austin, Austin, TX, [2]UNAM, Cuernavaca, Mexico

Higher levels of cognitive function, bipedalism, and other uniquely human traits evolved over millennia, but genomic evidence for these changes has been lacking. Recent high-quality reference genomes of 240 different primates, ancient DNA from archaic hominids, and comparative functional genomic data across primates have identified human gained brain enhancers and promoters (HGEPs), selective sweeps, neanderthal introgressed regions (NIRs), and human accelerated regions (HARs). These correspond to different evolutionary periods and can be associated with traits, contextualizing the evolutionary history of each trait. To do so, we used stratified LD score regression as well as protein coding gene enrichment analysis to integrate these key evolutionarily genomic regions with GWAS of several human traits and diseases and image derived morphological traits of the heart, brain, and skeletal structure. First we examined functional genomic annotations and saw differences in heritability enrichment across stages of development and traits we examined. Virtually all traits with enrichment for overlap in HGEPs were only significant in early time points but not in adults, suggesting that major differences in trait gene expression between humans and primates lie in early development. At least 1 psychiatric, dermatological, and respiratory trait showed enrichment while no immune or metabolic traits did. Similarly, we find significant enrichment for skeletal and brain morphology traits but not heart morphological traits. Next we examined annotations derived from comparative analysis of DNA sequence. Genes associated with various psychiatric traits were especially enriched in HARs, alongside brain visual cortex area and arm-to-leg ratio in line with known human anatomical evolution. Conversely, ancient selective sweeps and NIRs had greater enrichment with immune and reproductive traits, respectively. Further, genes associated with the thalamus, which is responsible for relaying sensory information in the brain, and heart atrial area were specifically enriched in NIRs. As HARs arise from differences between humans and great apes while selective sweeps and NIRs are related to archaic humans, our results paint a temporal map of which traits likely evolved before emergence of bipedal hominids as well as which traits continued to evolve or only began to evolve in more recent human history. Overall, we combined large datasets from comparative, ancient, medical, functional and imaging genomics to investigate the evolution of human traits and connect these to evolutionary timepoints, furthering our understanding of the emergence of human specific biology.

Session Title: Evolutionary and Population Genetics Poster Session I

PB3013 A tree-based algorithm for identifying mutator alleles from genetic variation data

Authors:

L. Zhu, K. Harris; Univ. of Washington, Seattle, WA

Mutator alleles are genetic variants that increase the mutation rate, for example by interfering with DNA repair. Although mutators have been hypothesized to play a role in the rapid divergence of mutation spectra between vertebrate populations and species, they are difficult to detect in sexual organisms because they quickly recombine away from the mutations they beget. In theory, mutator alleles segregating in large genetic variation databases should be detectable via linkage to excess rare variants; here, we describe a tree-based algorithm for identifying this signature. At a mutator locus, edges within the subtree subtending the mutator allele are expected to accumulate mutations at an elevated rate, as are linked edges in adjacent subtrees. Our algorithm uses a tree sequence to identify the "ARG footprint" of a candidate mutator allele, which is essentially the set of tracts inherited identical by descent between each pair of sequences carrying the candidate mutator. We use SLiM simulations to verify that ground-truth ARG footprints show good agreement with the ARG footprints that our algorithm efficiently infers, and we perform a power analysis showing that our approach is capable of identifying mutators of moderate effect sizes from large tree sequences constructed using *tsinfer*. We have started inferring for TOPMed individuals to potentially investigate the presence of mutator activity.

Session Title: Evolutionary and Population Genetics Poster Session II

PB3014 Adaptations to diving in the Haenyeo of Jeju, Korea

Authors:

M. Ilardo[1], R. Nielsen[2], J-Y. Lee[3]; [1]Univ. of Utah, Salt Lake City, UT, [2]Univ California Berkeley, Berkeley, CA, [3]Seoul Natl. Univ., Seoul, Korea, Republic of

Around the world, various human populations have adjusted to a marine-based lifestyle by adopting the practice of subsistence breath-hold diving. One such population is the Haenyeo, a group of all female breath-hold divers on the island of Jeju, Korea. Perhaps even more remarkable than their diving abilities are the temperatures at which the dives are performed: in the winter, water surface temperatures dip to 10ºC. We pursued for the first time an investigation of the genetics and diving physiology of the Haenyeo. We measured baseline metrics associated with diving physiology as well as cardiovascular measurements during a series of simulated dives. We then used whole genome sequencing (WGS) data to identify genetic loci that have been under selection in this population and examined whether these loci were associated with candidate phenotypes. We report the demographic history of the Haenyeo as well as physiological and genetic adaptations to diving.

Session Title: Evolutionary and Population Genetics Poster Session III

PB3015 An ancient DNA time-transect of the Western Steppe from 1000 BCE to 1000 CE disproves the Khazar hypothesis of Ashkenazi Jewish origins.

Authors:

A. Shaus[1,2], I. Shingiray[3], C. Lalueza-Fox[4], R. Pinhasi[5], D. Reich[1,6,7,8]; [1]Harvard Med. Sch., Boston, MA, [2]Tel Aviv Univ., Tel Aviv, Israel, [3]Univ. of Oxford, Oxford, United Kingdom, [4]Inst. of Evolutionary Biology of UPF and CSIC, Barcelona, Spain, [5]Univ. of Vienna, Vienna, Austria, [6]Harvard Univ., Cambridge, MA, [7]Broad Inst. of MIT and Harvard, Cambridge, MA, [8]Howard Hughes Med. Inst., Boston, MA

Unlike many literate civilizations of the past, most of the cultures of the Western Steppe during the millennia before and after the turn of the common era left little, if any, written evidence. Hence, the only historical knowledge of these cultures is provided by the writing of often unfriendly neighbors, which is at best fragmentary and certainly tendentious. Additionally, in many cases, the archaeological record of these societies is restricted to isolated burials, which are difficult to characterize and analyze based on frequently limited and fragmentary funerary material culture, and is further complicated by the common presence of intrusive graves from later periods. We report genome-wide data from 105 Steppe individuals from Moldova, Ukraine, Russia, Georgia and Kazakhstan, 83 of which were radiocarbon dated. We trace the intricate effects of centuries-long patterns of migrations, admixtures and population displacements for populations likely to correspond to the historically described Scythians, Sarmatians, Alans, Khazars, and Bulgars. We detect a major population shift beginning with the second half of the 4th century CE, which can be associated with the consequences of the so-called “Hunnic mayhem”, postulated by some historians. By analyzing multiple genomes from the core area of the Khazar Khaganate, we set to rest the “Khazar hypothesis” of Jewish Ashkenazi origins, which postulates that Khazars, converted to Judaism in the 8th and 9th centuries CE, were a major contributing source to present-day Ashkenazi Jews. Instead, we find that Khazar burials have a typical Western Turkic genetic profile, with no evidence at all for recent shared ancestry with present-day or medieval Jews.

Session Title: Evolutionary and Population Genetics Poster Session I

PB3016 An SFS-based model of negative selection predicts fine-scale constraint in the noncoding genome

Authors:

E. Koch, D. Lee, V. Seplyarskiy, S. Sunyaev; Harvard Med. Sch., Boston, MA

The availability of whole-genome sequencing data with sample sizes in the hundreds of thousands has led to attempts to map fine-scale selective constraint across the human genome. This is generally done by comparing the probability that a variant is present in a sample to neutral expectations. These depend strongly on mutation rates, and a poor rate map can therefore lead to biased or noisy estimates of selection. For instance, tRNA genes are nearly ten times more mutable than expected based on sequence context, yet they likely constitute an important class of constrained sites in the human genome. We first developed a mutation rate model (Roulette) that includes such local effects and provides better neutral predictions than previous approaches. A second limitation of most current methods is that they ignore the frequency distribution (SFS) of observed variants. Hypermutable sites are highly informative about selective constraint, but the shape of the SFS for these sites depends on the mutation rate due to recurrent mutation. Inference for these sites can be achieved by combining precise rate predictions with modeling of selection's effect on allele frequencies. We developed a method (RaKLette) to move beyond presence/absence by modeling the zero-inclusive distribution of allele frequencies using Roulette. This method uses a multinomial transform of the SFS to account for mutation rate dependence. Multinomial coefficients are then fit using a regression approach that can capture constraint associated with regions of the genome or quantitative covariates. We summarize the effects of selection on a per-mutation level using the KL divergence between neutral and expected allele frequencies. Simulations based on realistic human population histories and distributions of selection coefficients suggest that RaKLette outperforms presence/absence models, with greatest improvements for high mutation rates and strong to moderate selection. We then applied RaKLette to whole-genome sequencing data from gnomAD v3 and compared constraint estimates generated using 1kb windows, enhancer models, and phylogenetic conservation. We provide a per-mutation annotation track along with quantitative estimates of fitness effect for various functional annotations.

Session Title: Evolutionary and Population Genetics Poster Session II

PB3017 Ancient DNA from 576 prehistoric people reveals the genetic history of the Sambaqui shellmound builders of Brazil.

Authors:

Y-C. Liu[1,2], M. Okumura[3], K. Sirak[1,2], O. Cheronet[4], N. Rohland[2], R. Pinhasi[4], D. Reich[1,2,5,6]; [1]Dept. of Human Evolutionary Biology, Harvard Univ., Cambridge, MA, [2]Dept. of Genetics, Harvard Med. Sch., Boston, MA, [3]Dept. of Genetics and Evolutionary Biology, Inst. of BioSci.s, Univ. of São Paulo, São Paulo, Brazil, [4]Dept. of Evolutionary Anthropology, Univ. of Vienna, Vienna, Austria, [5]Broad Inst. of MIT and Harvard, Cambridge, MA, [6]Howard Hughes Med. Inst., Harvard Med. Sch., Boston, MA

The shellmounds built by prehistoric hunter-gatherers of southeast Brazil are a prominent feature of the Brazilian coastline; the largest ones were constructed 5000-3000 years before present (BP), with the earliest evidence going back to 8000 BP. Around 1000BP, archaeological research documents a major transition from hunter-gatherers to ceramic-users with the appearance of two types of pottery cultures in the same region, known as Itararé and Tupiguarani, likely to have been associated with Jê and Tupi-Guarani speakers. In the absence of ancient DNA data, the genetic origin and interrelationship of these groups has been unclear, and it remains uncertain whether these early ceramic users were related genetically to the prehistoric hunter-gatherers. We generated genome-wide ancient DNA data for 576 prehistoric individuals dating to 5100-400 BP from coastal shellmound sites in three states (Rio de Janeiro, Paraná, and Santa Catarina) and inland rockshelter sites in two states (São Paulo and Rio Grande do Sul), thereby more than tripling the amount of human ancient DNA data reported from the Americas. We show that coastal shellmound builders are closely related to inland groups dating to 6000 BP and earlier, such as the inland riverine shellmound builders from the Laranjal and the Moraes sites of the Vale do Ribeira de Iguape in southern São Paulo from this time. We find that the genetic structure of shellmound builders correlates tightly with their geographic locations regardless of chronology and material culture, following an isolation-by-distance pattern. We document genetic continuity over 5000 years in the coast even during the lifestyle transition from hunter-gatherers to farmers associated with the Itararé culture, indicating that this culture transition was largely due to local innovation or adoption of ideas rather than to movements of people. This contrasts with many other regions of the world where ancient DNA data often documents large population displacements correlating with significant cultural change and in particular the lifestyle shift from hunting to farming. Our study also reveals that prehistoric groups from the inland rockshelter sites that have been hypothesized to be associated with Jê speakers are indeed more closely related to present-day Jê speakers than to contemporaneous coastal groups, providing support for models that the spread of people related to present-day and historically attested Jê-speakers was mediated by long-distance migration. We detect high amounts of Runs Of Homozygosity and extensive Identity-By-Descent segments throughout the ancient DNA time transect, implying a long-term small effective population size.

Session Title: Evolutionary and Population Genetics Poster Session III

PB3018 Ancient viral detection using metagenomic data from ancient individuals

Authors:

L. Nishimura[1,2], R. Sugimoto[1], H. Kanzawa-Kiriyama[3], K. Shinoda[3], I. Inoue[1]; [1]Natl. Inst. of Genetics, Mishima, Japan, [2]The Graduate Univ. for Advanced Studies (SOKENDAI), Mishima, Japan, [3]Natl. Museum of Nature and Sci., Tsukuba, Japan

Ancient DNAs have been massively sequenced because of the recent development of sequence techniques. Ancient DNAs contain viral genomic information accompanied with host genomes. Researchers have discovered ancient viral sequences in historical samples like bones, teeth, and mummified tissues, providing insights into past pandemics and long-term viral evolution. However, the number of ancient viruses found has been limited. Here, we analyzed whole genomic sequencing (WGS) data from ancient individuals who lived in the Japanese archipelago for thousands of years. We downloaded WGS data of ancient individuals from NCBI database. We used various bioinformatic techniques to detect ancient viruses, including de novo assembly, read alignment, metagenomic profiling, and non-homologous viral detection using bacterial CRISPR immunological memories. We recovered an almost complete sequence of the Siphovirus contig89 (CT89) and compared ancient and modern CT89 genomes, revealing that the most recent common ancestor of CT89 existed around 7,900 years ago. We also detected more than 10,000 candidate ancient viral genomes and some of them did not match modern viral sequences, possibly indicating highly diverged or extinct ancient viral genomes. We then characterized the ancient viromes of each individual, revealing differences between ancient and modern Japanese and European viromes that may reflect different dietary behaviors. Overall, these findings suggest that analyzing ancient viruses can be useful for understanding their existence in the past and long-term viral evolution.

Session Title: Evolutionary and Population Genetics Poster Session I

PB3019 Archaic humans have contributed to large-scale variation in modern human T cell receptor genes.

Authors:

M. Chernyshev[1], M. Corcoran[1], M. Mandolesi[1], S. Narang[1], M. Kaduk[1], K. Ye[1], S. Christopher[1], A. Färnert[1], T. Kreslavsky[1], C. Bernhardsson[2], M. Larena[2], M. Jakobsson[2], G. B. Karlsson Hedestam[1]; [1]Karolinska Inst.t, Solna, Sweden, [2]Uppsala Univ., Uppsala, Sweden

Human T cell receptors (TCRs) are critical for mediating immune responses to pathogens and tumors as well as regulating self-antigen recognition. Yet, variations in the genes encoding TCRs remain insufficiently defined. Detailed analysis of expressed TCR alpha, beta, gamma, and delta genes in 45 donors from four human populations-African, East Asian, South Asian, and European-revealed 175 additional TCR variable and junctional alleles. Most of these contained coding changes and were present at widely differing frequencies in the populations, a finding confirmed using DNA samples from the 1000 Genomes Project. Importantly, we identified three Neanderthal-derived, introgressed TCR regions including a highly divergent TRGV4 variant, which mediated altered butyrophilin-like molecule 3 (BTNL3) ligand reactivity and was frequent in all modern Eurasian population groups. Our results demonstrate remarkable variation in TCR genes in both individuals and populations, providing a strong incentive for including allelic variation in studies of TCR function in human biology.

Session Title: Evolutionary and Population Genetics Poster Session II

PB3020 † Associations between genome-wide autozygosity and complex traits in an ancestrally diverse US cohort.

Authors:

E. Johnson[1], S. M. Colbert[2], M. C. Keller[3,4], J. F. Wilson[5,6]; [1]Dept. of Psychiatry, Washington Univ. Sch. of Med., St. Louis, MO, [2]Dept. of Genetics and Genomic Sci., Icahn Sch. of Med. at Mount Sinai, New York, NY, [3]Dept. of Psychology and NeuroSci., Univ. of Colorado Boulder, Boulder, CO, [4]Inst. for Behavioral Genetics, Univ. of Colorado Boulder, Boulder, CO, [5]Ctr. for Global Hlth.Res., Usher Inst., The Univ. of Edinburgh, Edinburgh, United Kingdom, [6]MRC Human Genetics Unit, Inst. of Genetics and Cancer, The Univ. of Edinburgh, Edinburgh, United Kingdom

While inbreeding depression is often considered with respect to Mendelian diseases caused by rare, recessive variants, it has been shown to affect complex traits in humans as well. Even in seemingly outbred populations, the net effect of recessive variants acting on a trait can be estimated by genome-wide autozygosity (the proportion of the genome that is homozygous due to inheritance of alleles identical-by-descent). Prior studies have demonstrated associations between greater autozygosity and generally decreased fitness-related trait values for complex traits such as height, weight, and fertility. Questions that remain include the degree to which rare versus common variants are driving these associations with autozygosity, and the extent to which confounding factors play a role in observed associations. In this abstract, we present data on the relationship between genome-wide autozygosity and complex traits in a large cohort of diverse ancestries from the US (All of Us; max N = 206,819). We focused on traits that were previously associated with autozygosity in work from the ROHgen consortium. In a meta-analysis across ancestries, we found significant associations between greater autozygosity and fewer years of education, shorter height, and greater waist-to-hip ratio, consistent with prior studies. When using broad continental ancestry assignments provided by the All of Us program, there were noticeable ancestry-specific differences in some trait associations with autozygosity (e.g., weight); we suspect these differences may be due to residual population structure and are currently investigating this. Controlling for potential confounding factors, such as religiosity, did not substantially attenuate the observed associations. Consistent with a study in the UK Biobank, we found that the extent to which common or rare variants drive associations between autozygosity and complex traits seems to be trait-dependent; for example, the relationship between educational attainment and autozygosity seemed to be largely driven by rare variants, while results were more mixed for height and weight. Future studies should focus on within-family designs as the gold standard for limiting the potential influence of confounders. Additionally, we note that statistical power was limited for ancestry-specific models and the more fine-grained consideration of genetic ancestry required for follow-up analyses. These constraints exemplify the need for targeted recruitment of more diverse samples.

Session Title: Evolutionary and Population Genetics Poster Session III

PB3021 Assortative mating: Disentangling mate choice from population stratification

Authors:

A. Mas Sandoval, N. Barban; Università di Bologna, Bologna, Italy

Human populations do not mate randomly. Positive assortative mating occurs when individuals tend to mate with other individuals that are similar to them, resulting in a genetic correlation between mates. Nevertheless, this correlation can be driven by at least two different scenarios: population stratification and mate choice. Geographical and sociocultural barriers restrict social interactions and limit the mating opportunities across population strata. Consequently, mate choice is restricted to the network of interactions of an individual. Despite evidence of population stratification in most human populations, the observation of assortative mating is often interpreted as a consequence of mate choice by phenotype.

Here, we aim to disentangle the contribution of mate choice and population stratification to assortative mating. With this purpose we analyze homozygosity to study assortative mating. Population stratification results in higher genome-wide homozygosity than expected in Hardy-Weinberg equilibrium. However, mate choice by phenotype results in increased homozygosity in specific genomic regions associated with the trait. We scan the genome to study the patterns of homozygosity deviations, focusing on genetic variants associated with different complex traits. In addition, following this methodology, we can infer assortative mating patterns from genetic data without information of the mating couples.

We do not find evidence of mate choice driving assortative mating beyond the effect of population stratification in traits previously described to be under active mate choice. Our results highlight the importance of an accurate analysis of the population structure in genetic studies.

Session Title: Evolutionary and Population Genetics Poster Session I

PB3022 † Bayesian inferences of admixture dynamics in Brazil

Authors:

C. Silva de Carvalho[1], M. Gouveia[2], E. Tarazona-Santos[1], H. P. Sant Anna[1]; [1]Univ.e Federal de Minas Gerais, Belo Horizonte, Brazil, [2]NHGRI - NIH, Bethesda, MD

Because Latin Americans mostly result from admixture during the last half-millennium between Europeans, Africans, and Indigenous, their genomes are like a mosaic of fragments (i.e. tracts) deriving from those ancestries. Different methods are being developed to infer admixture dynamics, which is one of the frontiers of the population genetics of Latin Americans. Mutatis mutandi, this is like passing from the era of photographs (i.e. the current picture of admixture Illustrated by na Admixture-like barplot) to that of filmmaking (i.e. how admixture evolved). We are combining two approaches based on Approximate Bayesian Computation to infer admixture dynamics. One approach relies on the distribution of autosomal tract lengths, which has revealed that the low Indigenous ancestry of admixed Brazilians, characterized by short tracts, was mostly introduced immediately after the first arrival of Europeans into the Americas, consistently with the decimation of indigenous in Brazil. Our second approach combines information from autosomal and X-chromosomes using tract lengths, as well as the gradual approximation of X-chromosome variants to Hardy-Weinberg equilibrium, which allows us to infer, in addition to the dynamics of admixture, that of sex/ancestry bias (i.e. different ancestry of mating males and females). This approach applied to data to the city of Salvador, the most African city in Brazil (1,246 unrelated individuals, 50.8%[SE=0.35] of African ancestry) allowed to infer female- and male-mediated contributions of Europeans, Africans, and Indigenous during the 16th-20th centuries. African contribution was mostly mediated by females, and we inferred a sex bias driven by African females and European males reinforced in the 17th century that lasted even after the abolition of slavery. Our model-based approaches illustrate how genomic data on autosomal and X-chromosome may be combined to fill the gaps in the social history of Latin American populations.

Session Title: Evolutionary and Population Genetics Poster Session II

PB3023 Characterization of novel genetic variation in the major secreted airway mucins and associations with respiratory disease

Authors:

E. Plender[1,2], P. Hsieh[1], J. Ebler[3], K. M. Munson[1], A. Sulvari[1], W. K. O'Neal[4], T. Marschall[3], J. D. Bloom[1,2,5], E. E. Eichler[1,6]; [1]Dept. of Genome Sci., Univ. of Washington Sch. of Med., Seattle, WA, [2]Basic Sci. Div., Fred Hutch Cancer Res. Ctr., Seattle, WA, [3]Inst. for Med. Biometry and Bioinformatics, Heinrich Heine Univ. Düsseldorf, Düsseldorf, Germany, [4]Cystic Fibrosis/Pulmonary Res. and Treatment Ctr., Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, [5]Howard Hughes Med. Inst., Fred Hutch Cancer Res. Ctr., Seattle, WA, [6]Howard Hughes Med. Inst., Univ. of Washington, Seattle, WA

BACKGROUND: Few loci in the human genome have been linked to severe respiratory disease outcomes; mucin genes, however, are potential candidates due to their critical roles in pathogen entrapment. MUC5AC and MUC5B, two secreted mucins in the airway, feature coding variable number tandem repeat sequences (VNTRs) that allow binding to pathogens. Additionally, these loci are implicated in the pathogenesis of asthma. Long-read sequencing (LRS) enables resolution of these repetitive gene structures and detection of novel variants, thereby providing a novel resource for future genome-wide association studies with airway disease. **METHODS:** Long-read assemblies from the Human Pangenome Reference Consortium (HPRC, n=50) and Human Genome Structural Variation Consortium (HGSVC, n=53) were used to characterize variation within *MUC5AC* and *MUC5B,* while non-human primate (NHP) genome assemblies (chimpanzee, bonobo, gorilla, and orangutan) were used to place this variation in an evolutionary context. **RESULTS:** Most human haplotypes of *MUC5B* (90%) yield proteins between 5761-5763 amino acids (aa) in length; however, five human haplotypes predict proteins of 7020aa that include an additional VNTR domain. Species-specific motif utilization and domain copy number was identified in non-human ape alleles. Together, these results challenge the long-held belief that the central exon of *MUC5B* is invariable. *MUC5AC* features far greater variation; in fact, 30 human allelic variants encode 16 distinct proteins with variation observed in cys domain copy number and VNTR copy/domain number. Phylogenetically, we identified three clades of human *MUC5AC* where the three most common variants encode proteins of 5654 (46%), 5742 (33%), and 6326aa (7%). In contrast, most NHP genomes encode larger MUC5AC proteins (6031-7888aa) with novel VNTR domains, suggesting protein size reduction in the ancestral human lineage or varying rates of VNTR expansion across ape lineages. Additionally, striking patterns of linkage disequilibrium in the *MUC5AC/5B* locus in East Asian genomes are evidence of recent positive selection. Together, these findings contextualize the observations that medium-length alleles of *MUC5AC* (5742aa, 33% of haplotypes) are in linkage with GWAS SNPs associated with increased risk for severe pneumonia and shorter-length alleles of *MUC5AC* (5654aa, 46% of haplotypes) are in linkage with those for asthma. **CONCLUSION:** LRS provides the first model for VNTR evolution and variation in the *MUC5AC* and *MUC5B* loci. We are using these findings to develop bespoke methods to capture polymorphisms in these genes and to improve disease associations for respiratory disease outcomes.

Session Title: Evolutionary and Population Genetics Poster Session III

PB3024 Characterizing the impact of deleterious recessive alleles on genetic diversity across the transition between weak and strong selection.

Authors:

S. Sun, R. Hernandez; Univ. of California San Francisco, San Francisco, CA

Understanding how recessive deleterious variants impact neutral genetic diversity through genetic linkage is an outstanding problem in population genomics. To date, studies have focused on co-dominant deleterious variants due to their additive effects which can be very intuitive and easier to model. In the context of linked selection, recessive deleterious variants are capable of driving background selection similar to their co-dominant counterparts and thus, decrease genetic diversity. However, under a weaker selection regime, recessive deleterious variants can greatly increase genetic diversity through the mechanism of associative overdominance. To understand the nature of this transition from strong to weak selection and its impact on genetic diversity driven by recessive deleterious variants, we conducted a systematic forward in time simulation study across a wide range of selection coefficients. We then incorporated demographic models into our simulations to assess how events such as population bottlenecks and expansion further affect these dynamics.

Session Title: Evolutionary and Population Genetics Poster Session I

PB3025 Comparative analysis of molecular variation in the three-prime repair exonuclease 1 TREX1 across mammalian species and in human disease.

Authors:

M. Konate Monnard, M. Shekfeh, J. Krushkal; NCI, Rockville, MD

The three-prime repair exonuclease 1, TREX1, plays a major role in degrading cytosolic DNA, which may be generated through a variety of processes, such as exogenous viral infections, activation of endogenous retrotransposon activities, DNA damage during replication, micronuclei formation, or rupture of mitochondrial membranes. TREX1 inactivation results in DNA accumulation in the cytosol, which activates the cGAS-STING DNA sensing pathway and downstream interferon-mediated innate immune signaling and inflammation. Furthermore, germline pathogenic mutational defects in the TREX1 gene lead to hereditary autoimmune and autoinflammatory disorders. To assess the functional importance of residues along the TREX1 sequence, we analyzed protein sequences of the functional TREX1 isoform, TREX1b, from 167 mammalian species. While placental mammals showed a considerable degree of conservation of the entire TREX1b protein, egg-laying mammals and marsupials each had their own unique C-terminal domains at nearly one-third length of the TREX1b protein. The observed variability in the C-terminal domains in different groups of mammals is notable, as the C-terminus of the human TREX1 plays an important role in TREX1 localization to the endoplasmic reticulum, is subjected to post-translational modifications, and highly penetrant mutations in this region were reported in individuals with systemic lupus erythematosus (SLE) and retinal vasculopathy with cerebral leukodystrophy (RVCL). We will discuss implications of TREX1 variation and conservation across mammalian species and potential effects of TREX1 variants on human disease, which we inferred from existing data on germline TREX1 variants in hereditary autoimmune and autoinflammatory disorders and on somatic TREX1 variants in human cancer cells and tumor samples.

Session Title: Evolutionary and Population Genetics Poster Session II

PB3026 Comparative study of the founder effect in different populations.

Authors:

L. Gagnon, C. Moreau, C. Laprise, H. Vézina, S. Girard; Univ. of Quebec in Chicoutimi, Chicoutimi, QC, Canada

Background/objectives: Populations with founder effects have been particularly useful in demonstrating how past demographic events have shaped current genetic structure and its consequences on human health. Several demographic and historical events can lead to a founder effect. However, the resulting founder effect will be unique in each population. Moreover, there are few comparisons between populations with founder effects in population genetics. For these reasons, we propose to compare five populations with a documented founder effect: the Ashkenazi Jews, the Himba, the Hutterites, the Quebec, and a group of populations from South America (PAGE). These populations were chosen for the variability of their historical and demographic characteristics that have led to different genetic structures. Methods: We propose to characterize the populations using a principal component analysis (PCA) and a uniform manifold approximation and projection (UMAP), segments of runs of homozygosity (ROH), identical-by-descent segments (IBD), and effective population size (Ne). Results and discussion: The analysis of the ROH and IBD segments informs about the population structure. Some populations like the Himba and the Hutterites have longer segments for both types; and the Quebec, PAGE and Ashkenazi Jews have numerous shorter segments. These results can be linked to the Ne analysis which reveals that the populations with longer segments tend to have a bottleneck that is more important whereas populations with shorter segments tend to have a less important or older bottleneck. However, the analysis of population structure with a UMAP shows sub-population structures within some populations with founder effects, namely: the Ashkenazi Jews, the Quebec, and the PAGE group. Therefore, these populations were reanalysed considering the division of the main population with founder effects into more sub-populations. These comparisons, even between sub-populations of the same origin, reveal important differences in IBD sharing, ROH and the effective population size. Conclusion: These analyses will provide a better understanding of the genetic differences between populations with different founder effects. In addition, in some of these populations, there are significant structural differences leading to sub-population divisions. It is therefore essential to take these sub-populations into account to avoid biases and properly design studies on both Mendelian diseases and complex traits.

Session Title: Evolutionary and Population Genetics Poster Session III

PB3027 Comparison and evaluation of HLA calling from African Whole Genome Sequencing data in the UK Biobank

Authors:

Y. Liu[1], Z. Meng[2], Y. Huang[1], L. Xiao[1], Y. Wu[1], S. Tao[1], I. Adrianto[2], A. Levin[2], Q. Wang[1], H. Gui[3,4]; [1]Mental Hlth.Ctr. and Psychiatric Lab., West China Hosp. of Sichuan Univ., Chengdu, China, [2]Dept. of Publ. Hlth.Sci., Henry Ford Hlth., Detroit, MI, [3]Ctr. for Hlth.Policy and Hlth.Services Res., Henry Ford Hlth., Detroit, MI, [4]Behavioral Hlth.Services and Psychiatry Res., Henry Ford Hlth., Detroit, MI

Background: The human leukocyte antigen (HLA) region significantly influences the risk of various diseases, especially immune-mediated diseases like autoimmune, allergy, and neuroinflammation. Fine-mapping of causal variants in this region, approximately 6 Megabases, is a daunting task due to its highly polymorphic nature and complex linkage disequilibrium patterns. HLA imputation has proven successfully in inferring HLA types from regional single nucleotide polymorphisms (SNP) for European or East Asian samples. The refinement of methods and enlargement of reference panels are underway to secure more precise and reliable HLA genotypes and haplotypes. However, a substantial population-scale HLA reference panel for African ancestry is still absent. **Methods:** We applied whole genome sequencing (WGS) from African participants (n=1,000) in the UK Biobank (UKB) to identify each participant's HLA alleles, using the HLA*LA ("linear alignments") software. Subsequently, we contrasted our results with the HLA alleles imputed from SNP arrays using HLA:IMP*2 software, performed by the UKB team itself. The concordance of the two methods was separately evaluated separately for 10 typical HLA genes (3 for class I and 7 for class II). A new reference panel was formulated by the HLA-TAPAS pipeline. **Results:** The consistency of HLA alleles were reported at both 2-digit and 4-digit resolution. While biallelic 2-digit concordance was moderate to good for HLA Class I genes (93% for HLA-A, 67% for HLA-B, and 80% for HLA-C), it significantly dipped for HLA-DP genes (ranging from 29% for HLA-DPA1 to 0.2% for HLA-DPB1). For 4-digit concordance, the biallelic accuracy match is 63-89% for class I genes and 0.2-86% for class II genes. This level of consistency was notably lower than that for European population, and necessitating caution when using imputed HLA types directly for African samples in UKB. **Discussion:** The low concordance rates for imputed HLA genotypes in African sample highlight the exiting issue of diversity deficiency in the currently accessible imputation reference panels. For further study, we will compare existing HLA calls with those from other WGS calling pipelines, and then use updated HLA allelic frequencies to detect the HLA diversity and evolution of the African population in the UK.

Session Title: Evolutionary and Population Genetics Poster Session I

PB3028 Decomposing polygenic adaptation of myopia on the sunlight exposure hypothesis.

Authors:

T. Xia, K. Nakayama; Graduate Sch. of Frontier Sci., The Univ. of Tokyo, Kashiwa-shi, Japan

Myopia is increasingly prevalent worldwide due to modern lifestyles. Spending more time outdoors under bright sunlight can inhibit myopia progression. We hypothesize that residents of high latitude regions, adapted to dim light environments, have a lower tendency to develop myopia. This can be supported by Europeans, who exhibit lower myopia prevalence than East Asians despite similar urbanization and education attainment that increase indoor activities.We aim to investigate the genetic basis of population discrepancies related to the sunlight exposure hypothesis. We computed population branch statistics (PBS) and number of segregating sites by length (nS_L) statistics in 1000 Genomes Project (1KGP) Europeans. We mapped the PBS and nSL outliers in myopia-associated SNPs to genes and conducted KEGG pathway enrichment analysis, "phototransduction" is significant and aligned with our interest. Representative loci for this term were *RHO* and *TSPAN10*. *RHO* encodes rhodopsin in the rod cells that are sensitive to dim light, and *TSPAN10* is significantly associated with both myopia and hair color.We utilized a previously established methodology, improvising the PBS selection index, to identify instances of polygenic adaptation by subtle allele frequency shifts enriched within a functional pathway. Myopia-associated SNPs were mapped to 4,006 nearby genes and subsequently mapped to gene ontology (GO) biological process terms, eventually 260 non-redundant GO terms we obtained. We observed significant shifts in per-gene PBS selection index distributions for 12 GO terms. Our focus was on "camera-type eye development" (GO:0043010) and "visual perception" (GO:0007601) due to their relevance. The most significant locus, *MED1*, has been previously associated with mammalian circadian rhythms.Haplotype and allele frequencies of *RHO*, *TSPAN10*, and *MED1* exhibit a geographical distribution similar to the pigmentation-associated locus, *SLC24A5*. In two-sample Mendelian randomization analyses, causal inferences to myopia were significant for tanning ability and time spent outdoors but not significant for circadian traits. Therefore, the genetic basis of myopia might be influenced by adaptations in pigmentation and light perception pathways.To investigate the overall genetic discrepancy of myopia, we improvised polygenic risk score (PRS) in 1KGP Europeans and concluded that the major population discrepancies were driven by population structure and demographic history. In summary, we suggest that the current prevalence disparity of myopia is collectively influenced by population structure, local adaptation, and pleiotropic effects.

Session Title: Evolutionary and Population Genetics Poster Session II

PB3029 Detecting evolutionary adaptation and admixed ancestry in the Taiwanese Siraya people

Authors:

W-Y. Ko[1], W-C. Pan[1], J-N. Zeng[1], K-T. Lin[1], J-H. Loo[2], Y. Satta[2], M. Lin[2]; [1]Natl. Yang Ming Chiao Tung Univ., Taipei, Taiwan, [2]Mackay Mem. Hosp., New Taipei City, Taiwan

While the indigenous people of Taiwan are of central interest in studying genetic origins of the Austronesian-speaking people who have populated throughout Southeast Asia, Oceania, and the Pacific, it is less known that there were many Plain indigenous tribes (known as "Pingpu" people) who once inhabited in the plains and coastal regions of western Taiwan. The Pingpu people had largely admixed with Han Taiwanese and nowadays it is almost impossible to distinguish them from the Han Taiwanese. Little is known about the time and degree of population admixture and scenarios of evolutionary adaptation in the Pingpu people. Here, we studied genetic ancestry and adaptation of the Pingpu Siraya people by analyzing the whole genome genotyping data in 43 individuals for a total of 2,612, 357 SNPs. We showed that the Siraya are genetically distinct from the Han Taiwanese and phylogenetically grouped together with the Austronesian populations with a strong gene flow occurred recently from the Han Taiwanese. We inferred genetic components and local genetic ancestry for each individual genome. The average proportions are 0.41 (ranging: 0.3 - 0.59) and 0.26 (0.14 - 0.36) for the Han and Austronesian ancestries, respectively. The admixture event was dated as early as 220 years ago (around 1798 AD). Finally, we detected a number of candidate genes targeted by positive selection likely related to brain-volume development and body height, schizophrenia, immunity, or olfactory function. The obtained results of genetic adaptation and admixed ancestry is expected to facilitate the genetic profiling of disease susceptibilities and development of statistical models for disease-risk prediction.

Session Title: Evolutionary and Population Genetics Poster Session III

PB3030 Detecting late-onset disease causing variants in the genomes of Han Taiwanese people

Authors:

J-L. Chen[1], M-L. Kang[1], W-C. Tseng[2], J-H. Loo[3], M. Lin[3], H-H. Lee[1], Y. Satta[4], W-Y. Ko[1]; [1]Dept. of Life Sci. and Inst. of Genome Sci., Natl. Yang Ming Chiao Tung Univ., Taipei, Taiwan, [2]Taipei Veterans Gen. Hosp., Taipei, Taiwan, [3]Mackay Mem. Hosp., New Taipei City, Taiwan, [4]Res. Ctr. for Integrative Evolutionary Sci., SOKENDAI, Hayama, Japan

Identifying genes/genetic variants responsible for disease susceptibilities is a key challenge in studying common diseases. Many of these variants are categorized as late-onset disease causing variants, of which the carriers can often survive through adulthood (and leave offspring). These variants segregate in the population as neutral variants since natural selection is ineffective to remove them from the population, which causes considerable genetic burden in our population. Here, we aim to identify late-onset disease causing genetic variants by examining allele-frequency differences at different age groups of healthy individuals collected by the Taiwan Biobank. We analyzed the whole-genome genotyping data for a total of 509,873 variants in 42,467 individuals (age range: 30-70), and identified a number of candidate variants whose allele frequencies differ significantly along age. While most identified candidate variants are novel findings from this study, we indeed detected a pathogenic mutation in *BRCA1* (rs80357377) and a deletion in *MSH6* (rs587779299) known to be associated with Lynch Syndrome. Interestingly, several variants linked to the haplotype that carries the selection-favored ALDH2 allele appear to increase in frequency with age despite having been reported associated with late-onset diseases (i.e., cardiovascular and chronic kidney diseases). In order to identify possible causal variants, we also conducted linkage disequilibrium (LD) analysis for a subset of 772 individuals whose whole genome sequences are available. To further detect the functional effects of each candidate SNP, we performed multiple runs of phenome-wide association analyses across 27 clinical traits that are related to cardiovascular, renal, hepatic, hematologic diseases and diabetes. Many of the identified candidate SNPs appear to have pleiotropic connections with these traits. The knowledge gained from this study is expected to help inform strategies for disease prevention and treatment.

Session Title: Evolutionary and Population Genetics Poster Session I

PB3031 Detecting recent natural selection in East Asia using ancient DNA

Authors:

L. Colbran, I. Mathieson; Univ. of Pennsylvania, Philadelphia, PA

Identifying regions of the human genome influenced by population-specific selection enables us to identif traits that influence fitness, learn about past selective pressures, and understand their effects on genetic and phenotypic variation. Ancient DNA (aDNA) provides direct access to the evolutionary history of individual variants and has been used extensively to detect and classify occurrences of natural selection. However, this research has focused almost exclusively on Europeans due to their well-understood demographic history and large sample sizes. While populations in East Asia have a similar history—affected by the out-of-Africa migration and the transition to agriculture in parallel with Europeans—they have been exposed to many independent evolutionary pressures. This provides a unique opportunity to study the common effects of shared selective pressures and to identify novel selected loci and traits.

In this study, we take advantage of the recent increase in available aDNA from East Asia to expand aDNA-based studies of selection to these populations. We generalize a previous approach for detecting selection that models present-day populations as a mixture of ancient populations. Since the ancient East Asian populations are admixed, we extend the model to allow both the source and target populations to be admixed. We model 241 individuals from 3 ancient groups dating to the past 33,000 years, and 509 individuals from 5 modern populations, assuming Northern and Southern East Asian sources for the present-day mainland populations, and Northern and Jomon sources for Japanese populations. We fit the expected frequencies of each variant given this history and test for significant deviations using a maximum likelihood approach. We identify novel loci at *POLR3H* and *TEDC1*, in addition to known loci such as the HLA region, *FADS1* and *ALDH1B*.

We also adapt this approach to test for polygenic selection on traits analyzed in genome-wide association studies by counting the proportion of independent trait-increasing alleles seen in each population and applying a similar test. We use the results of GWAS conducted in populations of European ancestry to avoid issues related to population stratification. We test for selection on a wide range of polygenic traits and, for example, find that LDL levels have been subject to selection (uncorrected P = 3.77e-6). Overall, this study provides a comprehensive examination of recent selection in East Asia, incorporating additional population data and new methods. It will enable comparison with results from studies in Europe and increase our understanding of the role of selection on traits across multiple modern populations.

Session Title: Evolutionary and Population Genetics Poster Session II

PB3032 Detection of novel amylase haplotypes and insights into the genetic variation of amylase genes using optical mapping and long-read sequencing.

Authors:

F. Yilmaz; The Jackson Lab. for Genomic Med., Farmington, CT

The amylase locus is a complex region of the genome that contains AMY1, AMY2A, and AMY2B genes that code for the amylase enzyme that breaks down starch. Human-specific duplications of amylase genes were adaptively evolved in response to increased human starch consumption. Copy number variations of these genes are associated with agricultural lifestyles and metabolic disorders such as obesity and diabetes. The true extent of the genomic variation in this locus is still unknown due to the presence of paralogous segmental duplications (>45 kbp) that are difficult to map using conventional methods. We investigated the genetic variation and haplotype structures of the amylase locus utilizing Bionano Genomics optical genome mapping and PacBio HiFi long-read sequencing. We detected 30 haplotypes (H1-H30), with sizes ranging from 111.4 kbp to 376.1 kbp, containing novel duplications and inversions in 81 individuals from Human Genome Structural Variation and Human Pangenome Reference Consortium samples, surpassing the previous discovery of 10 haplotypes. Notably, amylase haplotype prevalence varies significantly between populations. For instance, H1-111.4 kbp, H5-205.6 kbp, H6-205.5 kbp, and H8-205.5 kbp were the most common haplotypes among individuals of African and American descent, while H8 and H10 were dominant among East Asians. Europeans predominantly displayed H1 and H6 haplotypes, whereas H6 was the most prevalent in South Asians. We hypothesize that this locus has evolved rapidly, as evidenced by population divergence and population-specific haplotypes. Further, we characterized the mechanisms that give rise to the variation in the locus. Non-allelic homologous recombination underlies the majority of amylase structural rearrangements, while three FoSTeS events were also identified. The analysis of AMY1 from diploid genomes revealed four copies were the most common (n=11), while two individuals had ten copies. Our study provides the most comprehensive map of amylase gene haplotype structures and novel insights into genetic variation. These findings pave the way for further research into the mechanisms underlying genetic variation at this locus, as well as the roles of amylase genes in human dietary adaptations, metabolic disorders, and nutrition.

Session Title: Evolutionary and Population Genetics Poster Session III

PB3033 Development of an integrated pipeline for comprehensive population genetic analysis

Authors:

N. Park, H. Kim, J. Kim; Konkuk Univ., Seoul, Korea, Republic of

The availability of population genetic variant data resulting from the advances in next-generation sequencing (NGS) technologies has led to the development of various population analysis tools to enhance our understanding of population structure and evolution. Currently, available tools analyzing population genetic variant data generally require different environments, parameters, and formats of input data, which acts as a barrier to the widespread use of such tools by general researchers not familiar with bioinformatics. To address this problem, we developed an automated and comprehensive pipeline PAPipe to perform seven widely used population genetic analyses using population NGS data. PAPipe seamlessly interconnects and serializes multiple steps, such as read mapping, genetic variant calling, data filtering, and format converting, in addition to seven population genetic analyses including population structure analysis, linkage disequilibrium decay analysis, fixation index analysis, population admixture analysis, principal component analysis, phylogenetic analysis, and pairwise sequentially Markovian coalescent analysis. PAPipe can be used to generate extensive results that provide clues or insights to enhance user convenience and data usability. PAPipe is intended for use in the Linux operating system, and the Docker image files are also provided for reducing the challenges in the environment settings.

Session Title: Evolutionary and Population Genetics Poster Session I

PB3034 Diagnostic yield of genome and exome sequencing in the ancestrally diverse CSER Phase II consortium is not associated with genetic ancestry in a variety of clinical settings

Authors:

Y. Mavura[1], D. Crosslin[2], K. D. M. Ferar[3], J. Greally[4], L. Hindorff[5], G. Jarvik[6], S. Kalla[7], B. Koenig[1], M. Kvale[1], P-Y. Kwok[1], M. Norton[1], S. Plon[8], B. Powell[9], A. Slavotinek[1], M. Thompson[10], A. Popejoy[11], E. Kenny[12], N. Risch[1]; [1]Univ. of California San Francisco, San Francisco, CA, [2]Tulane Univ., New Orleans, LA, [3]Univ. of Washington, Seattle, WA, [4]Albert Einstein Coll. of Med., Bronx, NY, [5]NIH/NHGRI, Bethesda, MD, [6]Univ. of Washington Med Ctr., Seattle, WA, [7]Baylor Coll. of Med., Houston, TX, [8]Baylor Coll. Med., Houston, TX, [9]Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, [10]Hudson Alpha Inst. for Biotechnology, Huntsville, AL, [11]Univ. of California, Davis, Sacramento, CA, [12]Icahn Sch. of Med. at Mount Sinai, New York, NY

Background It has been suggested that diagnostic yield (DY) from genome (GS) and exome sequencing (ES) may be higher among patients with predominantly European ancestry than those whose ancestral backgrounds may not be as well represented in genomic research and reference databases. Here, we examined the association of DY with estimated proportion(s) of genetic ancestry across global reference populations in a large cohort of diverse prenatal, neonatal and pediatric participants in various clinical/ healthcare settings. **Methods** A total of 3015 participants and their families were enrolled at one of 5 independent centers across the US as part of the Clinical Sequencing Evidence-Generating Research (CSER) phase II consortium. The participants were recruited from prenatal, NICU and pediatric clinics and underwent GS or ES for diagnosis. Genetic ancestry proportions were estimated from GS & ES data using Human Genome Diversity Project GS data and ADMIXTURE. Analyses compared the distribution of genetic ancestry between the combined CSER categories of definitive or probable ("positive") diagnostic results versus all others (negative and inconclusive) by Kolmogorov-Smirnov tests and linear associations of ancestry with DY by Cochran-Armitage trend tests within each of the 5 centers. The analyses were conducted overall and stratified by mode of inheritance. We then conducted fixed effects meta-analyses using weighted sum of Z-scores. **Results** Across centers, 18.7% (Min 5.2%, Max 27.9%) had a positive case outcome (DY). The majority of positive cases, 72.9% (Min 60%, Max 90.7%) were of Autosomal dominant (AD) MOI. The average estimated genetic ancestry proportions overall were African: 21%, American: 17.1%, East Asian: 4.5%, European: 46%, Middle Eastern: 7.7%, South Asian: 3.7%. We observed no statistically significant association in overall DY with any of the genetic ancestries for any of the 5 centers or in the meta-analyses. However, we observed a relative increase in Autosomal recessive (AR) homozygous MOI cases with South Asian ancestry (C-A P-value: 0.0011) approaching statistical significance due to consanguinity. There was decreased African ancestry for AD de novo and increased African ancestry for AD unknown MOI possibly due to differences in parental sequencing, but there was no association with AD DY overall. **Conclusion** No geographic genetic ancestry was associated with the likelihood of a positive diagnosis, supporting the possibility that ES and GS in diagnosis of previously undiagnosed but potentially Mendelian disorders across all ancestral populations may be more equitable than hypothesized in a variety of healthcare/ clinical settings.

Session Title: Evolutionary and Population Genetics Poster Session II

PB3035 Discovery of high-impact variants in CARTaGENE whole-genome sequences from diverse populations of Quebec

Authors:

K. Lo[1], S. Gagliano Taliun[1,2], C. Bhérer[3,4,5], S. Gravel[3], G. Lettre[1,2]; [1]Montreal Heart Inst., Montreal, QC, Canada, [2]Dept. of Med., Université de Montréal, Montreal, QC, Canada, [3]Dept. of Human Genetics, McGill Univ., Montreal, QC, Canada, [4]Canada Excellence Res. Chair in Genomic Med., McGill Univ., Montreal, QC, Canada, [5]Victor Phillip Dahdaleh Inst. of Genomic Med., McGill Univ., Montreal, QC, Canada

CARTaGENE (CaG) is a population-based cohort that includes ~43,000 participants recruited in Quebec (Canada) and is characterized by a wealth of longitudinally collected phenotypic information (e.g. questionnaire, health records, imaging, lab results). While the majority of its participants descend from French settlers who immigrated to Quebec in the 17th and 18th centuries, CaG is a cosmopolitan cohort that captures the diverse ancestries of modern Quebec.

We recently completed deep whole-genome sequencing of 1,756, 163 and 131 CaG participants of French-Canadian (FC), Haitian (HAI) and Moroccan (MOR) ancestry, respectively. Following rigorous quality-control measures, we obtained a total of 80,407,530 single-nucleotide variants and small insertions or deletions (indels). For variant annotation, we utilized two resources: the well-established Variant Effect Predictor (VEP) and the more recent Functional Annotation of Variants Online Resource (FAVOR) tools. To prioritize phenotypically interesting variants, we focused on the "VEP high-impact" variants (n=26,874), which encompass frameshift, splice acceptor and donor, start lost, stop gain and stop lost variants. Look-ups in ClinVar revealed that 864 (3%) of these variants are annotated as pathogenic/likely pathogenic. For all the common (minor allele frequency >1%) high-impact variants, we compared CaG and gnomAD allele frequencies, selecting gnomAD non-Finnish European, African/African American and mid-Eastern populations to match the CaG FC, HAI and MOR subsets, respectively. We found 119 FC, 288 HAI and 399 MOR variants with a ratio CaG/gnomAD >2, suggesting a potential enrichment of these variants in these populations. We also identified 50 FC, 71 HAI, and 287 MOR high-impact variants absent from gnomAD. To gain insights into the potential functional implications of these variants, we will conduct phenome-wide association studies, taking advantage of the rich phenotypic information in CaG.

Our findings contribute to a deeper understanding of the genetic diversity and potential disease risk factors within the FC, HAI and MOR populations. These results have implications for personalized medicine, population health research, and the development of targeted interventions tailored to the unique genetic profiles of these populations.

Session Title: Evolutionary and Population Genetics Poster Session III

PB3036 Disentangling clock-like mutation processes in germline and soma

Authors:

N. Spisak, M. de Manuel, W. Milligan, G. Sella, M. Przeworski; Columbia Univ., New York, NY

Decompositions of mutations into "signatures" based on the 96 single base substitution (SBS) types provide insights into mutagenesis in cancer, non-cancerous soma and germline. Among the signatures cataloged in the COSMIC database, two are ubiquitous in the soma and are considered "clock-like," because of their linear increase with age in some cell types: SBS signature 1, which consists primarily of transitions at methylated CpG sites caused by spontaneous deamination, and the much more diffuse SBS signature 5, which is of unknown etiology. The standard view of these signatures is much like Tolstoy's view of happy families: it is often assumed that while each non-clock-like mutational signature has a distinct source, all clock-like mutations are clock-like in the same way. As we show, this need not be the case. Firstly, we introduce a modeling framework that includes different modes of mutagenesis. This model highlights that mutational signatures can exhibit clock-like behavior due to several underlying mechanisms, including DNA replication errors, unrepaired damage, or damage that is repaired incorrectly. Next, we use de novo germline mutations as well as somatic mutations detected in single cells or small monoclonal populations in order to estimate the age-dependency of these mutational signatures in different cell types. In post-mitotic cells such as neurons and oocytes, there is no discernable increase of the number of SBS signature 1 mutations with age, in contrast to what is observed in dividing cell lineages. Thus, the underlying mutagenic process tracks cell divisions. One implication is that we can use the rate of accumulation of CpG transitions in tissues with known turnover rates to learn about division rates in other cell types. In turn, the number of the SBS signature 5 mutations does increase with age in post-mitotic cell types, indicating that the source of mutation is independent of cell division. Instead, SBS signature 5 accumulates with cell-type specific DNA damage rates, implicating errors in DNA repair as a key underlying mechanism.

Session Title: Evolutionary and Population Genetics Poster Session I

PB3037 † Dissecting the genetic and evolutionary sources of phenotypic variation in East Polynesians

Authors:

D. Liu[1], M. Roux[1], G. Rijo de Leon[1], I. Teiti[2], C. Harmant[1], A. Tessier[2], A. Bisiaux[1], V. Richard[2], A. Jaquaniello[1], M. Aubry[2], V-M. Cao-Lormeau[2], E. Patin[1], L. Quintana-Murci[3]; [1]Inst. Pasteur, Paris, France, [2]Inst. Louis Malardé, Papeete, French Polynesia, [3]Pasteur Inst. and Coll. de France, Paris, France

Understanding the links between genetic diversity, local adaptation and phenotypic variation is of upmost importance in evolutionary genetics; yet, most human genomics studies have focused only on European-descent populations. Such an imbalance precludes not only the transferability of genomic findings to other populations, but it provides a partial view of the genetic changes experienced by our species to survive new environments. Polynesia is of particular interest in this context as it has a unique settlement history, made of long-distance voyaging and extensive admixture, and is thought to present among the highest prevalence worldwide of metabolic disorders, raising questions regarding the role of natural selection and genetic drift in driving the genetic architecture of human traits. Here, we generated whole-genome sequencing data for 1,770 individuals originating from 18 East Polynesian islands and collected a wide range of demographic and phenotypic data relating to metabolism and morphology. Our analyses revealed that East Polynesians from all islands carry genetic ancestries that are maximized in present-day Austronesian speakers from Taiwan and the Philippines, Papuans, western Europeans, and East Asians, as well as native South Americans in the Marquesas Islands only. We observed large variation in East Asian- and European-related ancestries within islands, supporting a common history of recent, extensive admixture. We confirmed the high prevalence of obesity and type 2 diabetes in the region: > 50% of the population has a BMI > 32, and > 10% has glycated hemoglobin (HbA1c) levels > 6.5%. Applying multiple linear regression models and controlling for sex, age and island of residence, we found that BMI, HbA1c, cholesterol, HDL, and skin pigmentation (melanin index) are associated with genetic ancestry, suggesting that admixture has contributed to phenotypic diversity in Polynesians. For example, we detected a European-specific *SLC24A5* variant as a strong determinant of skin pigmentation variation in Polynesians. Furthermore, we identified novel variants associated with variation in metabolic and morphological traits, including variants that are private to Pacific populations. Interestingly, we found no signals of selective sweeps on such variants, suggesting a predominant role of genetic drift in driving phenotypic variation in the region. Together, these results highlight how quantitative genomics in understudied populations can increase our understanding of the genetic and evolutionary sources of human phenotypic variation.

Session Title: Evolutionary and Population Genetics Poster Session II

PB3038 Diversity and Representation of South Asian Genomes

Authors:

A. Das, M. C. Schatz; Johns Hopkins Univ., Baltimore, MD

The rapid growth in genomics has not been uniform across the full range of human diversity, leaving the majority of the world's populations poorly represented and resulting in systemic biases that can have serious impacts on clinical interpretations and other analysis. In this work, we aim to utilize recent advancements in genome sequencing and assembly to better understand the variation present in South Asian (SA) populations.

Using high quality short read data from 635 SA individuals in the 1000 Genomes Project and Simons Genome Diversity Project, we investigated the variation between these individuals relative to linear and pangenome references. To do this, we follow a similar pipeline used in the creation of the African Pan-Genome (Sherman et al. 2019) which assembles contigs from reads that were unaligned and/or poorly aligned relative to a chosen reference genome. We then attempt to place the larger contigs in the reference, allowing us to identify variants and novel sequences.

We focus our analysis on the T2T CHM13 genome to rule out spurious contigs created from unresolved regions in GRCh38. Interestingly, when using CHM13 we assemble ~1 Mbp of sequence from unaligned reads and ~15 Mbp of sequence from poorly aligned reads per individual, highlighting widespread population-specific sequence missing. We are able to successfully place more of these contigs against CHM13 than GRCh38, and find that they place evenly across the chromosomes, with those placements overlapping a range of biologically significant regions. We also find the majority of these assembled contigs to be private to the individual they are assembled from.

We similarly evaluated the recently released draft human pan-genome references from the Human Pangenome Reference Consortium, built from 47 individual genomes plus GRCh38 or CHM13. We observe higher alignment rates (+0.3%) of SA reads in pangenomes than in their corresponding linear reference, and find that the contigs assembled from unaligned reads are a subset of those assembled from the corresponding linear references. Nevertheless we still find ~10 Mbp of sequence per individual assembled from poorly aligned reads, including some contigs that are over 60Kbp in size, and many of which still overlap biologically significant regions.

We are currently investigating the transcriptional potential of these contigs using RNA-seq data from these individuals, as well as long read data from ongoing sequencing efforts. These data will help validate our short read results, allow us both to evaluate the functional significance of our novel sequence and investigate the overlap between placed contigs and clinically relevant variants.

Session Title: Evolutionary and Population Genetics Poster Session III

PB3039 † Diversity in the All of Us Research Program: race, ethnicity, and genetic ancestry

Authors:

S. Sharma[1,2], S. Nagar[1], L. Lichtenstein[3], P. Pemu[4,5,6], K. Mayo[7], A. Ramirez[8], A. Able[9,10], A. Green[10], L. Sokny[10], S. Zuchner[11], The All of Us Research Program, SEEC Consortium, L. Mariño-Ramírez[2], R. Meller[4], I. Jordan[1]; [1]Georgia Inst. of Technology, Atlanta, GA, [2]Natl. Inst. on Minority Hlth.and Hlth.Disparities, NIH, Bethesda, MD, [3]Broad Inst. of Harvard and MIT, Cambridge, MA, [4]Morehouse Sch. of Med., Atlanta, GA, [5]The All of Us Res. Program, NIH, Bethesda, MD, [6]SEEC Investigators, Univ. of Miami, Coral Gables, FL, [7]Vanderbilt Univ. Med. Ctr., Nashville, TN, [8]NIH, Mclean, VA, [9]VUMC, Nashville, TN, [10]All of Us Demonstration Projects Subcommittee NIH, Bethesda, MD, [11]Univ. of Miami Miller Sch. of Med., Miami, FL

The NIH All of Us Research Program aims to build one of the world's most diverse population biobanks in support of equitable precision medicine. For this study, we characterized the diversity of the All of Us participant cohort (n=165,072) with respect to race, ethnicity, and genetic ancestry. We evaluated the relationship between All of Us participants' self-identified race and ethnicity (SIRE) and their genetically inferred ancestry, exploring how genetic ancestry changes over space and time in the US and how it is related to health outcomes. Overall, minority racial and ethnic groups are overrepresented in the All of Us participant cohort, particularly for participants who identify as Black or African American. All of Us participants show diverse genetic ancestry, with major contributions from European (66.3%), African (20.7%), Asian (7.4%), and Native American (5.4%) ancestry components. Participant racial and ethnic groups show group-specific genetic ancestry patterns, with a continuum of ancestry proportions within groups. African and Native American ancestry are enriched in the southeast and southwest regions of the country, whereas European ancestry is more evenly distributed across the US. The diversity of All of Us participants' genetic ancestry is negatively correlated with age; younger participants show higher levels of genetic admixture compared to older participants. African, European, and Native American ancestries were found to be associated with numerous health outcomes in the All of Us cohort. African and Native American ancestry are both positively associated with diabetes and hypertension and negatively associated with skin disorders, whereas European ancestry is positively associated with numerous neoplasms and skin disorders. For mental health conditions, African ancestry is positively associated with schizophrenia, while European ancestry is positively associated with anxiety disorder. Our results underscore the diversity of the AoU cohort and suggest how ancestry can be leveraged to support genomic health equity.

Session Title: Evolutionary and Population Genetics Poster Session I

PB3040 Diversity of Y-chromosome haplogroup C in the Japanese population: Inference of its ethnic originhaplogroup C

Authors:

M. Osawa[1], R. Zhang[1], A. Ueda[1], E. Ochiai[2], S. Isozaki[1], K. Minaguchi[1]; [1]Tokai Univ Sch Med., Kanagawa, Japan, [2]Kitasato Univ Sch Med., Kanagawa, Japan

Y-chromosome exhibits 20 major haplogroups of A to T clades, based on SNPs. The majority of Japanese male individuals have C, D, and O clades. The C clade is widely distributed in the Eurasia continent, which overlaps the area of Altaic-speaking populations. In this study, we analyzed the divergence in the C lineage from SNP and short tandem repeat (STR), which were genotyped in the Japanese population, to investigate the anthropological origin as the ancestry. In analysis to DNA samples of unrelated Japanese male individuals (n = 1603), a total of 142 samples (8.9%) of C lineage, 631 (39.4%) of D, 479 (29.9%) of O1b, 283 (17.7%) of O2, and 68 (4.2%) of others were typed by the amplified product length polymorphism method. Then, 121 samples of C lineage were fully subtyped for 65 Y-SNP markers, which were selected from database of the International Society of Genetic Genealogy, by the SNaPshot system. The C samples were subclassified into three major subgroups of C1a1a (50.4%), C2a1 (5.0%) and C2b1 (44.6%), in which the most frequent allele was C1a1a1a* (Z7180) with 23 cases (19.0%). Based on Y-STR haplotypes using the Y-filer plus multiplex system, a median-joining network was constructed using Network 10.2. Two separate clusters of C1a1a and C2b1 were evident. The number of generations to the most recent common ancestor (TMRCA) with 95% CIs was calculated for the two decent clusters using the Ytime package. The subgroups of C1a1a and C2b1 were calculated 128 (58-257) and 165 (79-323) generations, corresponding to 4.0 (1.8-8.0) and 5.1 (2.4-10.0) thousand years, respectively, to TMRCA. Based on the database, C1a1 lineage is a Japanese-specific type, while C2b1 is common in southern China. The phylogenetic tree confirmed two major ancestry flows. One group of C1a1 that migrated to Japan in an ancient age up to around 4.0 thousand years should be indigenous, which is known as the Jomon era. The other that should be later migrated from southern China, which is known as the Yayoi era. This study provides an insight into the ethnic origin of the Japanese population.

Session Title: Evolutionary and Population Genetics Poster Session II

PB3041 † Dynamic clustering of genomics cohorts beyond race, ethnicity-and ancestry

Authors:

H. Mohsen[1], K. Blenman[2], P. Emani[2], Q. Morris[1], J. Carrot-Zhang[1], L. Pusztai[2]; [1]Mem. Sloan Kettering Cancer Ctr., New York, NY, [2]Yale Univ., New Haven, CT

Recent years have witnessed a steady decrease in the use of race categories in genomic studies. While studies that still include race categories vary in type and goal, these categories already build on a history during which racial color lines have been enforced and adjusted in the service of social and political systems of power and disenfranchisement, and when biological knowledge and data collection abilities were highly limited. Fixed, discrete classification systems have also limited the study of human biodiversity and disrupted widely spread genetic and phenotypic continuums across geographic scales. Relatedly, the use of broad and pre-defined stratification schemes—e.g. continental—across traits can risk missing important trait-specific genomic signals. To address these issues, we introduce a dynamic approach to clustering human genomics cohorts on a trait-specific level and without imposing a set of pre-defined categories. We tested our approach on whole exome sequencing datasets in 8 cancer types and partitioned these based on germline variants in cancer-relevant genes that could confer cancer type-specific disease predisposition. Results demonstrate clustering patterns that transcend discrete continent-based ancestry categories across all cancer types. Functional analyses based on cancer type-specific clusterings were also able to capture the fundamental biology underlying cancer and to identify novel potential drivers overlooked by a continent-based clustering model.

Session Title: Evolutionary and Population Genetics Poster Session III

PB3042 † Entwined African and Asian genetic roots ofmedieval peoples of the Swahili coast

Authors:

E. Brielle[1], J. Fleisher[2], S. Wynne-Jones[3,4], K. Sirak[1,5], N. Broomandkhoshbacht[1,5], K. Callan[1,5], E. Curtis[1,5], L. Iliev[1,5], A. M. Lawson[1,5], J. Oppenheimer[1,5], L. Qiu[5], K. Stewardson[1,5], J. Workman[1,5], F. Zalzala[1,5], G. Ayodo[6], A. O. Gidna[7], A. Kabiru[8,9], A. Kwekason[7], A. Z. P. Mabulla[10], F. K. Manthi[8], E. Ndiema[8], C. Ogola[8], E. Sawchuk[11,12,13], L. Al-Gazali[14], B. Ali[15], S. Ben-Salem[15], T. Letellier[16], D. Pierron[16], C. Radimilahy[17], J-A. Rakotoarisoa[17], R. L. Raaum[18,19], B. J. Culleton[20], S. Mallick[5,21], N. Rohland[5], N. Patterson[21,1], M. A. Mwenje[22], K. B. Ahmed[23], M. M. Mohamed[24], S. R. Williams[25], J. Monge[26], S. Kusimba[27], M. E. Prendergast[2,5], D. Reich[1,5,21], C. M. Kusimba[8,27,28]; [1]Harvard, Cambridge, MA, [2]Rice Univ., Houston, TX, [3]Univ. of York, York, United Kingdom, [4]Univ. of South Africa, Pretoria, South Africa, [5]Harvard Med. Sch., Boston, MA, [6]Jaramogi Oginga Odinga Univ. of Sci. and Technology, Bondo, Kenya, [7]Natl. Museums of Tanzania, Dar es Salaam, Tanzania, United Republic of, [8]Natl. Museums of Kenya, Nairobi, Kenya, [9]British Inst. of Eastern Africa, Nairobi, Kenya, [10]Univ. of Dar es Salaam, Dar es Salaam, Tanzania, United Republic of, [11]Cleveland Museum of Natural History, Cleveland, OH, [12]Univ. of Alberta, Edmonton, AB, Canada, [13]Stony Brook Univ., Stony Brook, NY, [14]United Arab Emirates Univ., Al Ain, United Arab Emirates, [15]United Arab Emirates Univ., Al-Ain, United Arab Emirates, [16]Université Toulouse III—Paul Sabatier, Toulouse, France, [17]Université d'Antananarivo, Antananarivo, Madagascar, [18]Lehman Coll., CUNY, New York, NY, [19]The New York Consortium in Evolutionary Primatology, New York, NY, [20]The Pennsylvania State Univ., University Park, PA, [21]Broad Inst. of Harvard and MIT, Cambridge, MA, [22]Natl. Museums of Kenya, Lamu, Kenya, [23]Gede Natl. Monument, Gede, Kenya, [24]Fort Jesus Museum, Mombasa, Kenya, [25]Univ. of Illinois at Chicago, Chicago, IL, [26]Univ. of Pennsylvania Museum of Archaeology and Anthropology, Philadelphia, PA, [27]Univ. of South Florida, Tampa, FL, [28]Univ. of Nairobi, Nairobi, Kenya

The urban peoples of the Swahili coast traded across eastern Africa and the Indian Ocean and were among the first practitioners of Islam among sub-Saharan people. The extent to which these early interactions between Africans and non-Africans were accompanied by genetic exchange remains unknown. Here we report ancient DNA data for 80 individuals from 6 medieval and early modern (AD 1250-1800) coastal towns and an inland town after AD 1650. More than half of the DNA of many of the individuals from coastal towns originates from primarily female ancestors from Africa, with a large proportion—and occasionally more than half—of the DNA coming from Asian ancestors. The Asian ancestry includes components associated with Persia and India, with 80-90% of the Asian DNA originating from Persian men. Peoples of African and Asian origins began to mix by about AD 1000, coinciding with the large-scale adoption of Islam. Before about AD 1500, the Southwest Asian ancestry was mainly Persian-related, consistent with the narrative of the Kilwa Chronicle, the oldest history told by people of the Swahili coast. After this time, the sources of DNA became increasingly Arabian, consistent with evidence of growing interactions with southern Arabia. Subsequent interactions with Asian and African people further changed the ancestry of present-day people of the Swahili coast in relation to the medieval individuals whose DNA we sequenced.

Session Title: Evolutionary and Population Genetics Poster Session I

PB3043 Estimating birth prevalence of Fraser syndrome 1 in Quebec.

Authors:

G. Gagnon[1,2], J. Pelletier[1,2], D. Ashton-Beaucage[1,2], M. Sayahian Dekhordi[1,2], V. Chapdelaine[1,2], D. Taliun[1,2], P. Campeau[3,1], C. Bhérer[1,2]; [1]McGill Univ., Montreal, QC, Canada, [2]Victor Phillip Dahdaleh Inst. of Genomic Med. at McGill Univ., Montreal, QC, Canada, [3]CHU Sainte Justine, Montreal, QC, Canada

Fraser syndrome is an autosomal recessive malformation disorder caused by pathogenic variants in three different genes (FRAS1, FREM2, and GRIP1). The main features of this disease include cryptophthalmos, syndactyly, and anomalies of the respiratory and urogenital tract. In our previous work, analysis of pathogenic variants in the French-Canadian population of Quebec revealed that the variant c.370C>T (p.Arg124Ter) in the FRAS1 gene was found at an unusually high frequency in this founder population. This variant has been previously implicated in Fraser syndrome 1 (FS1; MIM 219000). However, no prevalence has been estimated for this syndrome in Quebec, as it has not yet been described in this population. Here we report carrier frequency and prevalence estimates for FS1 in Quebec using whole-genome sequencing (WGS) and genotyping data from the CARTaGENE (CaG) population-based cohort, a multi-ancestry cohort with most participants of French-Canadian ancestry. The p.Arg124Ter variant was the only pathogenic variant previously reported in ClinVar as implicated in FS1 to be identified in CaG. In WGS data (N = 2,157), it is observed in eight European individuals at an allele frequency (AF) of 2.14e-03 (allele count (AC) = 8). By comparing to AF reported in Non-Finnish Europeans from the gnomAD database (v3.1.2), we confirm that this variant is significantly enriched in Quebec (OR = 29.09; p-value = 5.57e-08). Moreover, using CaG genotyping data (N = 24,839), we observe an elevated AF of this variant in some regions of Quebec, suggesting potential regional disease concentrations. Specifically, we find carrier frequencies of 1/96 in Sherbrooke, 1/155 in Quebec City, 1/170 in Trois-Rivières, and 1/173 in Saguenay, while the carrier rate for unrelated individuals of European ancestry is 1/219. Based on AF from CaG WGS and assuming Hardy-Weinberg equilibrium, we estimate a birth prevalence of FS1 in Quebec of 0.34 per 100,000 births. In one mother-child tertiary care center covering 40% of the Quebec population (average 84,000 births per year for Quebec), we identified four cases of FS1 in a decade, which gives a birth prevalence of 0.48 per 100,000 births. Three homozygous cases had the p.Arg124Ter variant, for a birth prevalence of 0.36 per 100,000 births, as predicted from AF in CaG. In summary, our findings provide the first estimation of the birth prevalence of FS1 in Quebec and identify p.Arg124Ter as a candidate founder pathogenic variant in this population. These results have a potential impact on clinical and genetic counselling practices in Quebec.

Session Title: Evolutionary and Population Genetics Poster Session II

PB3044 Estimating the Extra-Pair Paternity Rate in Human Populations: A Case Study of a French-Canadian Population in Quebec

Authors:

G. Martínez; McGill Univ., Montréal, QC, Canada

Large-scale genealogies linked to genetic data offer a unique opportunity to learn about the interplay of demography and genetics. Here, we consider the genotype data of 1464 Y-chromosome-carrying individuals from the CARTaGENE cohort. This cohort includes genome-wide genotype data for 29330 individuals sampled from metropolitan areas in Québec, Canada, of whom 7896 are linked to a deep genealogical database (BALSAC). We used these data to investigate extra-pair paternity (EPP), i.e., the excess mismatch between genetic and genealogical ancestry along the male lineage, mutational processes along the Y chromosome, and relatedness among the first French settlers in Quebec.

In previous studies on European human populations, the per-generation EPP rate has been estimated to be between 1% and 2%. By counting the mismatches between Y chromosome haplogroups and shared genealogical ancestry along the Y chromosome, we obtained a per-generation EPP rate of less than 1%. Accounting for these errors, we estimated the mutation rates along the Y chromosome as a function of paternal age and average relatedness between genealogical founders.

Session Title: Evolutionary and Population Genetics Poster Session III

PB3045 Ethiopian genomic diversity is shaped by migration, admixture, and adaptation to local environments but not by linguistic affiliations

Authors:

T. Mersha[1], D. Sertse[2], J. Habtewold[3], A. Tilahun[4]; [1]Cincinnati Children S Hosp. Med. Ctr., Cincinnati, OH, [2]Dept. of Biology, Univ. of Ottawa, Ottawa, ON, Canada, Ottawa, ON, Canada, [3]Hlth.and Applied Microbiol. Lab., Dept. of Pathobiology, Ontario Vet. Coll., Univ. of Guelph, Ontario, ON, [4]5Dept. of Biology, Univ. of Northern Iowa,, Cedar Falls, IA

Ethiopia is a crossroad of human genetic diversity and religious-cultural hub because of its geographic proximity to the middle east which served as a historical trade route between Asia and Europe as well as to the rest of Africa. To get insight into the genetic landscape of the Horn of Africa against the rest of the world, we performed a genetic diversity and admixture analysis in geographically and linguistically diverse populations in Ethiopia, South Sudan, and Somalia and compared them with global populations data from Utah residents with Northern and Western European ancestry (CEU), Maasai in Kinyawa, Kenya (MKK), and Yoruba in Ibadan, Nigeria (YRI). On average, the ancestry proportion of CEU within Ethiopian gene pool ranges 10-15%. Although the Ethiopian Cushitic, Semitic and North Omotic language speakers harbored high genetic diversity on a global scale, they showed the lowest genetic differentiation among each other. The gradients of genetic variations among different populations within the region are likely to be due to continuous gene flow as a result of admixture driven by past and present intra- and inter- ethnic migration and movements. Individuals speaking different languages but live in the same geographic regions are genetically more related than individuals speaking the same language but isolated by boundary which limits gene flow. The genetic diversity, haplotype and allelic distribution among the populations in this study reflects in part historical movement of people and geographic adaptations to specific environmental factors rather than linguistic affiliations. Here, I will discuss genetic evidence about the relationships among Ethiopians and genetic information underlying adaptations to certain environmental conditions.

Session Title: Evolutionary and Population Genetics Poster Session I

PB3046 Evidence of potential natural selection in African Americans

Authors:

R. Durodoye[1], J. Jaworski[2], J. Bartlett[1], J. Ulloa[3], H. Diaz-Zabala[4], S. Williams[1], M. Aldrich[4]; [1]Case Western Reserve Univ, Cleveland, OH, [2]VUMC, Nashville, TN, [3]Vanderbilt Univ., Nashville, TN, [4]Vanderbilt Univ. Med. Ctr., Nashville, TN

Introduction. Admixed populations are formed when genetically distinct ancestral populations mix. After admixture, highly differentiated alleles can present in frequencies deviating from the average global ancestry. One explanation for this finding is natural selection. The extent of such selection in African Americans remains unresolved as prior work presented conflicting evidence with regard to putative selection. We looked for evidence of non-random patterns of local ancestry, and tested for evidence of selection in a subset of the Southern Community Cohort Study.

Methods. The sample consisted of 1,375 unrelated African American participants previously genotyped on Illumina arrays (1M to 2.5) from across the Southeastern US. After standard quality control metrics 604,224 markers were available for analysis. We used the software Efficient Local Ancestry Inference (ELAI) to measure African (AFR) and European (EUR) ancestry on an overlapping set of ~583K markers. We performed pairwise comparisons to search for regions exhibiting evidence of genome-wide selection. Tests were done using multiple reference AFR and EUR populations from the 1000 Genome reference dataset. RFMix (v1.5.4) was used to confirm results were not sensitive to the analytical method and potential selection was assessed using the R package 'rehh'.

Results. The average genome-wide AFR ancestry among our samples was 84.4%. A region on chromosome 2 showed an excess of EUR ancestry over 5 standard deviations above the genome-wide mean and was observed in all reference comparisons. In addition, regions on chromosome 11 and chromosome 14 also showed excess EUR ancestry in most of the AFR/EUR comparisons. Findings were replicated in the RFmix analysis. None of these regions have been previously reported as deviating from background ancestry. Additionally, integrated haplotype score analyses indicated evidence of potential selection in these regions and identified patterns consistent with selection in our sample that was not observed in ancestral populations.

Conclusion. Our analyses indicate three potential chromosomal regions under selection post admixture in African Americans. Future analyses include identification of nearby genes and pathway enrichment analysis that explain selection in regions of excess EUR ancestry.

Session Title: Evolutionary and Population Genetics Poster Session II

PB3047 Evolution of epigenetic marks: a neutral null model reveals the influence of natural selection and enables dissection of causal underlying pathways.

Authors:

K. Hansen[1], L. Boukas[2], A. Razi[1], H. T. Bjornsson[3]; [1]Johns Hopkins Univ., Baltimore, MD, [2]Children s Natl. Hosp., Washington, DC, [3]Univ. of Iceland, Reykjavik, Iceland

Natural selection acts on phenotypes, and detecting its presence yields profound insights into the evolution of living organisms. Yet, detecting selection on phenotypic traits requires understanding how these traits would evolve in the absence of selection; a neutral null model. This is notoriously challenging. In particular molecular traits other than gene expression have received little to no attention.

Here, we focus on epigenetic marks as an example of a "genome-wide" molecular trait. We formulate a general neutral null model for the evolution of any epigenetic mark associated with genes, and develop a test for selection. Our test respects the foundational aspect of epigenetics: trans-regulation by transcription factors and chromatin modifiers. It leverages estimates of genic intolerance to heterozygous loss-of-function mutations, and the fact that any given mark is observed at multiple genomic sites. Because of the latter, our test is free of many restrictive assumptions and does not require across-species or population-scale data. We illustrate the generality of our approach by applying it to 3 different marks (DNA methylation, H3K4me3, H3K36me3) and a broad range of genomic compartments (promoters, gene bodies, transcriptional end regions). We detect selection on several marks, including an unexpected finding of directional selection on the size of the hypomethylated region around proximal promoters ($p<9.9e-5$), even though this has a negligible correlation with expression.

Our approach also enables accounting for confounders or potential intermediate traits, thereby facilitating investigations into underlying causal processes. First, we find that selection on epigenetic marks is unlikely to be purely a passive consequence of selection on gene expression. This supports a causal role of these marks on gene regulation. Second, we provide evidence (by comparative analyses, trio genomic sequencing data, and simulations) that the selective pressure on DNA methylation and H3K36me3 may partly be explained by their effect on regional mutation rates in the germline, in addition to their involvement in gene regulation. We show that the cumulative effect of mutation rate modification on multiple genomic regions regulated in trans, is strong enough to overcome the drift barrier. Exemplifying the protection of important regions from high mutability, we demonstrate that in humans the more intolerant to loss-of-function mutations a gene is, the lower its coding mutation rate is.

Our framework for selection inference is simple but general, and we anticipate it to be useful for molecular traits beyond epigenetic marks.

Session Title: Evolutionary and Population Genetics Poster Session III

PB3048 Evolutionary history of ancient African to Neanderthal admixture through the analysis of Neanderthal Homologous Regions

Authors:

D. Harris[1], A. Platt[1], M. Hansen[1], S. Fan[2], M. McQuillan[1], W. Beggs[1], T. Nyambo[3], S. W. Mpoloka[4], G. Mokone[4], G. Belay[5], C. N. Fokunang[6], A. K. Njamnshi[6], S. Tishkoff[7]; [1]Univ. of Pennsylvania, Philadelphia, PA, [2]Fudan Univ., Shanghai, China, [3]Kampala Intl. Univ. in Tanzania (KIUT), Dar es Salaam, Tanzania, United Republic of, [4]Univ. of Botswana, Gaborone, Botswana, [5]Addis Ababa Univ., Addis Ababa, Ethiopia, [6]Faculty of Med. and BioMed. Sci., The Univ. of Yaoundé I, Yaoundé, Cameroon, [7]Univ. of Pennsylvania Sch. of Med., Penn Valley, PA

There is strong evidence that anatomically modern humans (AMH) and Neanderthals admixed in Eurasia after the Out-of-Africa expansion. However, recent analyses suggest that there were early AMH migrations from Africa that led to gene flow into Neanderthals, and more recent back migrations of West Eurasians that introduced Neanderthal ancestry to sub-Saharan Africans (SSA). Additional research also revealed extended regions of AMH genomes that are depleted of Neanderthal ancestry (Neanderthal deserts). These deserts are indicative of natural selection acting to remove Neanderthal ancestry due to hybrid incompatibilities. To investigate the admixture and natural selection dynamics between Neanderthals and AMH, we used a method called IBDmix to identify Neanderthal homologous regions (NHRs) in 180 AMH African individuals and developed a statistical model that uses the genetic distance between AMH and Neanderthals to assign the direction of introgression at each NHR as either being from Neanderthal to AMH or AMH to Neanderthal. From the cumulative length of Neanderthal to AMH introgressed NHRs we estimate that SSA individuals' range between 0% and 1% Neanderthal ancestry and that it is significantly positively correlated with their non-SSA ancestry ($R^2 = 0.95$; $p < 2 \times 10^{-16}$). From the cumulative length of AMH to Neanderthal introgressed regions we estimate that 3.79%-4.54% of the Altai Neanderthal genome is of AMH origin. The distribution of the lengths of these regions implies that this AMH introgression occurred ~235,000 years ago. This is long before the out-of-Africa expansion of AMH and may predate the oldest divergences of extant AMH populations (estimated between 150,000 and 285,000 years ago). We found a depletion of AMH introgression in the Neanderthal genome at the locations of Neanderthal deserts in Eurasians. This observation suggests that Neanderthal deserts in AMH genomes are also AMH deserts in Neanderthal genomes, and is consistent with hybrid incompatibilities evolving in similar regions of both AMH and Neanderthal genomes during divergence. We also observed a depletion of AMH introgression in Neanderthal genes, which is consistent with selection against hybrid protein-coding regions. Our analysis of NHRs in a diverse sampling of sub-Saharan Africans adds further support to a model of modern human history that includes: 1) ancient gene flow from AMH to Neanderthals, 2) Neanderthal to AMH gene flow after the out-of-Africa expansion, 3) recent migration and gene flow of non-SSA AMH populations into some SSA populations that introduced small quantities of Neanderthal ancestry. Supported by NIH grants R35GM134957-01 and 5T32DK007314-39.

Session Title: Evolutionary and Population Genetics Poster Session I

PB3049 Evolutionary impact of transposon insertions in ancient humans revealed by machine learning tool*grnTea*

Authors:

Y. Wang[1,2,3], L. Edwards[4], E. Lee[2,1,3]; [1]Harvard Med. Sch., Boston, MA, [2]Boston Children's Hosp., Boston, MA, [3]The Broad Inst. of Harvard and MIT, Cambridge, MA, [4]Massachusetts Inst. of Technology, Cambridge, MA

Detecting transposable element insertions (TEIs) in ancient human DNA (aDNA) can potentially reveal novel genetic mechanisms of human evolution. All existing TEI detection algorithms are designed for modern human genomes, making them unsuitable for TEI detection in highly fragmented and sparsely sequenced aDNA samples. We developed a machine-learning-based tool *grnTea*, that detects TEIs in either selected loci or the entire genome from whole-genome sequencing (WGS) data. We simulated *in silico* TEIs in aDNA WGS data and trained *grnTea* to detect TE families active in humans, *i.e.*, Alu, L1, SVA. Compared to existing methods *MELT* and *xTea*, *grnTea* detects the presence of TEIs with a significantly higher sensitivity and precision, reaching F1-scores of 0.72, 0.82, 0.46 for Alu, L1, SVA, respectively on simulated data. Applications of *grnTea* to published high-coverage Neanderthal and Denisovan WGS data found 862 Alu, 103 L1, and 23 SVA insertions, more than 3 times larger than the published aDNA TEIs while recovering 76.8% of TEIs shared with modern humans. Clustering of archaic humans, along with modern humans and non-human primates based on the TEIs aligns with known genetic relationship. Additionally, our preliminary analyses identified 168 Alu, 26 L1, 1 SVA insertions putatively passed into non-African modern humans through introgression, about 4 times larger than the previous report. These introgressed TEIs (iTEIs) are depleted in exons and intergenic regions, but enriched in 5' UTR, introns, and promoter regions, implying their regulatory roles. Gene ontology analysis shows iTEI enrichment in heterophilic cell-cell adhesion, brain development related processes such as neurogenesis, differentiation, migration, and muscle cell proliferation. Since polymorphic TEIs have been associated with GWAS traits, we mapped iTEIs to GWAS signals and found correlations with obesity-related traits, height, and immune response. Furthermore, we intersected iTEIs with TE-eQTL and TE-sQTL results in modern humans and identified mRNA expression changes in testis and whole blood, and mRNA splicing changes in the brain, skin, heart, whole blood, and adrenal glands. We will expand the QTL analyses with enhanced TEI genotyping in modern humans. We are profiling TEIs in 52 deeply sequenced ancient Eurasians to trace the population allele changes over the last 10,000 years. In summary, we developed *grnTea* as a tool to identify TEIs from ancient WGS data, which will provide valuable TEI resources from 56 ancient humans. Our analyses revealed potential genetic mechanisms of how TEs can impact human evolution.

Session Title: Evolutionary and Population Genetics Poster Session II

PB3050 Evolutionary patterns in historical genomes from Britain

Authors:

K. Anastasiadou, T. Booth, S. Johnston, M. Kelly, J. McCabe, M. Silva, L. Speidel, P. Swali, F. Tait, M. Williams, P. Skoglund; The Francis Crick Inst., London, United Kingdom

Ancient DNA provides a direct way to detect selection and adaptation throughout human history by tracking the trajectories of allelic frequencies over time and identifying changes in genetic variation that can be associated with changes in environmental conditions. A prime example of that is positive selection on lactase persistence in Europe that has been associated with the rise of dairying practices and seems to be driven by a single nucleotide variant, rs4988235. However, the detection of selection on polygenic traits is not as straightforward, as it requires robust genomic data and methods to be able to detect small changes on the whole genome level. In this study, we investigate signals of polygenic evolution with ancient DNA by combining whole genome sequencing, imputation and different approaches with dense sampling and rich phenotypic data, focusing at an initial stage on whole genomes sequenced from more than 200 individuals from Late Medieval (1000-1500 CE) and Post Medieval (1500-1800 CE) Britain, a population with close shared ancestry to the UK Biobank. We estimate polygenic risk scores (PRS) for a number of physiological and anthropometric traits and investigate whether shifts in the distribution of trait values on the population level can be attributed to documented cultural and environmental transitions. We test different methods of PRS calculation and set out a methodological approach of testing for statistically significant shifts in PRS variance over time. Meanwhile, we analyse the PRS in combination with osteological data and explore the use of discordance in genetic and realised value of a trait as an indication for the general health condition of the individual under study. Overall, we expect this study to set the foundations for large-scale studies of polygenic evolution using ancient DNA, providing a new dimension to the interactions between genotype, environment and phenotype over time.

Session Title: Evolutionary and Population Genetics Poster Session III

PB3051 Expanding maps of archaic DNA variants within centromere-spanning haplotypes in complete, telomere-to-telomere (T2T) genomes of modern humans

Authors:

H. Loucks[1], S. Langley[2,3], F. Ryabov[4], J. K. Lucas[5], G. Karpen[2], V. Slon[6], I. A. Alexandrov[6], C. Langley[3], K. H. Miga[5]; [1]Univ. of California Santa Cruz, Santa Cruz, CA, [2]Univ. of California Berkeley, Berkeley, CA, [3]Univ. of California Davis, Davis, CA, [4]Natl. research Univ. Higher Sch. of Economics, Moscor, Russian Federation, [5]UCSC, Santa Cruz, CA, [6]Tel Aviv Univ., Tel Aviv, Israel

Centromeric satellite arrays play an important role in nuclear organization, and kinetochore recruitment and fidelity. Genetic and epigenetic variation within centromeric regions can lead to chromosome instability, which can contribute to cancers and aneuploidies in early development. Despite their importance in cell biology and function, we are only beginning to understand the genetic architecture of these regions and the patterns of evolution across human populations. Previous efforts to study linkage blocks spanning these regions identified centromere-spanning haplotypes, or 'cenhaps'. Meiotic recombination is rare in these regions allowing for intact transmission of entire centromere regions through evolution. Dating based on datasets from the 1000 Genome Consortium (1000G) provided evidence for ancient cenhaps within modern humans. Some of these sequences can be traced back to archaic hominins, and appear to be the result of Neanderthal and Denisovan introgression.
Complete telomere-to-telomere (T2T) reference genomes from humans and non-human primate assemblies offer an unprecedented look into haplotypes and genetic variation within and between cenhaps. Here we present an initial analysis of cenhap structure across the 1000G dataset relative to the first complete human genome (T2T-CHM13). In doing so, we have established a panel of homolog-phased and assembled centromeric regions representing archaic cenhaps on chromosomes 8, 10, 11, and 12. Our initial study revealed new satellite variants of higher order repeats (HOR) that differ in structure and organization from modern cenhaps. Further, we provided refinement of the cenhap structure, uncovered evidence of rare recombination events within peri/centromeric regions, and performed a detailed investigation of genes within archaic cenhaps. The dynamics of repeat expansion, degradation, and mutation are not well understood in these complex regions, and this study provides a new system for studying repeat evolution in the context of shared haplogroup structures. Additionally, this work in its completion aims to provide millions of bases of archaic genetic information to advance our understanding of uncharacterized structural variation in humans.

Session Title: Evolutionary and Population Genetics Poster Session I

PB3052 Exploring recent genetic sharing in the diverse population of Quebec using identity by descent.

Authors:

J. Pelletier[1,2,3], M. Sayahian Dehkordi[1,2,3], A. Diaz-Papkovich[1,3,4], S. Gravel[1,3], D. Taliun[1,2,3], C. Bhérer[1,2,3]; [1]McGill Univ., Montreal, QC, Canada, [2]Canada Excellence Res. Chair in Genomic Med., McGill Univ., Montreal, QC, Canada, [3]Victor Phillip Dahdaleh Inst. of Genomic Med. at McGill Univ., Montreal, QC, Canada, [4]Quantitative Life Sci., McGill Univ., Montreal, QC, Canada

Segments of identity by descent (IBD) reflect genetic sharing between individuals from recent common ancestors. Analyses of IBD patterns in a population allow identification of clusters of individuals with recent genetic sharing, providing insights into relatedness, demographic history and enabling IBD mapping. We leverage IBD segments to study recent genetic sharing in the diverse population of Quebec, Canada. Most of Quebec's genetic studies have focused so far on the well-documented French-Canadian founder population ignoring minor population groups with rich genetic diversity. We used the prospective CARTaGENE population-based cohort, comprising 29,330 participants recruited from urban centers in Quebec and who underwent genome-wide genotyping. We first compared different methods to identify IBD segments shared between all individuals. We utilized iterations of Louvain clustering algorithm in combination with Hudson's FST to identify genetic clusters. We identified 21 clusters, including those of European, African, and Asian ancestry. Combining our IBD clustering results with the self-reported data, we identified eleven non-French Canadian ancestry groups, accounting for more than 10% of the cohort, including Chinese, Filipinos, Haitian, Italian, Latin-American, Lebanese, Moroccan and Vietnamese individuals all of which are known diaspora populations in Quebec, with major migration waves in the 20th century. IBD clustering further identified French-Canadian regional subgroups, replicating known fine-scale patterns of population structure relating to historical migration and subsequent founder effects. We further characterized IBD patterns within genetic clusters. The cluster showing highest average total pairwise IBD sharing length was the French-Canadian cluster representing the Saguenay regional founding group (32.49cM), followed by the cluster of potential Jewish origin (25.66cM) and the cluster of potential acadian origin (22.36cM). To improve our characterization of the ancestry composition of Quebec IBD clusters, we will estimate admixture proportions and compare our results to PCA-based ancestry inference methods and topological approaches such as UMAP and HDBSCAN. Using IBD analysis we identified and characterized population clusters that unveiled recent genetic sharing in the population. We described several population subgroups in CARTaGENE, some of which have been understudied so far. These results will allow us to investigate disease prevalence among genetic clusters and perform IBD mapping, using detailed phenotypic data in CARTaGENE.

Session Title: Evolutionary and Population Genetics Poster Session II

PB3053 Exploring the population history of Khoe-San descendant communities in South Africa

Authors:

S. Edington[1], D. R. Al-Hindi[2], N. Font-Porterias[2,3], A. Surowiec[2], J. Myrick[2], C. Uren[4], P. J. Norman[3], M. Möller[4], B. Henn[2], A. W. Reynolds[1]; [1]Baylor Univ., Waco, TX, [2]UC Davis, Davis, CA, [3]Univ. of Colorado Sch. of Med., Aurora, CO, [4]Stellenbosch Univ., Cape Town, South Africa

The former Cape Colony, centered on the present-day city of Cape Town, South Africa, was one of the first European colonies in Sub-Saharan Africa. Genetic ancestry analyses show that individuals living in Cape Town having substantial amounts of European, East and South Asian, and West African ancestry derived from colonization and the Dutch slave trade. These groups mixed with and absorbed the local Khoekhoe and San peoples during the 17th-18th centuries. However, it is currently unclear how this colonial history impacted the genetic ancestry of communities further into the interior, along the colonial frontier. In this study, we aimed to characterize the impacts of European colonization on the longstanding population history of indigenous populations of southern Africa. We sampled 827 members of Khoe-San and Khoe-San descendant communities across South Africa and genotyped for ~2.2 million genome-wide SNPs on the Illumina H3Africa Array. We analyze global patterns of population structure using ADMIXTURE to infer relative levels of recent admixture from European, Bantu-speaking African, and Asian populations into these communities. Our results demonstrate a cline of Khoe-San ancestry with distance from the Cape Colony. Using local ancestry inference, we estimate the timing of these admixture events for each of the non-Khoe-San ancestries and explored local changes in effective population size. We find differential Indonesian to South Asian ancestries across different communities, reflecting heterogenous migration sources. This project advances our knowledge of population history in Khoe-San and Khoe-San descendent communities for whom few historical records are reliable.

Session Title: Evolutionary and Population Genetics Poster Session III

PB3054 Extension of Viviani's theorem for the evaluation of the impact of population stratification on mating type frequencies

Authors:

R. Sebro; Mayo Clinic Florida, Jacksonville Beach, FL

Abstract
Background: Population stratification occurs when a study population is comprised of several subpopulations where there is random mating within subpopulation but not between subpopulations. Wahlund first showed that population stratification results in a predictable deviation of genotypes from that expected assuming Hardy-Weinberg Equilibrium (HWE). More specifically, the Wahlund effect results in decreased heterozygotes (or alternatively stated an excess homozygotes) in a population with population substructure compared to a population in HWE. The aim of this study was to evaluate how the mating type frequencies are altered in population stratification.
Materials and Methods:
Assume a biallelic marker with minor allele B, and major allele A, where allele A has frequency between 0 and 1. We first assume that there is a population with stratification comprised of n subpopulations, where the allele frequency varies between subpopulation, and evaluate the distribution of mating type frequencies in this population at this locus. Next, we extend Viviani's theorem and use de Finetti diagrams to evaluate the distribution of mating type frequencies in a stratified population.
Results:
Population structure results in another predictable phenomenon - an excess in the number of unions between individuals of the same genotype compared to that expected assuming random mating at all loci where the allele frequency differs between subpopulation. This manifests as assortative mating with respect to subpopulation.
Conclusion: This phenomenon can also be reliably confirmed using an extension of Viviani's theorem. Subpopulation-related assortative mating would be expected in a stratified population.

Session Title: Evolutionary and Population Genetics Poster Session I

PB3055 Finding chrY haplogroup in samples with ultra-low sequencing depth

Authors:

T. Puurand[1], M. Möls[1], L. Kaplinski[1], T. Kivisild[2], M. Remm[1]; [1]Univ. of Tartu, Tartu, Estonia, [2]KU Leuven, Leuven, Belgium

With our current work we aim to show how three diverse types of tandem repeats that correlate with previously described human chromosome Y haplogroups and how hundreds of thousands of short DNA strings (k-mers) combinatorics allow to find chrY haplogroups from WGS data that is sequenced with coverage less than 0.01x without additional PCR. We also prove considerable decrease of tandem repeats motifs copy numbers which indicate decrease of the repeats total length. Our k-mer selection started with creating 25-mers by step of one along three tandem repeats (DYZ1, DYZ2, DYZ3 and DYZ19) in the Y chromosome (CHM13v2.0 assembly). Each region contains k-mers located multiple times inside the repeated region and k-mers located in other chromosomes. After filtering out multiplications in other chromosomes our final count of chrY specific k-mers was 618393. We queried k-mer frequencies from individual's k-mer lists and normalized them with "NIPT" chrY k-mers to set median values. This normalized value should indicate copy numbers of k-mers in the Y chromosome. Note that there is no need to only have 0 or 1 for values. 23 haplogroups of k-mers separation for the training set of 1000G and Estonian Genome Center WGS individuals were based on Mann-Whitney statistic by normalized average of k-mer counts for every k-mer for one SNP based haplogroup (a) and other haplogroups (b). Copy numbers varied from zero to over 100000. Separation of R1a and R1b fails with full set of haplogroups, and this section was performed separately. Result of separation allowed us to choose k-mers with more and less presence in haplogroups. The union of more and less presented k-mers in 23 haplogroups has 406574 entries for the building model. This model calculates haplogroup specific distances based on a specific training set. Simulated dilutions from the training set of 30x WGS data, chrY haplogroup is detectable in 0.001x sequencing depth, samples that were not used in training were detectable at 0.01x. aDNA and cfDNA samples give promising results, but very few samples contain previously known haplogroups. SNP or exon targeted capture of mess sequencing depth prediction which is important to make an accurate prediction.

Session Title: Evolutionary and Population Genetics Poster Session II

PB3056 † Fine-scale population movements and geographic patterns of African Americans over the past 300 years revealed through IBD sharing and pedigree data from 1.7 million people.

Authors:

A. Wolf[1], A. Sedghifar[1], J. Zhang[1], C. Mason[2], P. Milos[2], R. Curtis[2], J. Lang[1], B. Donnelly[1]; [1]Ancestry.com DNA, LLC, San Francisco, CA, [2]Ancestry.com DNA, LLC, Lehi, UT

African Americans constitute the third largest population group in the United States and have a deep interest in exploring their family history. However, African Americans face many barriers to family history research including displacement from, and loss of, formal records—including census, voter lists, and church registries. To facilitate family history research for African Americans, we leveraged our unique collection of DNA and pedigree data to reveal fine-scale movement and geographic patterns of African Americans in the United States over the past 300 years. We used genotype data from approximately 1.7 million research-consented AncestryDNA members to find identity-by-descent (IBD) segments among them. We then recursively ran the *Louvain community detection* method on the resulting IBD networks (individuals as vertices, IBD connections among them as weighted edges) to discover groups of people with significantly more genetic connections to each other than to non-group members. We annotated the network vertices with information from the corresponding individuals' pedigree data. These include surnames, birth dates, and locations of ancestors dating back hundreds of years. The result is a striking set of genetic clusters that map detailed patterns of movement and settlement in the United States over the past 300 years. For example, we are able to visualize population movements from southern states to northern and western states in the mid-1900s that agree with historical records of the Great Migration. As well, we are able to distinguish geographically overlapping groups—such as Louisiana Creole and non-Creole groups—based on differences in genetic ancestry analysis and reported surnames. Overall, our results provide a significant insight into population structure and history of African Americans in the United States. Our results depict both shared historical patterns and varied personal experiences. These resources are an impactful new tool for African Americans to use in their family history research, empowering them to make new personal discoveries, such as pinpointing where their ancestors lived and locating relevant records collections.

Session Title: Evolutionary and Population Genetics Poster Session III

PB3057 Finnish Y chromosome sequencing data suggests dual paths of haplogroup N1a1 into Finland

Authors:

A. Preussner[1], J. Leinonen[1], J. Riikonen[1], M. Pirinen[1,2,3], T. Tukiainen[1]; [1]Inst. for Molecular Med. Finland (FIMM), Helsinki, Finland, [2]Dept. of Publ. Hlth., Faculty of Med., Univ. of Helsinki, Helsinki, Finland, [3]Dept. of Mathematics and Statistics, Univ. of Helsinki, Helsinki, Finland

Y chromosome (chrY) haplogroups differ geographically within Finland, with N1a1-TAT enriched in the northeast and I1-M253 in the southwest, suggested to reflect two separate founding populations. However, to date, the distribution of finer-scale chrY variation beyond these major haplogroups has not been characterized in the population. To better understand the Y-chromosomal landscape in Finland, we studied whole genome sequencing data for the Y chromosome in 1831 Finnish men (born between 1923-1979), with precise geographical origins from the FINRISK project. We assessed the geographical distribution of common chrY haplogroups (freq ≥ 1%) within 19 regions across Finland, determined by paternal birth places (6-392 samples per region), and further examined the relationship between chrY haplogroups and autosomal genetic variation. In addition to detecting the four previously described chrY haplogroups in Finland, N1a1-TAT (69%), I1-M253 (20%), R1a-M420 (5%) and R1b-M343 (4%), we increase the resolution by observing in total 56 common haplogroups in our data, with 27 displaying geographical differences within the country (chisq p<0.05). We show that haplogroup N1a1-TAT, traditionally associated with eastern Finnish ancestry, splits into two unique lineages: N1a1a1a1a2a1-CTS2733 (53%) enriched in the northeast, and N1a1a1a1a1a-VL29 (16%) more frequent in the southwestern coast, specifically in Southwest Finland (64% of N1a1, chisq p=5e-11). Carriers of these two N1a1-TAT lineages displayed differences in their autosomal genome (PC1) in Southwest Finland (wilcoxon p<2e-16) but not elsewhere in Finland, with the VL29 carriers resembling closely southwestern Finns, suggesting its arrival to the southwestern coast may occurred more likely via the Baltic Sea rather than through the mainland. Overall, our results suggest two separate routes of haplogroup N1a1-TAT into Finland, N1a1a1a1a2a1-CTS2733 from the east and N1a1a1a1a1a-VL29 from the southwest. Within other haplogroups, such as I1-M253, we did not observe a similar degree of heterogeneity, although we detected some regional enrichments in its sublineages. In summary, we have characterized detailed chrY variation and its distribution regionally within Finland, providing novel insights into the population history of Finns, particularly within haplogroup N1a1-TAT. Our research further elaborates on the relationship between autosomal genetic structure and chrY haplogroups, providing valuable insights for future studies in population health. Overall, this study highlights the value of reassessing population-level chrY variation with detailed geographical and sequencing data.

Session Title: Evolutionary and Population Genetics Poster Session I

PB3058 † Fundamental limitations on the transferability of polygenic risk scores across human populations.

Authors:

K. Zhang, J. Storey, J. Akey; Princeton Univ., Princeton, NJ

Genome wide association studies (GWAS) have provided insights into the genetic determinants of many traits and diseases and facilitated the construction of polygenic risk scores (PRS) to predict phenotypes in independent samples. However, it is well known that PRS constructed from European cohorts have limited transferability to individuals from other populations. There are many potential factors that contribute to poor transferability of PRS across populations, including differences in patterns of linkage disequilibrium, effect sizes of causal variants, and context specific effects such as gene-gene and gene-environment interactions. Thus, there is considerable interest in delineating the factors that influence transferability so that genomic methods of prediction can be developed that are applicable across populations. It remains unclear, however, whether addressing these potential determinants of transferability will allow predictions to be made across populations, or if there are fundamental limits to transferability. To address this question, we investigated the transferability of PRS in the simplest and most optimal scenario where causal variants and their effect sizes are known and are identical across populations. Using both theory and simulations, we show that transferability is influenced by differences in allele frequencies and non-genetic sources of variation across populations. We further show that even in this simple and ideal setting, transferability can be low across human populations and differences in transferability are a function of the genetic architecture of a trait. Finally, we demonstrate that transferability is improved when the target population has a higher heritability than the discovery population for the trait of interest. Our results provide important insights into how human evolutionary history has imposed fundamental limitations on the transferability of phenotypic predictions across populations.

Session Title: Evolutionary and Population Genetics Poster Session II

PB3059 *FXN* protomutation alleles determine population susceptibility to Friedreich ataxia.

Authors:

S. Bidichandani, M. Tackett, C. Lam; Univ. of Oklahoma Hlth.Sci. Ctr., Oklahoma City, OK

Friedreich ataxia (FRDA) is typically caused by homozygosity for an expanded GAA triplet repeat (GAA-TR) in intron 1 of the *FXN* gene that results in length-dependent epigenetic silencing. Disease-causing expanded (E) alleles have 100-1500 triplets while non-disease alleles show a bimodal allelic distribution of <12 triplets (short normal; SN) and >12 triplets (long normal; LN). A shared haplotype between LN and E alleles supports a multistep process in which a unique SN to LN allele transition was followed by subsequent transition to E alleles. Prevalence of FRDA, which is exclusively seen in Europe, the Middle East, North Africa and South Asia, is determined by the population frequency of asymptomatic heterozygous carriers of E alleles. In order to investigate the disparate population distribution, the entire *FXN* locus (~65 kb) was sequenced in multiple individuals. This identified a 23-SNP core haplotype unique to LN and E alleles in susceptible populations. East Asians lack this haplotype, both in existing people and in ancient DNAs dating back tens of thousands of years, consistent with the absence of FRDA in East Asia. A partial haplotype consisting of 22 of 23 core SNPs was seen in ~10% of sub-Saharan Africans, which was associated with LN alleles consisting of <20 triplets, indicating that the original SN to LN transition occurred in Africa. Completion of the 23-SNP core haplotype and transition to LN alleles with >20 triplets, termed protomutation alleles, occurred exclusively in West Eurasians and South Asians. Finally, E alleles have arisen multiple times from protomutation alleles, indicating that population susceptibility to FRDA is determined by the presence of protomutation alleles.

Session Title: Evolutionary and Population Genetics Poster Session III

PB3060 Genealogy-based ancestry modelling reveals fine-scale genetic history

Authors:

L. Speidel[1,2], P. Skoglund[2]; [1]UCL, London, United Kingdom, [2]Francis Crick Inst., London, United Kingdom

Ancient DNA has provided a new genetic history of humankind, informing our understanding of present-day genetic variation and disease. However, recent genetic history over short time scales is challenging to reconstruct as it often involves fine-scale ancestry differences, often leaving important historical events below the detection threshold. Methods based on sharing of haplotypes, large chromosomal segments and rare variants can improve on power to observe patterns, but can not be readily used in unbiased ancestry models. Here, we develop a framework for ancestry modelling based on recent genealogical history, and show that it can improve statistical power of ancestry models by an order of magnitude. We apply this framework to ancient genomes and present-day individuals in the UK Biobank to develop a detailed ancestry model of Migration period and Viking Age expansions from and into Scandinavia, where previous studies were largely unable to develop quantitative models. These results demonstrate the ability of genealogy-informed analysis to provide insights into fine-scale genetic history.

Session Title: Evolutionary and Population Genetics Poster Session I

PB3061 Genetic affinities of “White and Asian” and “Any other Asian” self-identifying UK Biobank participants.

Authors:

C. Cataldo-Ramirez, T. D. Weaver, B. M. Henn; UC Davis, Davis, CA

It is a well-known limitation in current human genetics research that non-European diversity is under-represented in biobanks which then limits the investigation of genetic architecture for complex traits. This issue is further compounded by the practice of excluding “admixed” participants or individuals who identify with an ethnic label that is ambiguously defined, in an effort to reduce the effects of population stratification on downstream analyses. Here, we characterize the genetic affinities of UK Biobank (UKB) participants who self-identify as “White and Asian” (WA) and “Any Other Asian” (AOA) (n = 2645) by comparing three methods: principal component analysis (PCA), ADMIXTURE, and support vector machines (SVM). All methods were trained or informed by 1000 Genomes Project (1KG) data as a global reference of genetic diversity. Results of the three methods reproduce consistent patterns of clustering. Projecting WA, AOA, and UKB South Asian-identifying participants (Bangladeshi, Indian, Pakistani [n = 6756 post-QC]) onto 1KG PC space identifies at least four clusters within AOA data, while the WA data appear as a cline between European and East Asian groups. The SVM model utilized these PC scores and was trained with 1KG and UKB Pakistani and Bangladeshi groups as reference populations. Classification probabilities of 75% and above resulted in 1878 AOA and WA participants being reassigned into subcontinental genetic groups. This substantially increased the sample size of the UKB South Asian group (by 1432 additional participants). Although human genetic diversity is not biologically discretized by categorical labels, we hope that by clarifying the subcontinental ancestry of these participants and relating their ancestry to other UKB and 1KG groups, research groups can more easily include them in analyses. As demonstrated here with the UKB South Asians, using subcontinental classifications as a starting point has the potential to increase the utility of non-European data, by retaining participants who have been traditionally excluded.

Session Title: Evolutionary and Population Genetics Poster Session III

PB3063 Genetic and environmental contributions to ancestry differences in gene expression in the human brain

Authors:

K. Benjamin[1], Q. Chen[2], N. J. Eagles[2], L. Huuki[2], L. Collado-Torres[3], J. Stolz[2], J. Shin[2], A. Paquola[4], T. Hyde[1], J. Kleinman[1], A. Jaffe[1], S. Han[1], D. R. Weinberger[1]; [1]Lieber Inst. for Brain Dev. / Johns Hopkins Med., Baltimore, MD, [2]Lieber Inst. for Brain Dev., Baltimore, MD, [3]Lieber Inst. for Brain Dev. / Johns Hopkins Univ., Baltimore, MD, [4]Lieber Inst. for Brain Dev. / Johns Hopkins Med., Baltimore, MD

Introduction: Health disparities have endured for centuries. In neuroscience and genomics, individuals with recent African ancestry account for less than 5% of large-scale research cohorts for brain disorders but are 20% more likely to experience a major mental health crisis. Insights gained from genome-wide association studies (GWAS) about disease risk are promising for clinical applications. However, the lack of diversity in GWAS limits the accuracy of genetic risk prediction and hinders the development of effective and equitable neurotherapeutics.
Rationale: While diversity in GWAS has increased in recent years, population-based genetic association studies do not directly elucidate the biological mechanisms of risk variants. Recent efforts to investigate the biological impact of genetic variation on molecular traits of diverse populations have focused on improved fine mapping. These prior studies implicitly assume that the genetic mechanisms of risk and resilience are mostly conserved across ancestries; however, differences in the pathogenic role of ApoE in Alzheimer's disease between individuals of African and European genetic ancestries breaks this assumption.
Results: Here, we examined the impact of genetic ancestry on gene expression and DNA methylation (DNAm) in admixed Black American neurotypical individuals to reduce confounding effects of ancestry-related environmental factors. Ancestry-associated differentially expressed genes (DEGs), transcripts, and gene networks, while notably not implicating neurons, are enriched for genes related to immune response and vascular tissue and explain up to 26% of heritability for ischemic stroke, 27% of heritability for Parkinson's disease, and 30% of heritability for Alzhemier's disease. Ancestry-associated DEGs also show general enrichment for heritability of diverse immune-related traits but depletion for psychiatric-related traits. The cell-type enrichments and direction of effects vary by brain region. These DEGs are less evolutionarily constrained and are largely explained by genetic variations; roughly 15% are predicted by DNAm variation implicating environmental exposures. We also compared Black and White Americans, confirming most of these ancestry-associated DEGs and identifying potentially environmentally associated DEGs.
Conclusion: Our results highlight how environment and genetic background affect genetic ancestry differences in gene expression in the human brain; opening new avenues to the development of ancestry-aware therapeutics and paving the way for equitable, personalized medicine.

Session Title: Evolutionary and Population Genetics Poster Session I

PB3064 Genetic and environmental drivers of telomere length variation in ethnically diverse Africans.

Authors:

M. McQuillan[1], S. Verhulst[2], M. Hansen[1], W. Beggs[3], D. Wolde Meskel[4], G. Belay[4], T. Nyambo[5], S. Mpoloka[6], G. Mokone[6], C. Fokunang[7], A. Njamnshi[7], A. Aviv[8], S. Tishkoff[9]; [1]Univ. of Pennsylvania, Philadelphia, PA, [2]Univ. of Groningen, Groningen, Netherlands, [3]Univ Pennsylvania, Philadelphia, PA, [4]Addis Ababa Univ., Addis Ababa, Ethiopia, [5]Kampala Intl. Univ. in Tanzania, Dares Salaam, Tanzania, United Republic of, [6]Univ. of Botswana, Gaborone, Botswana, [7]The Univ. of Yaoundé I, Yaoundé, Cameroon, [8]Rutgers Univ., Newark, NJ, [9]Univ. of Pennsylvania Sch. of Med., Penn Valley, PA

Leukocyte telomere length (LTL) varies significantly across human populations, with individuals of African ancestry having longer LTL than non-Africans. However, the genetic and environmental drivers of LTL variation in Africans remain largely unknown. We report here on the relationship between telomere length, genetics, and environmental factors in ethnically diverse African individuals (n=1818) originating from Botswana, Tanzania, Ethiopia, and Cameroon. We observe significant variation in telomere length among populations [Population Grouping: $F(17,1800) = 29.79$, $p < 2x10^{-16}$], finding that the San hunter-gatherers from Botswana have the longest telomeres (8.2 ± 0.72 kb; mean ± SD), and Fulani pastoralists from Cameroon have the shortest telomeres (6.9 ± 0.71 kb). Genome-wide association analyses using high coverage whole genome sequences reveal suggestive associations at novel variants located within aging-related genes, as well as at predicted loss-of-function alleles in genes that regulate telomerase. We also show that LTL-associated loci from the UK Biobank explain a significant amount of the inter-individual variation in LTL in Africa, suggesting some level of polygenic score transferability between Europe and Africa for this trait. Moreover, we find that *falciparum* malaria endemicity is significantly associated with LTL while adjusting for age, sex, and genetics. Within Africa, populations indigenous to areas with high malaria exposure have shorter LTL than populations indigenous to areas with low malaria exposure. Together, these results shed light on the genetic and environmental drivers of telomere length variation in under-represented African populations. This work is funded by NIH grant R35 GM134957-01, the American Diabetes Association Pathway to Stop Diabetes grant #1-19-VSN-02, and the Center of Excellence in Environmental Toxicology grant T32-ES019851.

Session Title: Evolutionary and Population Genetics Poster Session III

PB3066 Genetic identification of Slavs in Migration Period Europe using an IBD sharing graph

Authors:

P. Flegontov[1,2], L. Viazov[1], O. Flegontova[1], H. Ringbauer[2], D. Reich[3]; [1]Univ. of Ostrava, Ostrava, Czech Republic, [2]Dept. of Human Evolutionary Biology, Harvard Univ., Cambridge, MA, [3]Harvard Univ., Boston, MA

Popular methods of genetic analysis relying on allele frequencies such as PCA, ADMIXTURE and qpAdm are not suitable for distinguishing many populations that were important historical actors in the Migration Period Europe. For instance, differentiating Slavic, Germanic, and Celtic people is very difficult relying on these methods, but very helpful for archaeologists given a large proportion of graves with no inventory and frequent adoption of a different culture. To overcome these problems and to test if archaeological cultural groups in Early Medieval Central/Eastern Europe correspond to populations isolated to some degree, we applied a method based on autosomal haplotypes. Imputation of missing genotypes, phasing and IBD inference were performed for ancient Eurasians using the ancIBD method (Ringbauer et al. 2023, bioRxiv). IBD links for subsets of these individuals are represented as graphs, visualized with a force-directed layout algorithm, and most clusters of distant relatives that were inferred in these graphs with the Leiden algorithm are in remarkable agreement with archaeological evidence. One of these clusters includes nearly all individuals in the dataset annotated archaeologically as "Slavic". Most individuals in this cluster were not buried in the Slavic cultural context, but all of them come from times and places where Slavic presence is likely according to written sources and/or archaeological evidence. These results illustrate the power of the IBD graph approach since we were able to reveal a community of distantly related individuals from very diverse archaeological contexts who could not be identified using the standard archaeogenetic toolkit. Considering also results of other analyses based on allele frequencies (PCA, ADMIXTURE), a hypothesis for the origin of this population can be proposed: it was formed by admixture of a group related to Baltic speakers with East Germanic people and Sarmatians or Scythians, and underwent a major bottleneck. We suppose that the starting point of these demographic processes in Central and Eastern Europe were migrations of Sarmatian populations, pushed out from Volhynia by Goths in the late 2nd century CE.

Session Title: Evolutionary and Population Genetics Poster Session I

PB3067 Genetic insights into the recent demographic history of Ireland and Britain

Authors:

A. Shanmugam[1,2,3], A. M. Molloy[4], L. Brody[5], M. Merrigan[6], S. O'Reilly[6], G. L. Cavalleri[1,2,3], R. McLaughlin[7,2,3], R. P. Byrne[7], E. Gilbert[1,3]; [1]Royal Coll. of Surgeons in Ireland, Dublin, Ireland, [2]The SFI Ctr. for Res. Training in Genomics Data Sci., Galway, Ireland, [3]The SFI FutureNeuro Res. Ctr., Dublin, Ireland, [4]Sch. of Med., Trinity Coll. Dublin, Dublin, Ireland, [5]NHGRI (NIH), Bethesda, MD, [6]Genealogical Society of Ireland, Dún Laoghaire, Ireland, [7]Trinity Coll. Dublin, Dublin, Ireland

Background: While subtle, discrete clusters of genetic identity across Ireland and Britain have been identified, their demographic history is unclear.
Methods: Using genotype data from 6,574 individuals with associated regional Irish or British ancestry, we identified Irish and British genetic communities using network community detection. We segregated identity-by-descent (IBD) segments by length and (1) estimated temporal genetic relatedness between communities and (2) inferred recent migration rates over the islands across time-scales. Additionally, we estimated changes in effective population over time. For a subset of the Irish haplotypes, we determined the enrichment/depletion of surnames within the genetic communities.
Results: We find evidence of recent population bottlenecks in the Orcadian, Manx and Welsh communities through patterns of IBD-sharing and ROH while the structure in Ireland is subtler - the communities share relatively more IBD segments and the genetic differences between the Irish communities are more subtle on average when compared to the British communities. The regional effective population size trajectories indicate a similar demographic history throughout the island of Ireland. Further, we observe a stable migration corridor between north-east Ireland and south-west Scotland while there is a recent migration barrier between Leinster and Western Ireland. We observed an enrichment of Anglo-Norman and English surnames in the Wexford community while within West Ulster-Argyll community, we saw an enrichment of Gallowglass and Scottish surnames.
Conclusions: Using these new insights into the regional demographic history of Ireland and Britain across different time periods, we hope to understand the driving forces of rare allele frequencies and disease risk association within these populations.

Session Title: Evolutionary and Population Genetics Poster Session II

PB3068 Genetic landscape of HLA in the Emirati population enabled by long-read Oxford Nanopore whole-genome sequencing.

Authors:

J. Quilez Oliete[1], S. Elavalli[1], G. Katagi[1], T. Cardoso[1], S. Behl[1], P. Merel[1], L. Paulin[2], F. Sedlazeck[2], T. Magalhaes[1], F. Al Marzooqi[1], The Emirati Genome Program; [1]G42 Hlth.care, Abu Dhabi, United Arab Emirates, [2]Baylor Coll. Med., Houston, TX

The Human Leukocyte Antigen (HLA) genomic region is biologically and medically relevant. HLA is highly polymorphic and thus difficult to characterize. Besides, most attempts have been based on short-read and targeted sequencing, restricted to only a subset of HLA genes and targeted mostly European-descent populations, failing to represent worldwide genetic diversity in the entire HLA locus.

Here we first show that we can accurately genotype 13 HLA loci from blood-derived long-read Oxford Nanopore Technologies (ONT) whole-genome sequences (WGS). In 10 individuals whole-genome sequenced in both short-read Illumina and ONT, we observed high quality HLA allele calls (mean HLA*LA Q1>0.99) and >99% genotyping concordance at second-digit resolutions between the two sequencing platforms.

We next characterized HLA genetic diversity in the Emirati population using ~10,000 participants of the Emirati Genome Program (EGP) whole-genome sequenced on ONT at ~41X (38-57X). We observed high-quality HLA allele calls supported by 0.96 per-allele mean Q1 (median = 1.00), per-sample mean Q1 >0.90 for most samples and the coverage supporting allele calls (~30-50X) comparable to that observed genome-wide. We confirmed in the Emirati population that Class Ia HLA genes (A, B and C) are more polymorphic than Class Ib genes (E, F and G). The former show a higher number of unique alleles (N = ~30-50) at low similar allele frequency (AF_{mean} = 0.0148; AF_{SD} = 0.0394), in contrast with the more oligomorphic more frequent Class Ib (N = ~2-4; AF_{mean} = 0.3684; AF_{SD} = 0.401). Class II genes (DPA1, DPB1, DQB1, DRB1, DRB3 and DRB4) showed a broader range of unique alleles typically in <25% of the Emirati cohort. Furthermore, we determined that detecting 1-3 most common alleles captures 90% of the population diversity in E, F and G, while a higher and broader range of ~15-30 and 2-25 alleles are needed for Class Ia and II, respectively.

Finally, we investigated the accuracy of HLA typing from ONT WGS of DNA derived from buccal swab and saliva. Such tissues are less invasive than blood and may facilitate establishing routine HLA typing, e.g. clinical practice. Second-digit resolution HLA genotypes were, on average, >85% concordant between blood and buccal swab and less so (>75%) between blood and saliva.

In summary, (i) we demonstrated that ONT WGS can be used to accurately genotype HLA compared to Illumina, which is still considered as default for HLA typing; (ii) we generated high-quality HLA genotypes in ~10,000 Emirati ONT WGS and pioneered the analysis of HLA genetic diversity in Emiratis; and (iii) we provided a first step in enabling routine HLA typing from non-invasive biological samples.

Session Title: Evolutionary and Population Genetics Poster Session III

PB3069 Genetic network rewiring between distantly related eukaryotic organisms.

Authors:

V. Pereira[1], R. Dandage[1], E. Kuzmin[1,2]; [1]Dept. of Biology, Ctr. for Applied Synthetic Biology, Ctr. for Structural and Functional Genomics, Concordia Univ., Montreal, QC, Canada, [2]Dept. of Human Genetics, Rosalind & Morris Goodman Cancer Inst., McGill Univ., Montreal, QC, Canada

Synthetic lethality represents an extreme example of a genetic interaction that occurs when a combination of mutations in different genes results in lethality, which would not be expected from the combined effects of individual viable single mutants. The extent of genetic interaction network conservation differs from genome sequence conservation between species. Two distantly related eukaryotic organisms such as the two yeast species, the budding yeast, *S. cerevisiae*, and fission yeast, *S. pombe*, diverged ~500 Mya and despite 75% genome conservation, they display 29% genetic interaction network conservation. Other distantly related eukaryotes such as *C. elegans* and *H. sapiens* diverged ~600 Mya. Here, we investigate genetic network rewiring by studying the genetic interactions that underlie conditional essentiality of single mutants between *S. cerevisiae*, *S. pombe*, *C. elegans*, and *H. sapiens*, whereby a gene is essential (ES) in one species but nonessential (NES) in another. We have extensively studied the 2853 *S. cerevisiae* - *S. pombe* orthologs, where ~15% are conditional ES (ES in one species but NES in the other). From 269 conditional NES genes that were ES in *S. pombe* and NES in *S. cerevisiae*, we identified 124 cases, which are rewired by synthetic lethal digenic interactions that modify conditional NES single mutants to synthetic lethal double mutants in *S. cerevisiae*. Single mutant fitness, phenotype rate and genetic interaction degree differentiate conditional NES genes that were rewired by synthetic lethal interactions compared to those which are nonewired. To understand the functional relationship between conditional NES genes and their rewiring synthetic lethal interactions, we overlapped them with common functional standards and found that they are functionally related since they were co-expressed, co-localized, co-annotated, shared protein-protein interactions and showed similar phenotypic profiles suggesting that genetic rewiring of essential genes is local. We have begun to extend these findings to *C. elegans* and *H. sapiens*, where ~21% of *H. sapiens* - *S. cerevisiae* orthologs, ~23% of *C. elegan*s - *S. cerevisiae* orthologs, and ~14% *H. sapiens* - *C. elegans* orthologs are conditional ES. Mapping how these conditional ES genes are rewired by genetic interactions will provides insight into genetic network rewiring between distantly related eukaryotes and should reveal principles of genetic network conservation which can shed light on human diseases, such as cancers, and potential synthetic lethal therapeutic strategies.

Session Title: Evolutionary and Population Genetics Poster Session I

PB3070 Genome-wide characterization of de novo tandem repeat mutations in the human genome

Authors:

T. Mokveld[1], E. Dolzhenko[1], H. Dashnow[2], B. van der Sanden[3], B. Pedersen[2], Z. Kronenberg[1], T. Nicholas[2], C. Fanslow[1], C. Lambert[1], N. Koundinya[4], W. Harvey[4], K. Hoekzema[4], J. Knuth[4], G. Garcia[4], K. M. Munson[4], S. Watkins[2], D. W. Neklason[2], A. R. Quinlan[2], C. Gilissen[3], A. Hoischen[3], E. E. Eichler[4], M. A. Eberle[1]; [1]PacBio, Menlo Park, CA, [2]Univ. of Utah, Salt Lake City, UT, [3]Radboudumc, Nijmegen, Netherlands, [4]Univ. of Washington, Seattle, WA

Tandem repeat (TR) expansions at more than 60 loci have been associated with severe diseases, including Fragile X syndrome and Huntington disease. Many of these diseases initially arise in families by a *de novo* TR expansion and worsen in later generations via subsequent expansions and genetic anticipation. Studies aimed at identifying novel pathogenic repeats have been hampered by limitations of short-read sequencing data and shortcomings of existing informatics pipelines. To enable rare disease researchers to comprehensively profile the length and sequence context of TRs in long read sequencing data, we previously developed the Tandem Repeat Genotyping Tool (TRGT). In our trio assessment, TRGT achieves >98% mendelian length consistency for repeats with motifs longer than 2bp when applied to 1 million TR regions. However, the accuracy rate remains insufficient for detecting true *de novo* TR mutations due to the potential >10,000 false positives per trio. Therefore, accurate detection of *de novo* TR mutations remains an open problem.

With this goal in mind, we have formulated a new method, TRGTdn, that works in tandem with TRGT to call *de novo* TR mutations, including expansions and contractions, in parent-offspring trios using long read HiFi data. TRGTdn relies on analyzing familial dissimilarities, which are established by pedigree-aware realignment of sequencing data. Crucially, we significantly enhance accuracy by detecting allele dropout through phasing.

We applied TRGTdn to the four-generational kindred 1463 CEPH pedigree sequenced with PacBio HiFi long reads and Illumina short reads. This multigenerational assessment enabled us to estimate the *de novo* TR mutation rate per generation at >30-60 calls per trio within the profiled TR loci. Mutation lengths ranged from a 2bp expansion to a 260bp contraction. Our analysis of CEPH individual 2189 confirmed 33 of 34 candidate *de novo* mutations through inheritance by at least 1 of their 6 children. Manual inspection of short-read data validated 17 short and structurally simple repeats in this individual.

Additionally, we used TRGTdn to characterize *de novo* mutations in 8 rare disease trios, observing mutation rates comparable to the CEPH pedigree trios. We will present results of our in-depth validation of these events and the characterizations of their sizes, motifs, and mutation types.

In conclusion, TRGTdn is an effective workflow for *de novo* TR mutation detection and characterization, one of the most difficult - and understudied - variant classes in the human genome. The uptake of this approach will enhance our understanding of the structural dynamics of the human genome and the role of TRs in rare genetic diseases.

Session Title: Evolutionary and Population Genetics Poster Session II

PB3071 Genome-wide data from medieval German Jews show that the Ashkenazi founder event pre-dated the 14th century.

Authors:

S. Waldman[1], K. Sczech[2], S. Carmi[3], D. Reich[4,5,6]; [1]Harvard Univ., Newton, MA, [2]City of Erfurt, Erfurt, Germany, [3]The Hebrew Univ. of Jerusalem, Jerusalem, Israel, [4]Harvard Univ., Boston, MA, [5]Harvard Med. Sch., Boston, MA, [6]Broad Inst. of MIT and Harvard, Cambridge, MA

We report genome-wide data from 33 Ashkenazi Jews (AJ), dated to the 14th century, obtained following a salvage excavation at the medieval Jewish cemetery of Erfurt, Germany. The Erfurt individuals are genetically similar to modern AJ, but they show more variability in Eastern European-related ancestry than modern AJ. A third of the Erfurt individuals carried a mitochondrial lineage common in modern AJ and eight carried pathogenic variants known to affect AJ today. These observations, together with high levels of runs of homozygosity, suggest that the Erfurt community had already experienced the major reduction in size that affected modern AJ. The Erfurt bottleneck was more severe, implying substructure in medieval AJ. Overall, our results suggest that the AJ founder event and the acquisition of the main sources of ancestry pre-dated the 14th century and highlight late medieval genetic heterogeneity no longer present in modern AJ.
Our main emphasis here will be on the methodology used to analyze the history of the AJ population in this study. Additionally, I will discuss new data that has been collected since the publication of this study.

Session Title: Evolutionary and Population Genetics Poster Session III

PB3072 Genome-Wide Exploration of Positively Selected Loci and Their Association to Disease Phenotypes in 30,000 Individuals from Sri Lanka and Bangladesh

Authors:

M. Kapoor[1], J. Backman[1], M. Riaz[1], Regeneron Genetics Center, M. Mridha, Kanti[2], A. Kasturiratne[3], J. Chambers[4], A. Shuldiner[1]; [1]Regeneron Pharmaceuticals, Tarrytown, NY, [2]BRAC Univ., Dhaka, Bangladesh, [3]Univ. of Kelaniya, Kelaniya, Sri Lanka, [4]Imperial Coll. London, London, United Kingdom

Introduction: The study of diverse global populations is crucial in understanding the role of genetic factors that contribute to their adaptation and survival in distinct geographical regions with unique selective pressures. These genetic factors may illuminate biological mechanisms relevant to human health, including those that incur protection from disease as potential targets for therapeutic development.**Methodology:** The South Asia Biobank (SAB) includes rich baseline demographic, lifestyle, clinical, environmental, and phenotypic data, biological samples and longitudinal follow-up from more than 150,000 South Asian participants. Whole exome sequencing and target sequencing-based genotyping (Twist platform) was performed in first tranche of 30,000 subjects as part of an ongoing collaboration between the Regeneron Genetics Center and SAB. Here, we conducted a genome-wide scan of selection signatures in SAB by estimating locus-specific time to the most recent common ancestor using the ascertained sequentially Markovian coalescent (ASMC) model. **Results:** Principal components and Uniform Manifold Approximation and Projection (UMAP) based analyses mapped each participant to distinct groups or clusters which were highly correlated with self-reported geographical location and local language. A genome-wide scan of selection signatures identified 40 positively selected loci (32 novel) satisfying genome-wide significance. Strong selection signatures were observed at several known loci including *HBB* (thalassemia),*SLC24A5* (skin pigmentation), *HLA* (immune system), and *ENPP7* (metabolism). We also performed similar analyses on Europeans, Africans, East and South Asians from UKBB and identified several positively selected loci unique to each population. Interestingly, we observed enrichment for genes involved in zinc homeostasis that are important in context of South Asians given adaptations to zinc scarcity in soil and diet. Phenome-wide association and enrichment analysis of novel selected loci indicated strong associations with a variety of complex phenotypes including anthropometric traits, immune system and metabolic disorders. **Conclusion:** We conducted a largest genome-wide scan to identify naturally selected loci in South Asians. Our findings underscore the importance of such studies and their potential to yield valuable insights into human adaptation, evolution and population genetics with implications for health and disease.

Session Title: Evolutionary and Population Genetics Poster Session I

PB3073 Genomic history of Armenian population.

Authors:

A. Hovhannisyan[1], A. Manica[2]; [1]Trinity Coll. Dublin, Dublin, Ireland, [2]Univ. of Cambridge, Cambridge, United Kingdom

Armenians have remained underrepresented in modern population genomic studies, despite inhabiting an important region between Western and Eastern Eurasia. Using the first full genomes on Armenians, we studied their origins, population structure, and demographic history. We demonstrated that the Armenian populations from western, central, and eastern parts of the highlands are relatively homogeneous. The Sasun, a population in the south that had been argued to receive the major genetic contribution from Assyrians, was instead shown to have derived its slightly divergent genetic profile from a bottleneck occurred in the recent past. We investigated the debated question on the genetic origin of Armenians and failed to find any significant support for historical suggestions by Herodotus of their Balkan-related ancestry. We also checked the degree of continuity of the population and found that Armenians have remained unadmixed through the Neolithic and at least until the first part of the Bronze Age. However, by incorporating temporally spaced ancient genomes from Armenia, we detected a genetic input from a source linked to Levantine Early Farmers during, or just after the Late Bronze Age, indicating that this period has been marked by large-scale migrations in the whole region.

Session Title: Evolutionary and Population Genetics Poster Session II

PB3074 Geography and study design in the biobank era: a population genetics-informed approach to the discovery of rare deleterious variants in human populations

Authors:

M. Steiner[1], D. Rice[2], C. Porras[1], J. Novembre[2]; [1]Univ. of Chicago, Chicago, IL, [2]Univ Chicago, Chicago, IL

The discovery of rare, large effect alleles and their phenotypic associations in human populations is critical to our understanding of biological mechanisms and an increasingly important component of the drug discovery pipeline. The availability of human whole genome sequencing (WGS) data at biobank scale provides an unprecedented opportunity to characterize rare variants. Given the growing number of biobanks being established globally and investments available for human genetic sequencing, one may consider whether it is advantageous to sequence as many individuals as possible in one geographic location (which we refer to as "narrow sampling") or to sequence individuals from as many geographic locations as possible ("broad sampling"). This question is particularly interesting in the context of rare deleterious alleles, which are known to be clustered in geographic space near their area of origin due to underlying evolutionary processes. With this motivation in mind, we present a novel theoretical framework to study the impact of spatial sampling design on the discovery of rare variants. In particular, we develop a population genetic model for the distribution of carriers of deleterious alleles in a structured population - accounting for dispersal, drift, selection, mutation, and uneven spatial sampling simultaneously. Analysis of our model suggests that, for a fixed sample size, broader sampling will result in the capturing of a greater number of deleterious variants, but each variant will be at reduced frequency within the sample; in contrast, narrow sampling yields fewer variants but each is found at higher frequency within the sample. To investigate how these different discovery profiles impact the power to discover associations between phenotypes and rare deleterious variants, we then integrate our work with statistical genetic models to identify trade-offs for genome-wide association studies (GWAS) and burden test studies. We find that under our model, the effect of spatial sampling design on GWAS and burden test power depends on several interconnected factors related to the genetic architecture, including effect sizes of the variants and spatial autocorrelation in the environmental component of the trait. Lastly, we validate key results using WGS data from the UK Biobank. We anticipate that our quantitative analyses will help to inform a broader discussion regarding best practices for study design in the biobank era.

Session Title: Evolutionary and Population Genetics Poster Session III

PB3075 Identification of medically actionable variation in admixed individuals from Mexico

Authors:

C. Gonzaga-Jauregui, T. Sepulveda-Morales, I. A. Perez-Alvarado, M. A. Arciniega-Sanchez, C. Van Hout; Intl. Lab. for Human Genome Res., Univ. Natl. Autónoma de México (LIIGH-UNAM), Queretaro, Mexico

The field of human genomics suffers from unequal representation and diversity. Studies of European ancestry individuals comprise about 90% of all genomic data available, whilst less than 0.5% of genomic studies have been performed in Latin American individuals. In 2022 Ziyatdinov et al. reported the sequencing of 141,046 exomes and 9,950 genomes from Mexican individuals from the Mexico City Prospective Study (MCPS), the largest genomic sequencing study in a non-European ancestry population. Along with the preprint, aggregated frequency data on this dataset was publicly released.

To assess the potential health impact on individuals from this population and evaluate the benefit of performing genomic sequencing studies in underrepresented populations, we used the publicly available data for the MCPS to identify medically actionable variation in this poorly genomically characterized population. Using the list of genes curated and published by the ACMG for reporting of secondary findings that are medically actionable, we have identified 382 directly sequenced variants in MCPS impacting 52 of the 73 genes in the ACMG SF v3.0 list. The five genes with the greatest number of actionable variants were *BRCA2*, *BRCA1*, *GAA*, *MLH1*, *LDLR*, and *RYR1*. However, the greatest number of individuals were impacted by variants in *BRCA1*, *LDLR*, *BRCA2*, *TMEM43* and *DSG2*, suggesting a substantial risk for breast and ovarian cancer, familial hypercholesterolemia, and arrhythmogenic right ventricular cardiomyopathy in this population. Additionally, we identified novel loss-of-function likely pathogenic variants in these medically actionable genes that are unique to the Mexican population and not reported elsewhere to date. Overall, we estimate that about 5% of the individuals in this cohort are carriers for a medically actionable variant where early screening and therapeutic interventions may result in improved health outcomes and reduced mortality.

Our analyses present the best estimate of medically actionable variation that may be expected to be found in admixed Mexican individuals. Consistent with other studies and similar analyses, we estimate that approximately 2% to 5% of individuals from outbred populations will carry variants of medical relevance that warrant early interventions and may be useful to guide public health policies.

Session Title: Evolutionary and Population Genetics Poster Session I

PB3076 Identifying and interpreting lineage-specific fixed differences between modern humans, archaic humans, and non-human primates.

Authors:

D. Tagore, R. Bierman, J. M. Akey; Lewis-Sigler Inst. for Integrative Genomics, Princeton Univ., Princeton, NJ

The explosion of whole-genome sequence data from contemporary individuals and ancient DNA is enabling high-resolution catalogs of lineage-specific fixed differences between modern humans, archaic humans (Neanderthals and Denisovans), and non-human primates to be constructed. These catalogs contain the genetic substrates that define uniquely hominin and human phenotypes and the chronology of evolutionary changes that distinguish hominins from non-human primates and modern humans from archaic humans. Although the interpretation of lineage-specific fixed differences remains daunting, the development of powerful functional genomics resources is enabling novel approaches to prioritize, identify, and test hypotheses about putatively causal genetic changes between lineages. Here, we describe an updated and densely annotated catalog of fixed differences between modern humans, archaic humans, and non-human primates from over 500,000 whole-genome sequences derived from multiple large-scale sequencing projects and sequence data we have generated from understudied human populations. As expected, the predicted number of functionally important lineage-specific fixed differences is larger for regulatory versus protein-coding DNA. Through integrating myriad functional genomics data from ENCODE, NIH Roadmap Epigenomics, GTEx, and CRISPR screens, we identify lineage-specific fixed differences between hominins and non-human primates and between modern humans and archaic humans that we predict perturb specific regulatory networks and changes in gene expression. We also leverage our high-resolution catalog of lineage-specific fixed differences to test hypotheses about the evolution of compensatory changes, estimate rates of back mutations and predict their functional and phenotypic consequences, and interpret deserts of archaic ancestry in modern humans. In summary, we will describe a carefully curated and powerful resource of lineage-specific fixed differences that provides novel insights into the genetic changes over the past ~15 million years of human evolutionary history.

Session Title: Evolutionary and Population Genetics Poster Session II

PB3077 Identifying health disparities and differential healthcare usage within the biobank at the Colorado Center for Personalized Medicine using identity-by-descent clustering

Authors:

M. Fisher[1], J. Shortt[1], N. Rafaels[2], S. Chavan[3], N. Pozdeyev[3], H. Smith[1], C. Caggiano[4], R. Shemirani[5], N. Zaitlen[4], V. Arboleda[4], E. Kenny[6], K. Crooks[7], C. Gignoux[8]; [1]Univ. of Colorado Anschutz Med. Campus, Aurora, CO, [2]Univ. of Colorado Sch. of Med., Aurora, CO, [3]Univ. of Colorado, Aurora, CO, [4]UCLA, Los Angeles, CA, [5]Icahn Sch. of Med. at Mount Sinai, Palo Alto, CA, [6]Icahn Sch. of Med. at Mount Sinai, New York, NY, [7]Univ Colorado, Aurora, CO, [8]U Colorado, Anschutz Med. Campus, Aurora, CO

Population-scale biobanks increasingly power insights into both the genetic and environmental factors that influence health. Here, we use a non-parametric, data-driven approach to identify high-resolution population substructures among 73,346 participants in the Colorado Center for Personalized Medicine (CCPM) biobank, use these structures to better characterize the risk of disease, and discuss how this information can be used to understand the factors that influence health. We identified 148,142,648 identity by descent (IBD) segments indicative of recent shared ancestry using identity by descent LocAlity Sensitive Hashing (iLASH) between pairs of unrelated individuals enrolled in CCPM. We clustered individuals using cumulative IBD sharing with iterative Louvain clustering to identify 71 distinct genetic communities. Genetic communities have both overlapping and distinct genetic similarity profiles compared to global reference panels. Using linked electronic health record (EHR) information to derive 1563 phecodes, we investigate associations between genetic communities and different phecodes, generating hypotheses on health disparities experienced by specific genetic communities. At 5% FDR, we identified 395 associations between 270 unique phecodes and 60 unique genetic communities. Interestingly, we identified dozens of associations between phecodes and communities where communities with similar genetic similarity profiles had opposite directions of effect for the same phecode, including skin cancer, type 2 diabetes, and nonhypertensive congestive heart failure. Importantly, we believe that while genetic relatedness can identify this population structure, different environmental factors are the primary drivers of differential risk between communities. These findings offer new insights into health disparities and healthcare usage within the CCPM population.

Session Title: Evolutionary and Population Genetics Poster Session III

PB3078 † Identifying mutator alleles in humans based on a proxy genome-wide association study

Authors:

V. Getseva[1], I. Agarwal[1,2], M. Przeworski[1]; [1]Columbia Univ., New York, NY, [2]Univ. of Oxford, Oxford, United Kingdom

Germline mutation rates are known to vary across individuals, families, and populations, and to evolve over time, but the causes of these differences remain poorly understood. In humans, the variance in the number of de novo single nucleotide variants is largely explained by parental ages but appears to have additional sources as well. Moreover, rare “mutator alleles” that increase mutation rates have recently been mapped in mice, rhesus macaques, and human trios that were ascertained for offspring with severe disease phenotypes. These findings raise the possibility that mutator alleles contribute non-negligibly to inter-individual variation in mutation rates in humans. To evaluate this possibility, we conducted a form of proxy genome-wide association study (GWAS) in whole exome sequences of “white British” individuals from the UK Biobank, a sample that includes a large number of close relatives as well as more distantly related individuals. Specifically, we considered putatively-neutral mutations seen only in one individual in the sample (i.e., “private” or singleton mutations) as a proxy for individual mutation rates. As we show, these private mutations, while not all de novo, nonetheless accrued quite recently (typically, over the last several generations). To map modifiers, we then test for association between the number and type of private mutations and the loss of function of ~11K genes using a mixed linear model. In parallel, we consider whether missense and loss of function mutations in known DNA replication and repair genes sit in a genomic region of increased diversity, as expected if they are mutators. Thus, we provide the first genome-wide survey of mutator alleles in humans.

Session Title: Evolutionary and Population Genetics Poster Session I

PB3079 Identifying Sex-by-Ancestry Interactions of Asthma in African Descent Admixed Individuals

Authors:

K. Ellis[1], T. Miller Fleming[1], P. Straub[1], J. Jaworski[1], V. E. Kerchberger[1], K. Actkins[2], E. G. Atkinson[3], L. K. Davis[1]; [1]Vanderbilt Univ. Med. Ctr., Nashville, TN, [2]NIH, Cary, NC, [3]Baylor Coll. of Med., Houston, TX

Asthma is heterogeneous, complex, and demonstrates both sex and ancestry disparities in the United States. The prevalence ratio for women to men is 1.7:1, and African Americans experience asthma at higher rates compared to their white counterparts. Notably, Black women have the highest rate of asthma diagnoses and mortality. This disparity is thought to be due to a combination of ancestry specific genetic risk factors, racial health disparities, and gender disparities in social determinants of health. Prior studies indicate that African descent admixed individuals bear greater and disparate global genetic risk for asthma than non-African descent admixed individuals. Effects of asthma risk variants among African descent admixed individuals may also vary, by local ancestry and by sex, as local ancestry estimates vary widely according to geographic origin of African descent and demographic histories. Despite the awareness of how sex and ancestry separately influence asthma susceptibility, existing literature fails to examine the interaction between these two risk factors. Moreover, existing literature often fails to thoroughly examine asthma susceptibility in African descent admixed individuals, leaving a detrimental gap in the literature that stalls equitable precision medicine advancements and widens health disparities. **We hypothesize that sex moderates the effect of ancestry-specific X-chromosome risk factors for asthma susceptibility.** By leveraging genetic and phenotypic data from the *All of Us* Research Program, our cohort consists of 49,005 (4,845 cases, 44,160 controls) African descent admixed individuals, with 28,199 individuals having two X chromosomes and 20,806 individuals having one X chromosome and one Y chromosome. We aim to quantify the independent, joint, and interaction effects of sex and genetic ancestry on asthma diagnosis in African descent admixed individuals. We will characterize local ancestry across the X-chromosome to jointly test local ancestry and risk allele counts associated with asthma. We will then fit the resulting sex-stratified covariate-adjusted effect estimates for SNP risk alleles and local ancestry to logistic regression models. Results will be shared along with lessons learned regarding the unique challenges of local ancestry assignment on the X chromosome in a large biobank sample.

Session Title: Evolutionary and Population Genetics Poster Session II

PB3080 Incorporating functional information into ancient DNA based selection scans.

Authors:

L. Poyraz[1], L. Colbran[2], I. Mathieson[2]; [1]Columbia Univ., New York, NY, [2]Univ. of Pennsylvania, Philadelphia, PA

Ancient DNA can directly reveal the contribution of natural selection to human genomic variation. However, while ancient DNA based methods have been successful in identifying genomic signals of selection, inferring the phenotypic effects of that selection has been difficult. Evidence from genome-wide association studies suggests that most trait-associated variants are non-coding, indicating that a large proportion of the phenotypic effects of selection will also act through non-coding variation and the transcriptome.

Since we cannot measure gene expression directly in ancient individuals, we used an approach (PrediXcan) developed to predict gene expression from genotype data. We developed this approach into a gene-level test for regulatory shifts due to selection across a time transect of 4500 years. We predicted the expression of 17,834 protein coding genes for each individual in a population of 91 modern and 684 ancient individuals from Britain using PrediXcan models trained on the genotypes and transcriptomes of 49 tissues from the GTEx project. Based on linear regression models of expression against time, 36 genes (FDR < 0.05) had evidence for significant shifts in expression levels. We compared the results from our transcriptome-wide scan to a genome-wide scan based on estimating per-SNP selection coefficients from time series data, and to an approach based on combining classical selection statistics (iHS and SDS) with PrediXcan models.

We highlight several genes with directional regulatory change, including the previously known targets of selection LCT and FADS1, and other genes in those regions. The transcriptome-wide scan identified 14 genes out of 150 in the genomic selection peak in the HLA region. Moreover, we identified a signal for reduced expression of *OAS3*, corroborating experimental evidence for an association between the adaptively introgressed Neandertal haplotype in the *OAS* region and the reduced expression of *OAS3* in response to viral immune triggers. Finally, we found 10 novel signals of selection on gene expression not captured by scans based on genomic signatures of selection that do not incorporate functional information. These results demonstrate the potential of combining this information and ancient DNA to elucidate the consequences of selection and highlight the fact that most of the gene expression changes resulting from selection are of genes that are not themselves the direct targets of selection.

Session Title: Evolutionary and Population Genetics Poster Session III

PB3081 Integration of Gene-Level Annotation and Visualization in the Updated Functional Annotation of Variants Online Resource (FAVOR)

Authors:

H. Zhou[1], V. Verma[1], X. Li[1], L. Zilin[1], S. McCabe[2], E. Van Buren[1], X. Lin[3]; [1]Harvard, Boston, MA, [2]Indiana Univ. Sch. of Med., Indianapolis, IN, [3]Harvard T.H. Chan Sch Pub Hlth, Boston, MA

Rapid advancements in large-scale whole genome sequencing (WGS) studies conducted on biobanks have led to the identification of numerous coding and non-coding variants. These studies offer an unparalleled resource for gaining insights into the genetic underpinnings of human diseases. In this context, variant functional annotations play a crucial role in the analysis of WGS data, interpretation of results, and prioritization of causal variants associated with diseases or traits. Furthermore, understanding gene functions becomes essential due to the strong relationship between variant functions and the genes they impact. We develop a new version of the Functional Annotation of Variants Online Resources (FAVOR). This new version of FAVOR offers an extensive and multifaceted online portal for variant and gene functional annotation. The latest version of FAVOR offers a comprehensive summary and visualization of findings for all nine billion possible single nucleotide variants (SNVs) across the entire genome, as well as for all well-studied human genes. In addition to supporting rapid variant-, gene-, and region-level queries of variant functional annotations, the new version of FAVOR also provides highly detailed gene-level functional annotation. This means that users can delve into comprehensive information about the functional impact of variants on specific genes, enabling a thorough analysis and understanding of gene-level effects. FAVOR now offers an enhanced platform that facilitates both broad-scale queries and in-depth exploration of variant and gene functional annotations. The updated version of FAVOR are available at https://favor.genohub.org.

Session Title: Evolutionary and Population Genetics Poster Session I

PB3082 Introgression of structural variation in human *OCA2* informs archaic hominins pigmentation.

Authors:

K. Kim, P. Hallast, F. Yilmaz, C. Lee; The Jackson Lab. for Genomic Med., Farmington, CT

Over the last decade, there is growing evidence that part of the modern human gene pool originated from now-extinct, archaic hominins (AH) such as Neanderthal and Denisovan. However, most of the evidence has been in the form of single-nucleotide variations (SNVs) due to the inherent limitations of short-read sequencing as well as the degraded nature of AH genomes. Here, we explore AH-originated structural variations (SVs) by genotyping SVs in two high-coverage AH genomes (Altai Neanderthal and Denisovan) and comparing them to the comprehensive SV call sets from the 1000 Genomes Project and the Human Genome Diversity Project. We identified 153 AH-originated SVs that are absent in both African and great ape genomes, but present in both AH and non-African genomes. 49 of these SVs were not detectable using SNV information alone due to their weak linkage disequilibrium with surrounding SNVs. Among the 153 potentially introgressed SVs, we found a Denisovan-originated intronic insertion in *OCA2*, a gene that has been shown to play a role in human pigmentation including iris, skin, and hair color. The insertion has an ~64% allele frequency in Bougainville, one of the Solomon Islands populations, and less than 18% frequency in other human populations. The insertion in *OCA2* is linked (genotype correlation r^2>0.6) to hair color and skin-associated SNVs in the Genotype-Tissue Expression (GTEx) project and the UK biobank data. Given that indigenous people with blonde hair and dark skin are commonly observed in the Solomon Islands, these results suggest a new allele that may explain pigmentation variation in the area and give us insights into the hair color of Denisovans.

Session Title: Evolutionary and Population Genetics Poster Session II

PB3083 Investigating Mitochondrial Lineages in a Historic Afro-descendant Peruvian Population with Ancient DNA

Authors:

L. Pott[1], J. Zolik[1], J. L. Santa Cruz Alcalá[2,3,4], C. Maass[5], M. A. Nieves-Colón[1]; [1]Univ. of Minnesota, Minneapolis, MN, [2]Univ. Natl. Mayor de San Marcos, Lima, Peru, [3]APACHETA Consultores, Lima, Peru, [4]Museo Virtual de la Cultura Afro Peruana, San Luis de Cañete, Peru, [5]Defense POW/MIA Accounting Agency, Washington, DC

Human populations in Latin America have a long history of voluntary and forced migration from Europe, Asia, and Africa, which created highly complex patterns of admixture. However, Afro-descendant populations remain understudied. Between the 15th and 19th centuries, European colonizers forcibly transported over 12 million Africans to the Americas through the Transatlantic Slave Trade. Around 500 Africans were brought to the *Hacienda La Quebrada* sugar plantation in San Luis District, Peru. Recent excavations uncovered 245 individuals in an 18th-19th century burial ground for the enslaved people who worked on the plantation and their descendants. We are collaborating with archaeologists and community members in a community-engaged research project to learn more about patterns of genetic diversity among the people buried in this cemetery. In one of the first ancient DNA (aDNA) studies of an Afro-diasporic community in Peru, we assess the maternal lineages of these individuals. To investigate the preservation of skeletal remains in the cemetery and their suitability for aDNA analysis, we identified a subset of individuals (n=30) with good macroscopic tooth preservation and sampled dentin from them. aDNA extractions were successful, and we detected quantifiable DNA in 28 out of 30 samples. Following shotgun sequencing of 28 single-stranded libraries, sequenced reads exhibit characteristic patterns of aDNA damage that authenticate them as ancient. We recovered nuclear genomes at an average depth of 0.00071X and mitochondrial genomes at an average depth of 2.3X. The proportion of endogenous DNA varied greatly among the samples. Two samples with excellent preservation had an endogenous DNA content over 1% and will be candidates for target enrichment and deeper sequencing. We also identified the mitochondrial haplogroups of three individuals with >5X mitochondrial coverage, revealing diversity in their maternal ancestries. Two individuals belong to haplogroup subclades B2 and C1b, which are commonly found among Indigenous Americans. One individual belongs to subclade L2c2a1, which is associated with southern populations in Africa and the Bantu expansion. Future work will include nuclear and mitochondrial target enrichment to characterize additional haplogroups, as well as to identify the biogeographical origins and admixture patterns of these individuals. This work will contribute to a better understanding of the genomic impacts of slavery and the demographic history of this understudied Afro-descendant population.

Session Title: Evolutionary and Population Genetics Poster Session III

PB3084 Investigating the contrast in ROH between Newfoundland founder population and ancestral British-Irish population based on 1000 Genomes data

Authors:

S. Demiriz, M. E. MacMillan, T. Furneaux, B. Ballew; Sequence Bio, St. John's, NL, Canada

The Newfoundland population exhibits characteristics of a founder effect due to a genetic bottleneck caused by the initial settlement of the island by British and Irish settlers circa 1600. Due to factors including a lack of easy means of travel around the island and cultural expectations, settlers mostly lived and married within the same town and religion, resulting in the preservation of genetically isolated subpopulations.
Runs of homozygosity (ROH) are two contiguous homozygous segments of a common ancestor in an individual, which enables us to evaluate the relative distance of a population from its ancestral origin. However, ROH data of any given population carries meaning only when juxtaposed with that of other populations. The 1000 Genomes project, with its collection of genomic data primarily from Europe and other continents, allowed us to compare our findings in Newfoundland against it.
As in other populations with founder characteristics, the Newfoundland population shows differences in ROH lengths compared to other European populations. Our aims in this study were to (i) highlight the close connection between Newfoundland and its genetic heritage from its British-Irish ancestry; (ii) contrast the lengths of Newfoundland ROH with those of populations included in the 1000 Genomes project; and (iii) investigate the extent of the Newfoundland founder effect by comparing it with other known founder populations such as the Finnish. Finally, we attempted to define the three main subpopulations of Newfoundland - Irish, English and admixed - based on available data such as PCA, reported religion, and ancestry, and investigate any correlations with each subpopulation’s respective ROH patterns.
We assessed the effects of various computational methods such as measuring ROH with and without LD pruning, whether to filter on MAF and what threshold to use, and compared orthogonal methods of ROH calculation via functionality from different software packages. The entire process is configured to run reproducibly and at scale, and can be easily adapted to display different sets of comparisons with minimal changes in configuration.

Session Title: Evolutionary and Population Genetics Poster Session I

PB3085 † Investigating the ecological suicide ("ecocide") theory in Rapa Nui with ancient DNA data

Authors:

B. Sousa Da Mota[1], J. Moreno-Mayar[2], T. Higham[3], S. Klemm[2], M. Gorman Edmunds[4], F. Torres Hochstetter[5], M. Friess[6], J. Stenderup[2], M. I. Orbegozo[2], D. I. Cruz Dávalos[1], V. Laborde[6], E. Heyer[6], M. E. Allentoft[7], H. Schroeder[2], O. Delaneau[1], A-S. Malaspinas[1]; [1]Univ. of Lausanne, Lausanne, Switzerland, [2]Univ. of Copenhagen, Copenhagen, Denmark, [3]Univ. of Vienna, Vienna, Austria, [4]Ka Haka Hoki Mai Te Mana Tupuna, Hanga Roa, Rapa Nui, Chile, [5]Rapa Nui Museum (previous affiliation), Hanga Roa, Rapa Nui, Chile, [6]Musée Natl. d'Histoire Naturelle, Paris, France, [7]Curtin Univ., Bentley, Australia

Rapa Nui (Easter Island) is a speck of land in the Pacific Ocean, at the easternmost vertex of the Polynesian triangle. Polynesians peopled the island between 1,500 and 800 years ago, which is famous for its massive stone statues, the moai. However, the history of its people has been told time and time again as an example of ecological suicide ("ecocide"), with resources overexploitation that led to deforestation, famine, wars and population collapse in the 1600s. Compared to other Polynesian islands, natural resources are scarce, biodiversity is low, and most of the land is barren without trees. Although it has been shown that the island was once covered by a dense forest, evidence against the ecocide is growing. Firstly, the Polynesian rats (*Rattus exulans*) are the most likely culprit of deforestation. Secondly, it is disputed that such environmental change led to a population collapse in the light of archaeological and demographic studies. However, up until now, genetics has been absent from this debate. Here, we use ancient DNA to investigate a potential collapse of the Rapanui population in the 1600s. We sequenced the genomes of 15 ancient Rapanui, radiocarbon dated to the beginning of the 19th century, thus pre-dating the 1860's slave trade raids that resulted in a *de facto* genocide. We showed that they had a Polynesian origin, with 10% of Native American gene flow dated to the 15th century. We found no close relatives (up to 3rd degree) among these ancient Rapanui and estimated very low amounts of inbreeding, despite a large proportion of their genomes being in runs of homozygosity. Our estimates of recent effective population size (Ne) suggest that the Rapanui were a small population (less than 3000 individuals), which steadily grew after peopling the island, with no evidence of a second strong bottleneck associated with the ecocide. To further characterize the population history, we simulated genomes that replicated the Rapanui population history. We employed one- or two-bottleneck demographic models, with the former representing the island founding and the latter corresponding to a potential population collapse. By performing a permutation test, we rejected that weak and strong second bottlenecks generated data that were similarly distant to the observed Ne (1,000,000 permutations, p-value=0). Together with the analysis of distance metric values, these findings suggest that a second strong bottleneck is inconsistent with the observed data. Altogether, these results add to the mounting evidence that the Rapanui history is an example of human resilience and ingenuity in the face of environmental changes, and not the most well-known case of self-inflicted collapse.

Session Title: Evolutionary and Population Genetics Poster Session II

PB3086 Leveraging Large Genealogies for Realistic Population Genetics Simulations.

Authors:

L. Anderson-Trocme[1], S. Zabad[2], A. Wohns[3], K. Tayeb[1], S. Gravel[2], J. Novembre[1]; [1]Univ. of Chicago, Chicago, IL, [2]McGill Univ., Montreal, QC, Canada, [3]Broad Inst., Cambridge, MA

Population scale genealogies encode the process by which individuals in a population move through space and time. Although well-studied models such as the Wright-Fisher model create genealogies that capture the mean and variance in offspring number, they do not capture the idiosyncratic dynamics of real populations. To move beyond this, we took advantage of the French Canadian pedigree -- with over 5M digitized records, it is one of the largest and most complete spatially linked population scale genealogies available for research.

Anderson-Trocmé et al. (2023) illustrated how genealogy-aware simulations can accurately capture real population structure. However, there remains a wide gap between theoretical models used to understand evolution and empirical genealogies, which encode a wide range of demographic effects.

To address this gap, we computed demographic parameters (e.g. population sizes, migration rates, etc.) from the French Canadian genealogy and encoded them into the Demes format so as to be readily compatible with popular tools like SLiM and msprime. Furthermore, we developed a forwards-in-time genealogy simulation framework that accommodates additional features of complex assortative mating preferences, such as age disparity and relatedness rates.

The combination of empirical demographic parameters and our simulation framework enables an unprecedented ability to tease apart the relative influence demographic factors have on broader population structure. Moreover, because we use the Demes format, we can directly compare simulation tools. Finally, because the French Canadian genealogy is spatially resolved, we explore deme models at various scales of geographic resolutions to gain insights into concepts of effective populations.

In conclusion, this project allows us to assess the accuracy of standard population genetic assumptions in large cohorts. Our work bridges an important gap between empirical populations and theoretical models through a more nuanced understanding of how subtle demographic forces shape population structure at both fine and broad scales.

Session Title: Evolutionary and Population Genetics Poster Session III

PB3087 Lifetime reproductive success in 14 million sibling pairs reveals novel genetic insights into human complex diseases

Authors:

A. Liu[1,2,3,4], E. T. Akimova[5], Z. Yang[1], G. Genovese[3,6,4], T. Fall[7], M. C. Mills[5], A. Ganna[1,8,3,4]; [1]Inst. for Molecular Med. Finland, Helsinki, Finland, [2]Analytic and Translational Genetics Unit, Massachusetts Gen. Hosp., Boston, MA, [3]Program in Med. and Population Genetics, Broad Inst. of Harvard and MIT, Cambridge, MA, [4]Stanley Ctr. for Psychiatric Res., Broad Inst. of Harvard and MIT, Cambridge, MA, [5]Leverhulme Ctr. for Demographic Sci. and Nuffield Coll., Univ. of Oxford, Oxford, United Kingdom, [6]Dept. of Genetics, Harvard Med. Sch., Boston, MA, [7]Molecular Epidemiology, Dept. of Med. Sci., and Sci. for Life Lab., Uppsala Univ., Uppsala, Sweden, [8]Analytic and Translational Genetics Unit, Massachusetts Gen. Hosp.,, Boston, MA

Sibling similarities have long been used to quantify the genetic contribution to phenotypic variability, the so-called heritability. When suggested to be heritable, for a common polygenic disease GWAS is employed to identify its associated common variants, and for a Mendelian or monogenic disease, a rare-variant association analysis, with one disease being studied per analysis. Sibling similarities in lifetime reproductive success (LRS), however, offer a unique opportunity to systematically reveal the genetic architecture of many diseases in a hypothesis-free manner. Leveraging 14 million well-phenotyped sibling pairs available in the Finnish and Swedish nationwide multi-generation registry, we propose a novel framework aiming to answer how complex the disease architecture is, genetically (G) and environmentally (E). For a total of 1,646 diseases (14 categories), we quantified the LRS of disease-affected individuals ($LRS_{affected}$) and their unaffected siblings ($LRS_{unaffected}$) separately for each sex and then grouped diseases based on the estimated LRS. We hypothesized that: (a) If a disease largely reduces $LRS_{affected}$ but has no impact on $LRS_{unaffected}$, it's likely to be driven by G (eg, de novo mutation, monogenic variants) or E that are specific to the affected individuals; (b) If a disease slightly reduces both $LRS_{affected}$ and $LRS_{unaffected}$, it's likely to be driven by G (polygenic variants) or E that are shared between siblings; (c) If a disease reduces $LRS_{affected}$ but increases $LRS_{unaffected}$, it's under balancing selection possibly because of G by E; and (d) if a disease exhibits different patterns in males and females, we then consider this disease is selected in a sex-dependent manner. We saw 40 diseases in males and 65 in females largely reduced $LRS_{affected}$. However, unlike model (a) assuming no change in $LRS_{unaffected}$, for many such diseases, $LRS_{unaffected}$ was significantly reduced as well (model (b)), suggesting that many Mendelian diseases might also be impacted by polygenic variations. Balancing selection is seen for several diseases (model (c)) and is usually coupled with a sex-specific manner (model (d)). For example, alcohol dependence significantly reduced LRS in affected men and women and unaffected brothers, but with favorable LRS in unaffected sisters. Diseases with lower LRS are generally rarer in prevalence (a consequence of contemporary selection) and more heritable, especially in men. By studying fitness in 14 million sibling pairs from two nations, we inferred a comprehensive picture of the complex architecture of human diseases. The findings of this work can be used to inform the design of future genetic studies.

Session Title: Evolutionary and Population Genetics Poster Session I

PB3088 Local ancestry inference in the MHC locus.

Authors:

H. Andrade[1], C. Passos[1], K. Nunes[1], C. Dinardo[2], A. Ferraz[3], C. Teixeira[4], S. Garcia[5], B. Custer[6,7], E. Sabino[8], S. Kelly[6,9], D. Meyer[1]; [1]Univ.e de São Paulo, São Paulo, Brazil, [2]Fundação Pró-sangue, Hemocentro de São Paulo, São Paulo, Brazil, [3]HEMOPE, Recife, Brazil, [4]Hemominas, Belo Horizonte, Brazil, [5]HEMORIO, Rio de Janeiro, Brazil, [6]Epidemiology, Vitalant Res. Inst., San Francisco, CA, [7]Dept. of Lab. Med., Univ. of California San Francisco, San Francisco, CA, [8]Inst. de Med. Tropical, Departamento de Moléstias Infecciosas e Parasitárias da Faculdade de Med. da Univ.e de São Paulo, São Pulo, Brazil, [9]Univ. of California San Francisco Benioff Children's Hosp. Oakland, Oakland, CA

Using methods that assign continental-level ancestries to genomic segments (i.e., Local Ancestry Inference LAI), studies identified an excess of African ancestry in the MHC locus in human populations (e.g., Mexicans, Colombians). A hypothesis to explain this finding is that African ancestry in the MHC is favored in these admixed populations. However, the MHC has features that raise concerns about the reliability of LAI: difficulty in phasing and extremely high (and extensively shared) SNP-level polymorphism. This has led to concerns that LAI calls for the MHC may be biased to overestimate the most diverse parental population. Here, we analyze the reliability and reproducibility of LAI using different subsets of SNPs, to measure the impact of features of the data upon downstream findings. We used the whole genome sequencing data of 2629 individuals from the Recipient Epidemiology and Donor Evaluation Study-III Brazil Sickle Cell Disease Cohort and applied the Gnomix tool to infer LAI. We compared LAI that: (a) included all the 400,000 SNPs in chromosome 6 (WGS data); (b) included only a subset of 30,000 6 SNPs in chromosome 6, corresponding to those in the Axiom Transfusion Medicine Array (Array data); (c) removed SNPs within HLA genes, or in their flanking regions (which have potential calling errors). For regions outside the MHC, ancestry estimates are unaltered when using any filter (46.8% African, 45,1% European, and 8.1% Native American in all scenarios). However, the results for the MHC are extremely sensitive to the inclusion criteria: for the lower SNP density Array analysis (regardless of the inclusion of HLA SNPs), we find an excess of African ancestry in the MHC (mean of 51.4% AFR, 40.4% EUR, and 8.2% NAM), whereas with high SNP density WGS data (with or without the HLA SNPs) we no longer find an excess of African ancestry in the MHC (46.5% AFR, 45% EUR, and 8.5% NAM, similar to values for the whole chromosome). We compared the WGS and the array set of SNPs for other chromosomes, and only the MHC presented marked differences in ancestry pattern, reaching a difference of 8% in African ancestry depending on the inclusion criteria. Because of the extremely high density of SNPs in MHC, analyses that define windows by a fixed number of variants will have unusually small windows in the MHC (50% smaller in the WGS set and 10% in the array set), with higher non-independence between the SNPs, potentially contributing to unreliable estimation of LAI in the MHC region. While our approach showed that LAI in the MHC is highly sensitive to SNP density, our lack of a gold standard means we cannot associate the different densities to accuracies. Simulation approaches will be key to answer this question.

Session Title: Evolutionary and Population Genetics Poster Session II

PB3089 Local Ancestry-Aware Genotype Principal Component Analysis on Chronic Kidney Disease GWAS signals.

Authors:

W. Gonzalez-Rivera, P. Woo-Yeong, K. Frazer, M. Gymrek, M. D'Antonio; Univ. of California, San Diego, San Diego, CA

Genomic research has traditionally underrepresented the genetic diversity of the United States, particularly the Latinx population, which has resulted in significant biases and limitations in understanding the genetic basis of diseases and traits. Challenges arise in including Latinx individuals in genetic studies because of their complex admixed genomic structure. While previous studies have leveraged local ancestry inference to better model admixed individuals in association studies, this approach relies on using a number of predefined ancestry labels from a reference panel. Here, we developed Local Ancestry-Aware Genotype Principal Component Analysis (LA-GPCA), which instead uses a quantitative notion that improves the detection of genetic variants associated with diseases and traits in admixed populations.

As a test case, we conducted a genome wide association study (GWAS) using global ancestry principal components (PCs) as covariates on the UK Biobank for Chronic Kidney Disease (CKD) and Glomerular Filtration Rate (eGFR), which measures kidney function. From this analysis, we identified significant signals on chromosomes 15 and 4 in multiple populations. To further explore these signals, we applied LA-GPCA, which involved extracting a 1 Mb region surrounding each of the lead variants, rs2453533 and rs4859682, respectively in the chromosome 15 and 4 loci. Subsequently, we applied a pruning technique by analyzing linkage disequilibrium (LD) patterns within windows of 50kb, shifting the window by 10 variants after each analysis, and pruning variants with an $R^2 \geq 0.1$, performing a PC analysis on the remaining variants to obtain LA-GPCs. We next repeated our GWAS on these two regions using both global genotype PCs and LA-GPCs and compared the summary statistics with the original GWAS. Our analysis showed that chromosome 15 locus harbors GWAS signals specific to the EUR, SAS and AFR populations, whereas the signal in chromosome 4 locus was shared across all populations.

We are further exploring refining LA-GPCA as an approach to analyze local haplotype structures in genetic association studies. The goal is to develop a method for conducting genetic studies in diverse and admixed populations that is independent of assigning local ancestry labels but rather deconvolutes the local pattern of common genetic variations into shared and divergent haplotypes which can be directly utilized in GWAS. Removing ancestry labels and developing an ancestry-agnostic GWAS could overcome limitations in studying admixed populations, leading to the identification of universally associated genetic variants for disease and traits across diverse populations.

Session Title: Evolutionary and Population Genetics Poster Session III

PB3090 LoGicAl: Local ancestry and genotype calling uncertainty adjusted ancestry-specific allele frequency estimation from admixed populations.

Authors:

J. Wang, S. Zollner; Univ. of Michigan, Ann Arbor, MI

Knowing allele frequency distributions across different populations provides important insights into population-enriched genetic drivers of disease etiologies, design of replication genome-wide association studies, transferability of polygenic risk scores, and interpopulation demographic histories. The allele frequency itself is insufficient to describe the heterogeneous patterns of genetic variation across different admixed populations since allele frequency does not account for their complex population histories and inherent heterogeneous ancestry background. Existing computational programs developed to deconstruct allele frequencies of the underlying ancestral populations from admixed populations are typically based on "best-guess" local ancestry and genotype, thereby ignoring uncertainty in upstream local ancestry and genotype calling steps, which will eventually bias the estimates.

Here, we introduce LoGicAl, a novel ancestry-specific allele frequency estimation tool, which adjusts for both ancestry calling uncertainty and genotype calling uncertainty. LoGicAl models probabilistic local ancestry and genotypes while accounting for phasing uncertainties with a EM algorithm. It informs ancestry-specific allele frequency estimates by averaging alternate allele dosages across each individual, which are defined by probabilistic local ancestry calls, genotype likelihoods and allele frequency distributions. LoGicAl is capable of interpreting large-scale (N > 10,000) phased or unphased multi-way admixture data generated from array-based or sequence-based genotyping techniques.

In simulation studies, we evaluated the impact of sample size, allele frequency, population structure, and uncertainty in genotyping and ancestry calls on estimates obtained from LoGicAl and from methods not adjusting for the uncertainties. The results demonstrate high estimation accuracy of LoGicAl in both array-based and low-coverage or high-coverage sequence-based genotyping data with different degrees of ancestry calling uncertainty (Pearson's R with true values > 0.999). The simulation studies also demonstrate that ignoring the uncertainties inflates bias in estimating the allele frequencies 12-fold on average, even with large sample size (~10K). In the absence of uncertainties, LoGicAl reduces running time by 66% over other methods while maintaining estimation accuracy and reduces bias for rare varaints. We provide a new, computationally efficient program that provides precise estimates of ancestral allele frequencies, thus improving the genomic analysis of admixed populations.

Session Title: Evolutionary and Population Genetics Poster Session I

PB3091 Measuring constraint against missense variation in the context of AlphaFold2 protein structures

Authors:

M. Silk[1,2], L. Thibaut[1,2], Genome Aggregation Database (gnomAD) Consortium, K. M. de Lange[1,2], D. G. MacArthur[1,2,3]; [1]Garvan Inst. of Med. Res., Sydney, Australia, [2]Murdoch Children's Res. Inst., Melbourne, Australia, [3]Univ. of New South Wales, Sydney, Australia

Catalogues of genetic variation such as gnomAD and UK Biobank have proven invaluable in the diagnosis of rare disease. Beyond simply identifying patient-ascertained variants that are rare or absent in the population, these databases can also be used to identify genes and sub-regions of genes depleted of variation (constrained), and therefore likely functionally important. Missense constraint is of particular interest in healthcare due to the diverse effects missense variants can have on protein structure and function, and the difficulty in interpreting their consequences.

Currently, the majority of region-specific missense constraint measurements are applied to linear DNA sequence. However, genetic variation mapped to the protein tertiary structure can uncover additional functionally important regions, as distant parts of a gene sequence often exist close together in the protein tertiary structure. Advances in protein structure prediction, including AlphaFold2 and ESMFold, have massively expanded on the proportion of the proteome with a valid, high-confidence structure prediction, allowing constraint to be assessed across the full proteome.

We present Protein Observed/Expected missense constraint in 3D structures (PrOEmis3D), a measure of constraint to missense variation across 18,609 AlphaFold2 protein tertiary structures using the gnomAD v2 database of over 125,000 human exomes. Constrained regions are dynamically defined based on genomic context using well-established mutational models that correct for genomic context, sequence coverage and CpG methylation. We show that regions with low PrOEmis3D scores (constrained) are highly enriched for ClinVar variants compared with regions with high PrOEmis3D scores (tolerant) (Mann Whitney U test; $p < 10^{-50}$). These measurements are accurate across proteins of varying sizes and configurations. Additionally, measurements are calculated for each codon within the protein's genetic sequence, providing direct and easily accessible clinical utility.

Session Title: Evolutionary and Population Genetics Poster Session II

PB3092 Microevolutionary processes analysis in the human genome

Authors:

A. Urnikyte, L. Pranckeniene, V. Kucinskas; Vilnius Univ., Vilnius, Lithuania

Introduction: Differences in the relative fitness of genomic variants are foundational, without these, neither natural selection nor adaption can exist. This research aimed to analyze two major microevolutionary forces, mutations and positive selection, using the whole genome sequencing data of Lithuanians collected from two generations. Here, we aim to answer the main question - how commonly are mutations under selection in modern humans, and how do allele frequencies change from generation to generation. Material and methods: The detection of mutations and the signatures of positive selection and relative fitness across different age groups, including newborns (generation I), their parents (generation II), and the root ancestor (considered as a generation of the 1000 genome project (generation III)), was performed using whole genome sequencing data. Signs of positive natural selection were infered using Tajima's D, FST and cross-population extended haplotype homozygosity (XP-EHH) statistical methods. Selected SNPs were annotated using ANNOVAR in GRCh37 (hg19), RefSeqGene, dbSNP151 and CADD v.1.3. Visual Studio 2017 and C# language were used to write the calculation software for relative fitness, and a graphical presentation and analysis of the results was performed using the Rcmdr and ggplot packages.Results: This study demonstrated that when exploring the landscapes of relative fitness on each chromosome, the overall pattern of relative fitness background remains consistent across both generations. However, a general tendency of relative fitness decrease was observed. We hypothesize that the newly formed genome variants or those with a very low frequency from the previous generation had insufficient time to be significantly influenced by natural selection. Consequently, in the subsequent generation, the force of natural selection acting upon them is stronger, resulting in a decrease in their cumulative relative fitness. Furthermore, the strong natural selection pressure on the genetic regions that encode the NEGR1 and PTPN1/PTNP21 genes were also identified, highlighting the evolution of the Lithuanian population’s genome over generations, and possible genomic “deficiencies” for better adaptation. Conclusions: This research contributes to the advancement of scientific knowledge and holds particular significance in elucidating the relationship between natural selection and adaptation.

Session Title: Evolutionary and Population Genetics Poster Session III

PB3093 Mixed Turkic and Mongolic ancestry reveals complex history and origin of the Tu and other Chinese populations.

Authors:

D. Tabin[1], N. Patterson[2], T. Zeng[1], D. Reich[3]; [1]Harvard, Dept. of Human Evolutionary biology, Cambrdige, MA, [2]Broad Inst MIT & Harvard, Cambridge, MA, [3]Harvard Univ, Boston, MA

Turkic and Mongolic speakers have had a long and complex history of interaction with each other and with other neighboring populations. Due in part to this complex history, there is debate about whether or not the two groups share a common linguistic origin. We provide new genetic evidence regarding these patterns by leveraging the fact that Turkic and Mongolic peoples have differential sources of both West and East Eurasian ancestry, which we show can be used to disentangle Turkic and Mongolic ancestry in admixed populations. We find variable proportions of Turkic- and Mongolic-associate ancestry in modern Chinese Turko-Mongolic ethnic minorities including the Salar, Bonan, Dongxiang, Daur, Mongola, and Tu. Most have ancestry from both groups with notable exceptions such as the Salar who are consistent with having no Mongolic ancestry. While Chinese connections to and interactions with the Eurasian Steppe can be seen as early as the Shang Dynasty, genetic dating of the most recent wave of admixture suggests that the modern mixture of ancestries present in these populations were formed during the heyday of the Mongol empire. A plausible scenario is that these groups were formed by Turkic and Mongolic speakers, possibly soldiers, settling and admixing with Chinese and Tibetan peoples. Many of these admixed populations would have then undergone a language shift towards the Mongolic prestige language, a scenario that would be consistent with the joint genetic, linguistic, and historical evidence.

Session Title: Evolutionary and Population Genetics Poster Session I

PB3094 Molecular insights from telomere-to-telomere assemblies of diverse human Y chromosomes across 185,000 years of evolution

Authors:

P. Hallast[1], P. Ebert[2], M. Loftus[3], F. Yilmaz[1], P. Audano[1], G. Logsdon[4], M. Bonder[5], W. Zhou[6], W. Höps[7], K. Kim[1], C. Li[8], S. Hoyt[9], P. Dishuck[4], D. Porubsky[4], F. Tsetsos[1], Q. Zhu[1], K. Munson[4], P. Hasenfeld[7], W. Harvey[10], A. Lewis[4], K. Hoekzema[4], Human Genome Structural Variation Consortium (HGSVC), R. O'Neill[11], J. Korbel[7], C. Tyler-Smith[12], E. Eichler[13], X. Shi[8], C. Beck[14], T. Marschall[15], M. Konkel[3], C. Lee[1]; [1]The Jackson Lab. for Genomic Med., Farmington, CT, [2]Inst. for Med. Biometry and Bioinformatics, Core Unit Bioinformatics, Ctr. for Digital Med., Med. Faculty, Heinrich Heine Univ., Dusseldorf, Germany, [3]Dept. of Genetics & Biochemistry, Ctr. for Human Genetics, Clemson Univ., Clemson, SC, [4]Univ. of Washington Sch. of Med., Dept. of Genome Sci., Seattle, WA, [5]German Cancer Res. Ctr. (DKFZ), Div. of Computational Genomics and Systems Genetics, Heidelberg, Germany, [6]Univ. of Michigan, ANN ARBOR, MI, [7]European Molecular Biology Lab. (EMBL), Heidelberg, Germany, [8]Temple Univ., Dept. of Computer and Information Sci., Philadelphia, PA, [9]Univ. of Connecticut, Dept. of Molecular and Cell Biology, Storrs, CT, [10]7Univ. of Washington Sch. of Med., Dept. of Genome Sci., Seattle, WA, [11]Dept. of Molecular and Cell Biology, Inst. for Systems Genomics, Univ. of Connecticut, The Univ. of Connecticut Hlth.Ctr., Storrs, CT, [12]Wellcome Trust Sanger Inst, Hinxton, United Kingdom, [13]Univ. of Washington Sch. of Med., Dept. of Genome Sci., Howard Hughes Med. Inst., Seattle, WA, [14]The Jackson Lab. and UConn Hlth., Farmington, CT, [15]Inst. for Med. Biometry and Bioinformatics, Ctr. for Digital Med., Med. Faculty, Heinrich Heine Univ., Dusseldorf, Germany

The complex sequence composition of mammalian Y chromosomes, containing dense repeats and high-copy duplications, has led to their systematic omission from genomic analyses. While more than half of the Y chromosome is missing from the GRCh38 human reference (approx. 31 Mbp), the first complete human Y has recently been assembled in by the Telomere-to-Telomere (T2T) Consortium (Rhie et al. manuscript under review). To improve our understanding of human Y chromosome evolution and genetic diversity, we assembled an additional 43 Y chromosomes by combining high-coverage PacBio HiFi and Oxford Nanopore ultra-long sequence data from the Human Genome Structural Variation Consortium (HGSVC) and Human Pangenome Reference Consortium (HPRC). We selected samples to include major African and non-African Y haplotypes, with the root of the phylogenetic tree dated to approximately 185 thousand years ago (kya) (95% HPD interval: 162-212 kya), thus covering genetic variation across a substantial period of modern human Y chromosome evolution.

We report remarkable complexity and diversity in chromosome size and structure, in contrast with its low level of base substitution variation. Specifically, the size of the Y chromosome assemblies varies extensively from 45.2 to 84.9 Mbp, mostly due to differences in the Yq12 heterochromatic block. Other regions, including the centromere and the ampliconic regions, revealed bursts of DNA expansions and contractions over short periods of time. Half of the male-specific euchromatic region is subject to large inversions with a >2-fold higher recurrence rate compared to inversions in the rest of the human genome. Ampliconic sequences associated with these inversions further show differing mutation rates that are sequence context-dependent and some ampliconic genes show evidence for concerted evolution with the acquisition and purging of lineage-specific pseudogenes. The largest heterochromatic region in the human genome, the Yq12, is composed of alternating arrays of *DYZ1* and *DYZ2* repeat units that show extensive variation in the number, size and distribution of these arrays, but retain a 1:1 copy number ratio of the monomer repeats, consistent with the notion that functional or evolutionary forces are acting to balance this chromosomal region.

The availability of sequence-resolved Y chromosomes from multiple individuals provides a unique opportunity for identifying new associations of specific traits with Y-chromosomal variants and garnering novel insights into the evolution and function of complex regions of the human genome.

Session Title: Evolutionary and Population Genetics Poster Session II

PB3095 Noncoding variant effect prediction with probabilistic machine learning models

Authors:

C. Shearer, A. Amin, D. Marks; Harvard Med. Sch., Boston, MA

Noncoding regions make up the majority of the genome and are causally related to human disease. Unfortunately we are often unable to distinguish noncoding variants that cause disease from "tag along" variants due to low statistical power. This presents a clear need for additional evidence to uncover pathogenic noncoding variants, both to extend clinical datasets and to identify specific variants that cause common disease. Given that substantial progress has been made in the development of computational methods for determining the effects of variants in proteins and design based on evolutionary selection, we hypothesize that similar techniques can be used for the noncoding genome. We present preliminary results for predicting pathogenic noncoding human variants unaligned genomic sequences from across evolutionary history.

Session Title: Evolutionary and Population Genetics Poster Session III

PB3096 Pacific Islander Founder Populations.

Authors:

M. Penjueli[1], J. Blanco-Portillo[2], J. Sokol[3,4], F. Liu[5], O. Garcia[3], K. Taparra[3], C. Barberena-Jonás[6], S. G. Medina-Muñoz[6], G. Venkataraman[3], A. Moreno-Estrada[6], P. Kirch[7], M. Feldman[2], C. Gignoux[8], G. Wojcik[9], K. Fox[10], A. R. Jha[1], A. Ioannidis[3,5]; [1]New York Univ. Abu Dhabi, Abu Dhabi, United Arab Emirates, [2]Stanford Univ., Palo Alto, CA, [3]Stanford Sch. of Med., Palo Alto, CA, [4]Boston Univ. Sch. of Med., Boston, MA, [5]Stanford Inst. for Computational and Mathematical Engineering, Palo Alto, CA, [6]CINVESTAV, Irapuato, Mexico, [7]Univ. of Hawaii at Manoa, Honolulu, HI, [8]Univ. of Colorado, Boulder, CO, [9]Johns Hopkins Bloomberg Sch. of Publ. Hlth., Baltimore, MD, [10]UCSD, La Jolla, CA

We describe genetic founder effects in Pacific Islanders and Native Hawaiians, an under-studied population, who currently experience the shortest life expectancy across all populations in the United States. We characterize the genetic bottlenecks and founding population sizes of islands in the primary settlement path to Hawaii, together with the proximate geographic origin and timing of settlement of Hawaii, which has long been a subject of debate. This allows us to characterize signals of selection between the islands over this time period. In particular, we use ancestry-specific approaches to explore the genomic variation of over four thousand modern-day Hawaiians and other Pacific Islanders using more than 864,892 single nucleotide polymorphism markers. Based on Polynesian-derived allele frequencies and segments, we find these master navigators arrived to Hawaii from a proximate geographic origin in the Tuamotu Archipelago, which is in concordance with oral traditions. This contrasts previous hypotheses, which posited an origin in the Marquesas or Society Islands. We also characterize the admixture patterns between several major ancestries found across these remote Polynesian islands, and together with our selection analysis, are able to highlight a number of loci, providing targets for further study to extend personalized genetic health results to this important founder population within the United States.

Session Title: Evolutionary and Population Genetics Poster Session I

PB3097 † Pattern of genetic variation and population structure in Native Hawaiians

Authors:

S. Jeon[1], B. Dinh[2], C. Fan[3], X. Sheng[4], F. Chen[3], L. Wilkens[5], L. L. Machand[5], C. Haiman[4], C. Chiang[3]; [1]Keck Sch. of Med., USC, Pasadena, CA, [2]Univ. of Southern California, West Hills, CA, [3]Univ. of Southern California, Los Angeles, CA, [4]Keck Sch. of Med., USC, Los Angeles, CA, [5]Univ. of Hawaii cancer center, Honolulu, HI

Populations with a special evolutionary history can lead to unique patterns of genetic variation and population structure, which, in turn, dictate the allelic architecture of complex traits and diseases. In particular, a population like the Native Hawaiians exemplifies all major features of human evolutionary history, including bottleneck, admixture, and natural selection. We investigated the impact of demographic history on the patterns of genetic variation and population structure using a total of 10,721 whole-genome sequenced individuals from five ethnic groups in the Multiethnic Cohort (MEC). In total, we identified 82 million SNPs and 5 million indels after quality control in this dataset, including 26 million SNPs and 2 million indels segregating in 1,065 self-reported Native Hawaiians. We found a large number of variants enriched in the Polynesian-ancestry population like Native Hawaiians, but rare elsewhere in the world. For instance, among 63 million variants found to be <0.5% in minor allele frequency in 9,656 MEC White, African American, Latino, and Japanese individuals, approximately 5 million of them have a frequency >5% in Native Hawaiians, suggesting that we would have increased power to detect associations at these variants only in the Polynesian-ancestry population. We leveraged the haplotypic information from the genome-wide sequence of genealogical trees to derive the expected genetic relationship matrix (eGRM). Based on the eGRM, we investigated the fine-scale structure in the Native Hawaiian population through principal components analysis (PCA) and uniform manifold approximation and projection (UMAP). We identified three major population subclusters within the data and found these subclusters to be differentiated among measures, such as runs of homozygosity, population size trajectories, and enrichment of functionally important alleles. Our results help provide insights into the fine-scale structure in the Native Hawaiian population and could aid in interpreting any differences in disease risk between Native Hawaiians and other continental populations.

Session Title: Evolutionary and Population Genetics Poster Session II

PB3098 Patterns of population structure & genetic variation within the Saudi Arabian population.

Authors:

D. Malomane[1], S. Mangul[2], M. Abedalthagafi[3], C. W. K. Chiang[1]; [1]Ctr. for Genetic Epidemiology, Dept. of Population and Publ. Hlth.Sci., Keck Sch. of Med., Univ. of Southern California, Los Angeles, CA, [2]Sch. of Pharmacy, Univ. of Southern California, Los Angeles, CA, [3]Emory Univ., Atlanta, GA

The Arabian Peninsula is considered the initial site of historic human migration out of Africa. The modern-day indigenous Arabians are believed to be the descendants which remained from the ancient split of the migrants into Eurasia. Following the split, the Arabians have experienced limited admixture not only externally but also within their country borders. In Saudi Arabia, tribalism has been practiced for centuries, encouraging endogamy while limiting inter-tribal marriages. Here, we investigated how such population history and cultural practices have shaped the genetic variation of the Saudi population. We genotyped 3,352 individuals and whole genome sequenced 341 from the Saudi population. Principal component analysis (PCA) and Uniform Manifold Approximation and Projection (UMAP) revealed the presence of genetic structure within the Saudi population. We detected twelve genetic subclusters that corresponded to the geographical distribution of different tribal regions. The distinction between genetic clusters was further supported by the components of ancestry based on unsupervised admixture analysis. We observed clear differences in the proportion of ancestral lineages in the Saudi clusters, with some clusters showing more homogeneity with a single dominant ancestry component with an estimated proportion up to 94% of the genome, while others are admixed with multiple ancestry components with maximum proportion no more than 27%. Furthermore, we used the f_3-statistics to test for evidence of admixture in these clusters, finding five out of twelve clusters with admixture sources proxied by various HGDP reference populations. We found differences in the runs of homozygosity (ROH) across clusters with the median total sum of ROH ranging from 38.12Mb to 232.6Mb within a cluster, while the median number of ROH ranged from 42 to 150 ROHs. Consistent with the out-of-Africa event, all clusters experienced and subsequently recovered from a decline in the effective population size (N_e) about a 100,000 years ago. A second decline in N_e occurred around 6-7,000 years ago, and while N_e rebounded for most clusters, the N_e for one cluster remained persistently low until the present time. This particular cluster is from the western part of Saudi Arabia, contains the most numerous and the longest ROH, and has the highest proportion of a single ancestral lineage, consistent with long-term isolation of these people. Together, these results provide insights into the patterns of genetic variation present within the Saudi Arabian population and establish a foundation on which to interpret medical- and pharmaco- genomic findings from this population.

Session Title: Evolutionary and Population Genetics Poster Session III

PB3099 Phasing of millions of samples achieves near perfect accuracy, enabling parent-of-origin classification of variants

Authors:

C. Williams[1], J. O'Connell[2], W. A. Freyman[2], 23andMe Research Team, C. R. Gignoux[3], S. Ramachandran[1], A. L. Williams[2]; [1]Brown Univ., Providence, RI, [2]23andMe, Inc., Sunnyvale, CA, [3]Univ. of Colorado Anschutz, Aurora, CO

Given the availability of large, ancestrally diverse genetic datasets and the potential for parent-of-origin association studies, we investigated state of the art intra-chromosomal phasing methods and developed a new inter-chromosomal phasing approach. Specifically, we applied SHAPEIT5 and Beagle 5 to chromosome 2 genotype data from both the UK Biobank (UKB) and >8 million diverse, research-consented 23andMe customers, finding that both perform exceptionally well. Beagle's median switch error rate (SER) (after excluding single SNP switches) in white British trios from UKB is 0.031% compared to zero for European ancestry 23andMe research participants (55.6% have zero phase errors). South Asian ancestry 23andMe research participants have the highest median SER amongst the 23andMe populations, but it is still remarkably low at 0.46%. SERs in non-European ancestry samples in UKB are up to ten times higher than their 23andMe counterparts, leading us to investigate best practices for phasing datasets with imbalanced ancestry compositions, e.g., the UKB where 94% of participants are described as having European ancestry. We find that SERs for African, South Asian, and East Asian ancestry samples from the UKB and 1000 Genomes Project are lowest when the white British are downsampled ~100-fold. In 23andMe we found mixed benefits of phasing with population-only samples vs. the 8 million research participants, motivating further inquiry. Building on these findings, we developed an inter-chromosomal phasing method that assigns paternal and maternal variants discretely genome-wide (whereas SHAPEIT and Beagle phase each chromosome separately). Our approach uses identity-by-descent (IBD) segments to phase variants on different chromosomes, representing the segments a focal individual shares with their relatives as nodes in a signed graph. We perform bipartite clustering on the signed graph using spectral clustering and tested our method on 1059 UKB trios, where it clustered IBD segments with a median precision of 99.7%, yielding a median phase accuracy of 99.8% in regions covered by IBD segments. We also ran the method in the 23andMe database and found a median global phase accuracy of 97.8% in Europeans and 98.2% in African Americans. The method's precision depends heavily on data from relatives, so will increase as datasets grow larger and more diverse. Our method enables analyses that require the parent-of-origin of variants, such as association studies and ancestry inference of untyped parents.

Session Title: Evolutionary and Population Genetics Poster Session I

PB3100 † Population identification based on shared ancestry captures fine-scale population structure and population-specific medical insights across the US

Authors:

M. Isshiki[1,2], I. Raghuram[3], A. Griffen[1], P. Meissner[1], J. Greally[1], S. M. Raj[1]; [1]Albert Einstein Coll. of Med., Bronx, NY, [2]The Univ. of Tokyo, Tokyo, Japan, [3]Univ. of California Berkeley, Berkeley, CA

More than 90% of genomic studies have been conducted on European ancestry individuals. Due to the lack of knowledge, the performance of genomic prediction of health is still not good in underrepresented racial or ethnic groups. We aim to expand the benefits of medical genomics to underserved populations in the US, using population genetics approaches to understand ancestry, and investigate their associations with different phenotypic outcomes. We focused our initial research on the participants of the All of Us Research Program (AoU) living in New York City (NYC) (n= 13,817), one of the most diverse cities in the world. We conducted Identity-by-descent (IBD) network analysis to ascertain fine-scale population structure within NYC. By combining it with the BioMe biobank (n=11,549), another dataset including NYC residents, we detected 16 distinct IBD population groups based on the amount of recent shared ancestry. Of these groups, eight showed signatures of strong founder effect: Puerto Ricans (PR), Ashkenazi Jewish (AJ), Central/South Americans (CSA), South Asians, Southeast Asians, South Caribbean African ancestry, Roma-Southeast European and Other Jewish. Across the eight founder populations, we identified 53 Pathogenic/Likely Pathogenic variants (curated in the ClinVar database) that had frequencies ≥ 0.005, the typical threshold of inclusion to genetic screening. Of these variants, 27 were previously unknown to have a founder effect. We next asked whether the IBD segments used to determine the population groups were associated with phenotypes. We used an IBD mapping approach to identify associations between IBD segments and phenotypes registered in electronic health records in AoU, in founder IBD groups with $n > 1000$ in our sample, namely PR, AJ and CSA. We also conducted Phenome-wide association studies (PheWAS) for each population to identify phenotypes enriched in each population group. We expanded this NYC-specific analysis to the entire AoU dataset across the US (n~245K). As expected, we found regional differences in population structure and the strength of founder effect. In particular, variation of IBD groups of Hispanic and Latino individuals were different among regions. We also examined differences in interaction with environment (e.g. pollution, sunlight) between IBD groups. We will make the IBD population groups available for other researchers working on the AoU dataset.

Session Title: Evolutionary and Population Genetics Poster Session II

PB3101 Population structure and genetic relatedness analysis of the UAE population.

Authors:

I. Morjane, On behalf of the UAE Healthy Future Study Team, Investigators and Collaborators; Publ. Hlth.Res. Ctr. , New York Univ. Abu Dhabi, Abu Dhabi, United Arab Emirates

Assessing genetic risk in multi-ethnic and admixed populations has yielded significant insights and uncovered various genetic variants associated with complex phenotypes and diseases. However, conducting robust association mapping studies in admixed populations requires comprehensive genetic profiling and careful consideration of factors such as relatedness and population structure. In this regard, populations of the Middle East provide both challenges and opportunities. Here we report the results of the analysis of genotyping data of study participants of the UAE Healthy Future Study (UAEHFS), the first prospective cohort study in the UAE, aiming to examine the association between multiple exposures (including environmental, lifestyle and genetic risk factors) and multiple outcomes with a focus on obesity, diabetes and CVD. We developed a custom version of the Precision Medicine Diversity Array (>800,000 genetic variants), generated genotyping data for the UAEHFS cohort and developed a database with annotation and allelic frequency data. The data was used to perform population genetic and relatedness analyses using different methods. The results highlight the impact of past and recent demographic events on the genetic make-up of the UAE population. We also test the performance of various measures of relatedness and test the impact of accounting for relatedness on mapping of quantitative traits.

Session Title: Evolutionary and Population Genetics Poster Session III

PB3102 Portability of tag SNPs used to determine ABO alleles across diverse populations: A systematic review.

Authors:

K. Martinez, A. Klein, J. R. Martin, C. Sampson, J. Giles, M. Beck, K. Bhakta, G. Quatraro, J. Farol, J. Karnes; Univ. of Arizona, Tucson, AZ

Introduction: The ABO gene codes for a major human alloantigen system. ABO blood types have widespread clinical use and are robustly associated with myriad phenotypes. Many studies genetically infer ABO blood type, but there is potential for misinference through use of inappropriate tag SNPs across diverse populations. This systematic review aimed to evaluate the portability and suitability of tag SNPs used to determine ABO alleles across diverse populations.

Methods: PubMed, Embase, Scopus, Medline/Ovid, CINAHL, and Dissertations and Theses were searched from 2005-2022 using keywords to identify tag SNPs to determine ABO alleles (O, A, and B). Primary data was extracted independently by two investigators and included: tag SNPs used for ABO alleles, presence of haplotype analysis, cohort size, population description, and primary phenotype. As almost all studies used race/ethnicity descriptors, we inferred continental ancestry for each study population based on 1000 Genomes superpopulations (AFR, AMR, EAS, EUR). We calculated linkage disequilibrium (r^2) between study-selected tag SNPs and ABO functional (O vs non-O and A vs B allele-determining) SNPs using inferred continental ancestry groups on a per study basis. Median r^2s for tag SNPs were compared between inferred continental ancestry groups using Kruskal Wallis followed by Dunn's test for multiple comparisons using rank sums.

Results: Inclusion criteria were met for 136 articles. Most studies were done in a EUR population (n=74) followed by EAS (n=31), AFR (n=19), AMR (n=5). Study tag SNPs had a wide range of linkage (r^2=0.387-0.975 for O vs non-O tag SNPs and r^2=0.187-1.00 for A vs B tag SNPs). 49.3% of the studies used the functional SNP to determine O vs non-O alleles and 66.2% of the studies a functional SNPs to determine A vs B alleles. The proportion of O allele tag SNPs with r^2<0.9 was 100% (AFR), 100% (AMR), 0% (EAS), and 89% (EUR). The proportion of O allele tag SNPs with r^2<0.5 was 100% (AFR), and 0% for AMR, EAS, and EUR. O vs non-O tag SNPs performed significantly worse in African populations (AFR median r^2=0.443) than in East Asian (EAS median r^2= 0.946; $p=1.1x10^{-5}$) and European populations (EUR median r^2=0.867; p=0.02).

Conclusion: Our results suggest a lack of portability of ABO tag SNPs across diverse populations. While continental ancestry is not equivalent to race and ethnicity, prior studies may have used inappropriate tag SNPs for their populations. This may have led to erroneous ABO allele determination and inaccurate associations, especially in individuals with African ancestry. This practice may exacerbate existing disparities in genomic research for underrepresented populations.

Session Title: Evolutionary and Population Genetics Poster Session I

PB3103 Predicting stature and estimating allele effect size in ancient humans.

Authors:

K. Kaymak-Loveless, S. Cox, I. Mathieson, C. Shin; Univ. of Pennsylvania, Philadelphia, PA

Ancient DNA has made it possible to use genomics to understand the evolutionary history of complex traits. Here, we assemble the largest dataset to-date of 444 ancient West Eurasian individuals with paired stature and whole-genome genetic data to investigate the applicability of polygenic scores derived from present-day populations, the importance of environment, and the effect of individual variants on height.

We first confirm previous results that polygenic scores can predict height in ancient populations. Using summary statistics from UK Biobank, we observe that 6.9% of variation in femur length is predicted by PRS (SD=0.29%). Importantly, this remains consistent when PRS are calculated using GWAS from sibling pairs (R2=0.059, SD=0.28%), indicating these results are not inflated by population stratification. We also show that effects are highly transferable by replicating this analysis using GWAS conducted in present-day populations of European and East Asian ancestry (R2=0.045 and 0.057 respectively).

Previous work by Marciniak et al. (2022) suggested that early Neolithic farming populations had lower stature than would be expected based on genetics, potentially due to relatively poor nutrition. However, concerns remained about the small sample size and population stratification. In our dataset, approximately three times larger, we robustly replicate these results including using sibling GWAS. Specifically, we estimate that Neolithic individuals were significantly shorter than non-Neolithic individuals with the same genetic height (p=0.0092), emphasizing the importance of the environment for understanding complex trait evolution.

Finally, due to our large sample size we performed association tests for stature to measure the effects of individual variants. We find that effect sizes of SNPs identified in GWAS are highly correlated with those in ancient samples (β=1.525, SE=0.326). However, some SNPs shown to have been under selection but that do not have an effect on stature in present-day populations, do have an effect in ancient populations. These include the SNP responsible for adult lactase persistence (β=0.599 SD, SE=0.200 SD), emphasizing the difficulty of using present-day SNP effects to draw conclusions about effects in the past.

Session Title: Evolutionary and Population Genetics Poster Session II

PB3104 Prehistoric ancestries are consistently associated with the complex trait landscape in European Biobanks.

Authors:

D. Marnetto[1], V. Pankratov[2], M. Mezzavilla[3], S. Aneli[4], D. Fusco[5], L. Pagani[3]; [1]Univ. of Turin, Torino, Italy, [2]Univ. of Tartu, Tartu, Estonia, [3]Univ. of Padua, Padova, Italy, [4]Univ. of Turin, Torino, Italy, [5]Università di Torino, Torino, Italy

The contemporary European genetic makeup formed in the last 8,000 years when local Western Hunter-Gatherers (WHGs) mixed with incoming Anatolian Neolithic farmers and Pontic Steppe pastoralists. Previous research efforts have indeed characterized evolutionary events specific to these populations which affected their phenotype, through the tracking of few highly characterized SNPs or polygenic scores. Nevertheless, while the first approach is limited in the number of variants analyzed and largely blind with regards to complex polygenic traits, the second builds on population-dependent effect sizes. Here we investigate how the unique tiling of genetic variants inherited from different ancestral components impacts the phenotypic landscape of 50,000 UK Biobank donors taken as representative of the Western European population. This work expands an analysis previously run on the Estonian Biobank which found significant trait-ancestry associations for 11 of the 27 complex traits investigated.
Using matching individual-level genotype and phenotype data from contemporary samples and genotype data from the ancient source populations, we quantify the correlation between each ancestry and present-day phenotypic variation in each complex trait. We rely on a metric (CovA) based on trait-associated genomic regions, without leveraging potentially population-specific effect sizes. The analysis is supported by a simulation framework to test the accuracy of our metric in a demographic scenario matching contemporary UK and probing several variables for their effect.
Among others, all anthropometric, hair pigmentation, heart rate trait-ancestry associations found in the Estonian Biobank were replicated in the UKBB population. Other complex traits reached significance possibly due to increased sample size: e.g we find a novel association between WHG ancestry and increased blood pressure. Furthermore we overcome the limitations of the previous analysis by a) increasing the number of traits analyzed and finding new associations e.g. between WHG ancestry and asthma; b) including a more rigorous simulation to help the signal interpretation; c) verifying that genome-wide ancestry associations are explained by either confounding factors or driven by highly polygenic local signals.
Our results show that these ancient components were differentiated enough, possibly through pre-admixture selection, to reveal ancestry-specific associations with the complex trait variability displayed by contemporary Europeans. Our findings are robust across biobanks and therefore representative of a continental trend rather than a population-specific result.

Session Title: Evolutionary and Population Genetics Poster Session III

PB3105 Probing the role of structural variants in adaptation in Holocene Western Eurasia.

Authors:

J. Maravall López[1], A. Akbari[2], D. Reich[1,2,3,4], H. Li[2,5]; [1]Harvard Univ., Cambridge, MA, [2]Harvard Med. Sch., Boston, MA, [3]Broad Inst. of Harvard and MIT, Cambridge, MA, [4]Howard Hughes Med. Inst., Cambridge, MA, [5]Dana-Farber Cancer Inst., Boston, MA

Complex genetic variation beyond single nucleotide polymorphisms is notoriously difficult to resolve, which hinders understanding of their phenotypic effects and evolutionary dynamics. Here, we leveraged new bioinformatic methodology and the exponentially increasing number of available ancient genomes to study the frequency changes of structural variants over time in about 6000 Holocene West Eurasians. We leveraged k-mer based genotyping utilizing a pangenome graph built using data from the Human Pangenome Reference Consortium cohort to study structural variants in high association with SNPs that have been established as significant hits in genome-wide scans for selection in ancient DNA time-series data. We identified a set of structural variants, largely unreported, for which co-localization of these selection scans and orthogonal GWAS results align, consistent with the haplotypes carrying the structural variants having important phenotypic effects which were acted upon by selection.

Session Title: Evolutionary and Population Genetics Poster Session I

PB3106 Reconstructing parent genotypes at direct testing-level accuracy using siblings and other relatives

Authors:

A. Williams[1], Y. Qiao[2], E. Jewett[3], K. McManus[4], W. A. Freyman[1], J. Blangero[5], The 23andMe Research Team; [1]23andMe, Inc., Sunnyvale, CA, [2]Cornell Univ., Ithaca, NY, [3]23andMe, Sunnyvale, CA, [4]23andMe, Inc., San Francisco, CA, [5]Univ. of Texas Rio Grande Valley Sch. of Med., Brownsville, TX

Reconstructing ancestors' DNA has tremendous potential to empower genome-wide association studies by increasing sample size, can improve direct genetic effect estimates, and can enhance forensic and genetic genealogical research. In expectation, n full siblings inherit a fraction of $1\text{-}1/2^n$ of each parents' genome or 93.8% when $n = 4$, thus providing rich information for reconstruction.

We developed HAPI2, a method for jointly phasing $n \geq 2$ siblings that implicitly reconstructs their parents' DNA and also built HAPI-RECAP, a tool that combines HAPI2's output with identity-by-descent (IBD) segments from relatives to distinguish which haplotypes belong to each parent. More specifically, sibling data does not immediately provide inter-chromosomal parent assignments of the haplotypes, and intra-chromosomal parent assignments are sometimes ambiguous, but IBD segments to maternal/paternal relatives reveal cross- and within-chromosome linkage. Moreover, recombination breakpoints in the inferred transmitted haplotypes carry information about the sex of the transmitting parent since locus-specific recombination rates differ markedly between males and females.

We analyzed 72 nuclear families from the San Antonio Mexican American Family Studies (SAMAFS), each with data from both parents and ≥ 4 full siblings, and censored the parents' data before running HAPI2 to reconstruct parent genotypes. When analyzing 10 families with data for ≥ 8 full siblings, HAPI2 reconstructs an average of 98.9% of each parent's genotypes with a mean error rate of $< 10^{-3}$, or roughly equal to genotyping error rates. In turn, when applied to all 72 sibling sets, HAPI2 reconstructs an average of 92.5% of each parent's genotypes with a mean error rate $< 10^{-3}$.

To link the genotypes to each parent inter- and intra-chromosomally, we applied HAPI-RECAP and effectively linked the majority of the reconstructed DNA to one of the parents, again with a mean error rate of $< 10^{-3}$. We have previously found that use of sex-specific genetic maps can correctly infer the parent sexes using IBD break points from half-sibling data and will extend this methodology to infer the parent sexes using the recombinations inferred by HAPI2. Both IBD sharing with relatives and sex-specific maps together have the power to enable precise sex-specific assignments of reconstructed genotypes to parents.

Besides these results from the SAMAFS data, we will apply HAPI-RECAP to a collection of ancestrally diverse families from the 23andMe research participant database and present results quantifying the impact of including various numbers of close and distant relatives.

Session Title: Evolutionary and Population Genetics Poster Session II

PB3107 Reconstructing the genetic history of Ancestral South Indians and Ancestral Austro-Asiatics

Authors:

A. Dey[1,2], A. Basu[1], D. Nandi[1]; [1]Natl. Inst. of BioMed. Genomics, Kalyani, India, [2]Univ. of Calcutta, Kolkata, India

South Asia has a complex population history. Using previously published genomic data of present-day and ancient individuals from South Asia and its neighboring regions, we identified an ancestral component that is ubiquitous in the genomes of many present-day South Asian tribal populations, henceforth Tribal South Asian (TSA) ancestry. TSA was also present in South Asian ancient individuals, but almost absent in the ancient individuals from the neighboring regions. Most importantly, it is the only ancestry that is found in all present-day individuals from mainland South Asia. Hence, it is likely to represent one of the oldest ancestries of this region. TSA is highly correlated with Ancestral South Indian (ASI) and Ancestral Austro-Asiatic (AAA) ancestries (r=0.98 and 0.88 respectively), which can only be inferred exclusively from present-day South Asian individuals, and hence they provide better resolution. Due to their centrality in the peopling of South Asia, we aim to understand the formation of AAA and ASI using the genomes of present-day Austro-Asiatic and Dravidian-speaking tribal populations, which have AAA and ASI as their major ancestries respectively. These tribal populations are mostly isolated and they have a scattered distribution in Central and South India. We have analyzed GenomeAsia 100K-generated whole genome sequencing data of 497 unrelated South Asian individuals, that included 141 unrelated individuals from 11 Dravidian tribal populations and 4 Austro-Asiatic tribal populations. Despite being mostly isolated, these tribal populations showed high genetic similarity relative to other South Asian populations in Principal Component Analysis and ADMIXTURE analysis. We constructed a phylogenetic tree and performed D-statistics and IBD-based analyses which suggested that the Dravidian tribes from Central India are genetically closer to the Austro-Asiatic tribes, than they are to the Dravidian tribes from South India. We have used site frequency spectrum-based analysis to distinguish among various plausible demographic models in order to understand the population history of Austro-Asiatic and Dravidian tribes and estimate the separation time between ASI and AAA. This study elucidates the interplay between ASI and AAA ancestries and their role in the peopling of South Asia.

Session Title: Evolutionary and Population Genetics Poster Session III

PB3108 Reconstructing Western and Northern Mexico's Past through the PIPANOM Project.

Authors:

R. Curry[1], J. Sedig[2], J. Punzo Diaz[3], D. Reich[4]; [1]Harvard, Cambridge, MA, [2]Harvard Med. Sch., Cambridge, MA, [3]Inst. Natl. de Antropología e Historia, Mexico City, Mexico, [4]Harvard Univ., Cambridge, MA

Though ancient DNA has revolutionized our understanding of the past, America/Mexico remain understudied. The Proyecto de investigación de poblaciones antiguas en el norte y occidente de México (PIPANOM) addressed this imbalance through the examination of over 300 ancient individuals from central, western and northern Mexico. Data from these individuals have revealed population structure that is broadly associated with geographic distance across what is now modern Mexico, Belize, Guatemala, and the southwest United States. These data have also provided new insights into long-standing questions about migration and interaction of different archaeologically defined cultures in key eras of Mexico's past, such as the movement of people during the Postclassic period. Along providing new insights in an area that has been understudied with aDNA, PIPANOM has helped correct institutional imbalances in archaeological and paleogenetic research. PIPANOM is an archaeologist-led project that has brought together more than 20 collaborators. As part of the project, there have been numerous workshops, trainings, and visits bringing together Mexican and American archaeologists, geneticists, researchers, analysts, and students. As such, PIPANOM provides a roadmap for equitable large-scale projects that produce robust scientific results.

Session Title: Evolutionary and Population Genetics Poster Session I

PB3109 Roles of Genetic Ancestry and Hypopigmenting Alleles on Skin Color in a Caribbean Native American Population

Authors:

K. Ang[1], V. A. Canfield[1], T. C. Foster[1], T. Harbaugh[2], K. Early[1], R. Harter[3], K. Reid[1], S. Leong[1], Y. Imamura[4], D. Liu[1], J. Hawley[5], K. Cheng[6]; [1]Penn State Univ., Hershey, PA, [2]Penn State Hlth., Milton S. Hershey Med. Ctr., Hershey, PA, [3]Penn State Univ., State College, PA, [4]Penn State Coll. of Med., Hershey, PA, [5]Salybia Mission Project, Kalinago Territory, Dominica, [6]Penn State Col Med, Hershey, PA

Shared genetic ancestry suggests the sharing of one or more hypopigmenting skin color alleles between Native Americans and East Asians. Quantitating the effect sizes of candidate alleles requires an admixed study population with African but minimal European admixture. We genotyped 458 individuals (~15%) from a Caribbean Native American population, the Kalinago from the Commonwealth of Dominica, using Illumina Infinium Omni2.5-8 BeadChip. Admixture-based genetic ancestry was confirmed to be shared with East Asians at K=3. Genetic ancestry at K=6 was 55% Native American 32% African, and 12% European. Skin pigmentation ranged from 20 to 80 melanin units, averaging 46. Three albino individuals were homozygous for a multi-nucleotide polymorphism $OCA2^{NW273KV}$ of African origin whose single allele effect size was -8 melanin units. European hypopigmenting allele frequencies for $SLC24A5^{A111T}$ and $SLC45A2^{L374F}$ were 0.14 and 0.06, with effect sizes per allele of -6 and -4, respectively. Native American genetic ancestry by itself reduced pigmentation by more than 20 melanin units (range 24 - 29). Taken together, these results suggest the presence of to-be-identified hypopigmenting alleles of high impact that may be uniquely Native American or shared with East Asians

Session Title: Evolutionary and Population Genetics Poster Session II

PB3110 Selective dynamics of interruptions at short tandem repeats.

Authors:

M. Goldberg[1], H. Dashnow[1], K. Harris[2,3], A. Quinlan[4]; [1]Univ. of Utah, Salt Lake City, UT, [2]Univ. of Washington, Seattle, WA, [3]Fred Hutchinson Cancer Res. Ctr., Seattle, WA, [4]Univ of Utah, Salt Lake City, UT

Short tandem repeats (STRs) are hotspots of genomic instability that mutate at rates orders of magnitude greater than non-repetitive genomic loci. In the germline, STRs typically mutate through replication slippage, which manifests as expansions or contractions of a locus by one or more repeat units. A subset of STRs are likely functional: somatic and germline expansions at some loci are linked to certain Mendelian diseases, while variation at other loci has recently demonstrated association with complex traits. Accordingly, some STRs are inferred to be under purifying selection, regardless of their instability. A major determinant of an STR's mutation rate is the homology of its repeat units; one or more 'interruptions', or bases that disrupt the locus's canonical repeat, have been shown to significantly decrease mutation rate in both the germline and the soma. More broadly, interruptions may act as *cis*-acting anti-mutator alleles, in perfect linkage disequilibrium with the repeat allele whose mutation rate they decrease. We therefore hypothesized that interruptions would occur more frequently at loci under purifying selection, where the fitness effects of expansions would be more deleterious than at neutral loci. To test this hypothesis, we examined the distribution of high quality, segregating single nucleotide variants that occur within STRs. Intriguingly, we find the strength of purifying selection on loci significantly covary with the number of interrupting polymorphisms, regardless of SNV allele frequency and STR locus characteristics. Our findings may indicate that interruptions' abundance may be partially explained by resulting lower mutational burden at their linked loci.

Session Title: Evolutionary and Population Genetics Poster Session III

PB3111 Sexual dimorphism in Early Neolithic Europe was driven by culture, not genetics or environment.

Authors:

S. Cox[1], N. Nicklisch[2], M. Francken[3], J. Wahl[4], H. Meller[5], W. Haak[6], K. W. Alt[2], E. Rosenstock[7], I. Mathieson[1]; [1]Univ. of Pennsylvania, Philadelphia, PA, [2]Danube Private Univ., Krems-Stein, Germany, [3]State Office for Cultural Heritage Management Baden-Württemberg, Konstanz, Germany, [4]{Inst. fur Naturwissenschaftliche Archa}ologie, Tübingen, Germany, [5]State Office for Heritage Management and Archaeology Saxony-Anhalt, Halle, Germany, [6]Max Planck Inst. for the Sci. of Human History, Jena, Germany, [7]Univ. Bonn, Bonn, Germany

Human culture impacts health and disease as much as genetics or the natural environment, but their effects are often confounded. The combination of ancient DNA and polygenic scores derived from present-day populations provides an opportunity to estimate genetic effects directly and thus isolate the effects of culture and environment on health. Here, we take a multidisciplinary approach by combining published ancient DNA with osteological data on height, health, and diet to analyze cultural, environmental, and genetic contributions to variation in stature in four geographically defined populations of Early Neolithic Europe (8000-6000 years before present): North Central (n=203), South Central (n=1067), Southern (Mediterranean; n=127), and Southeastern (Balkan; n=139) Europeans. In individuals from North Central Europe, female stature is low ($p=9x10^{-7}$; β=-2.0cm), despite polygenic scores for height identical to males and to neighboring regions (p &gt 0.9). Dietary stable isotopes and paleopathological stress markers indicate increased environmental stress in this region that is equal in both sexes, even though it is only the females who appear stunted. This, combined with the high stature sexual dimorphism ratio (M:F=1.14) suggests strong effects from a cultural factor which preferentially supported the males to recover from environmental stress. In contrast, shorter average stature in Mediterranean Neolithic populations have been previously observed and reported to be associated with genetic differences; however, our analysis indicates this is likely an artifact of residual population structure in the genome-wide association studies (GWAS) as decreased polygenic scores do not replicate when calculated using summary statistics from between-sibling GWAS. Instead, we observe decreased height only in males ($p=5.5x10^{-7}$; β=-1.44cm), leading to reduced sexual dimorphism in the region (M:F=1.05), and which indicates a degree of male vulnerability in response to general environmental stress.
We conclude that while population-level stature trends in some cases reflect genetic factors, differences in sexual dimorphism in Early Neolithic Europe were largely driven by culture, or the interaction of culture and environment. Understanding these patterns is key to interpreting the evolution of genetic and socio-cultural determinants of health and the origins of social inequality.

Session Title: Evolutionary and Population Genetics Poster Session I

PB3112 Shared and distinct natural selection signatures within the Japanese population.

Authors:

X. Liu[1], M. Matsunami[2], M. Horikoshi[3], S. Ito[4], Y. Ishikawa[5], K. Suzuki[1], Y. Momozawa[1], S. Niida[6], R. Kimura[7], K. Ozaki[8], S. Maeda[7], M. Imamura[9], C. Terao[10]; [1]RIKEN, Yokohama, Japan, [2]Univ. of the Ryukyus, Nishihara-cho, Japan, [3]RIKEN Ctr. for Integrative Med. Sci., Yokohama, Japan, [4]Shimane Univ., Matsue, Japan, [5]Riken, Ctr. for Integrative Med. Sci., Yokohama, Japan, [6]NCGG, Obu, Japan, [7]Univ. of the Ryukyus, Nishihara, Japan, [8]The Natl. Ctr. for Geriatrics and Gerontology, Obu, Aichi, Japan, [9]Univ. of the Ryukyus, Nishihara, Okinawa, Japan, [10]RIKEN IMS, Yokohama, Japan

Differential natural selection signatures among the Japanese subpopulations are largely unexplored. Here we conducted genome-wide selection scans with 622,926 single nucleotide polymorphisms for 20,366 Japanese individuals, who were recruited from the main-islands of Japanese Archipelago (Hondo) and the Ryukyu Archipelago (Ryukyu), the two representative subpopulations. The integrated haplotype score analysis identified shared and distinct signals under positive selection. We found a novel locus at *IKZF2* especially in Ryukyu. Although the MHC showed significant signals in both subpopulations, the variants with the strongest signals exhibited the largest allele frequency differences between the subpopulations and tagged distinct haplotypes of HLA alleles, namely, *DQB1*06:02-DRB1*15:01* and *DQB1*06:04-DRB1*13:02* in Ryukyu and Hondo, respectively. The haplotype *DQB1*06:02-DRB1*15:01* is reported to be protective for human T lymphotropic virus type 1 (HTLV-1)-associated myelopathy/tropical spastic paraparesis (HAM/TSP), possibly mirroring the high prevalence of HTLV-1 infection in Ryukyu. In supporting to this link, *IKZF2* is frequently deleted in HTLV-1-associated adult T cell leukemia/lymphoma (ATL). The fastSMC analysis identified 14 loci under positive selection within the past 20-150 generations, and uncovered region-specific and time-dependent selection profiles, especially in *ALDH2*, a well-known East-Asian-specific signal. In summary, our study provided insights into the selection signatures within the Japanese and revealed potential sources of selection pressure.

Session Title: Evolutionary and Population Genetics Poster Session II

PB3113 Single-locus imputation of ancient African DNA using novel regression-based approach

Authors:

R. Pornmongkolsuk, A. Ochoa; Duke Univ., Durham, NC

In the study of adaptations, ancient DNA has the unique potential of providing ancestral haplotypes that are not present among modern populations. Indeed, ancient samples are highly susceptible to degradation that often results in low-coverage sequenced data-a characteristic that calls for genotype imputation to improve the accuracy of downstream analyses. However, most of the common imputation approaches require high-quality reference haplotypes-generally from modern genomes, which excludes a majority of informative ancient genomes from the reference panel. Here, we present a new logistic regression-based model for imputing ancient DNA on a single locus. The model is trained on a set of reference genomes tailored for a target sample based on its missing loci pattern. As a result, the model allows the reference genomes to retain an arbitrarily high missing rate-as long as the target locus is not missing. To account for linkage disequilibrium (LD), we use a forward selection algorithm to identify independent SNP predictors. In this project, the method will be used to impute the Duffy null polymorphism on the *DARC* gene (chr1:rs2814778)-a well-known adaptive variant protective against one type of malaria-to study its evolutionary history in ancient African populations, which are notoriously challenging for existing imputation approaches. This model will be validated 1) on simulated data that resembles African demographic models and ancient DNA missing rate patterns and 2) by the leave-one-out cross-validation test with real data of ancient African genomes. Overall, this approach is promising for imputing individual variants of interest with high accuracy and improving our understanding of single-locus adaptation by leveraging ancient DNA.

Session Title: Evolutionary and Population Genetics Poster Session III

PB3114 Sperm sequencing reveals extensive positive selection in the human germline.

Authors:

M. Neville, A. Cagan, A. Lawson, R. Sanghvi, T. Bayzetinova, F. Abascal, P. Nicola, I. Martincorena, M. Hurles, R. Rahbari; Wellcome Sanger Inst., Cambridge, United Kingdom

Mutations that occur in the germline are the source of all heritable genetic variation. Understanding the mechanisms creating germline mutations and the selection pressures shaping them are therefore crucial to the study of evolution and disease. Although the majority (~80%) of inherited mutations in humans originate from the paternal germline, direct observation of mutations in sperm has been limited by the need for an extremely low error rate sequencing technology. In this study, we utilized NanoSeq, a duplex sequencing method with an error rate below five per billion base pairs, to sequence 104 bulk sperm samples from individuals aged 24 to 75 years. Our findings revealed a steady accumulation of 1.4 mutations per year, consistent with classical trio studies. By conducting deep targeted sequencing on the bulk sperm samples, we identified over 10,000 coding mutations, equivalent to sequencing the coding regions in hundreds of parent-child trios for each sample. Leveraging these variants, we detected 7 known and 14 novel genes subject to significant positive selection in the male germline. Interestingly, our study shows that germline positive selection, which occurs via a proliferative advantage of mutant spermatogonia in the testis, extends beyond the RAS/MAPK pathway and can also involve loss-of-function mechanisms. Notably, nearly all positively selected genes identified in our study are associated with pathogenic disorders in children when inherited, likely leading to an increased birth prevalence of these disorders, especially to older fathers. Furthermore, we quantified the fraction of sperm carrying pathogenic variants and demonstrated that although no single pathogenic variant is typically found in a high fraction of an individual's sperm (all < 0.5%), estimates of the total pathogenic burden exhibited a strong correlation with age, reaching over 5% of sperm in certain older individuals. These findings shed light on the dynamics of germline mutations and have important implications for our understanding of human disease.

Session Title: Evolutionary and Population Genetics Poster Session I

PB3115 Strategies for improving variant calling on the sex chromosomes

Authors:

M. Wilson[1], S. Plaisier[2], T. Phung[1], A. Oill[3]; [1]Arizona State Univ., Tempe, AZ, [2]Arizona State Univ., Chandler, AZ, [3]Translational Genomics Res. Inst., Phoenix, AZ

Mammalian sex chromosomes have a shared evolutionary history, as they were once a pair of homologous autosomes. Though X and Y are differentiated, they still share high levels of sequence similarity in some regions, like the pseudoautosomal regions (PARs) and the X-transposed region (XTR). Females typically have two X chromosomes and males typically have an X and a Y chromosome. Despite the importance of X and Y in human health and disease and their unique evolutionary history, there is limited understanding on how taking a standard approach for alignment and variant calling - one that is designed for the autosomes/diploid chromosomes - affects variant calling on the sex chromosomes. Here, we undertook a simulation study to assess the effects of standard autosomal versus sex chromosome complement-informed alignment, variant calling and variant filtering strategies on variants called on the human sex chromosomes. We find that aligning samples to a reference genome informed by the sex chromosome complement of the sample - either the entire Y hard masked when aligning genetic females (46, XX) or hard masking one copy of the PARs when aligning genetic males (46, XY) - increases the number of true positives called in the PARs, and additionally, in females only, the XTR. Unexpectedly, masking one copy of the XTR when aligning male samples results in a ten-fold higher rate of false positives in the XTR. We find that haploid calling on X and Y in male samples reduces the number of false positives compared to diploid calling but does not decrease the number of false negatives. We also find that using diploid-based filter thresholds on haploid chromosomes worsens variant calling for some filters. Finally, we find that - using a default alignment approach - joint genotyping the X chromosome between males and females slightly reduces the number of true positives called in females, while dramatically increasing true positives called in males, but this effect is mitigated when using a sex chromosome-complement informed alignment. By taking into account sequence similarities and ploidy differences on the sex chromosomes, we outline best practices for mapping, calling and filtering variants on the sex chromosomes. We recommend aligning samples to versions of the human reference genome informed on the sex chromosome complement of the sample and to use biologically accurate ploidy parameters when calling variants and setting filtering thresholds. This approach results in a more accurate way of identifying variants on the sex chromosome and will help improve to account for the unique biology of chromosomes X and Y.

Session Title: Evolutionary and Population Genetics Poster Session II

PB3116 Strengths and limitations of using the CARTaGENE population-based cohort to assess carrier rates for recessive disorders in Québec

Authors:

A-M. Laberge[1], K. Mohler[1,2], R. Laflamme[3], G. Lettre[4]; [1]CHU Sainte-Justine, Service de génétique médicale, Montreal, QC, Canada, [2]CHUM, Montreal, QC, Canada, [3]Montreal Heart Inst., Montreal, QC, Canada, [4]Université de Montréal, Montreal, QC, Canada

Background: The French-Canadian (FC) population is known to have an increased frequency of specific monogenic conditions, due to a founder effect. Reported carrier frequencies of common Mendelian conditions in the Québec FC population are based on disease incidence using Hardy-Weinberg equilibrium. Accurate carrier rates for specific autosomal recessive conditions can inform population-based screening programs and individual risk assessment. CARTaGENE (CaG) is a large population-based cohort of 43 000 adults aged 40-69 recruited randomly to represent the Québec population, including individuals of FC, Haitian and Moroccan descent. **Methods:** We aim to estimate carrier frequencies of specific mutations associated with autosomal recessive conditions in the FC, Haitian and Moroccan populations of Québec. Based on a literature review, we identified recurrent variants specific to each population (102 FC, 5 Haitian, 33 Moroccan). Genotyping data on FC variants is available for 8 740 FC participants and genome sequencing data on 2 173 participants (1 756 FC, 163 Haitian and 131 Moroccan). Variant frequencies were calculated using the open-source genome analysis toolset PLINK and were compared to available reported carrier rates. **Results:** Observed variant frequencies confirmed reported carrier rates for some variants, however others were discordant with reported carrier rates. For example, carrier rates for four recessive conditions for which carrier screening is recommended in individuals from the Saguenay Lac-St-Jean (SLSJ) region were significantly lower than the carrier rates found in individuals eligible for carrier testing based on SLSJ origin. Potential explanations include a mismatch between representation of subregions of SLSJ in CaG and distribution of carriers within that region or that CaG may have recruited in SLSJ individuals who are not from SLSJ. Some variants were not found at all in the CaG cohort (26 FC, 3 Haitian, 13 Moroccan), suggesting that reported carrier rates may be overestimated, that some so-called founder mutations are not recurrent in the population or that the CaG cohort does not have the power to detect regional differences in variant frequencies. **Conclusion:** Our analyses show the strengths and limitations of using a population-based cohort to assess carrier rates when carriers are not uniformly geographically distributed in the population, for example if located in specific subregions. We plan for more detailed analyses to assess provincial and regional allele frequencies, including an assessment of regional carrier rates of current population instead of carrier rates based on region of origin of individuals.

Session Title: Evolutionary and Population Genetics Poster Session III

PB3117 Structural variant adaptation between the Peruvian Andes and Amazon using long-read sequencing

Authors:

V. Borda Pua[1], S. M. Yan[2], M. Gouveia[3], S. Roy[1], K. Cifuentes Godinez[1], R. Juscamaita[4], C. Sanchez[5], O. Caceres[5], C. Padilla[5], H. Guio[6], E. Tarazona-Santos[7], C. Rotimi[3], R. McCoy[2], M. Santolalla[4], T. O'Connor[1]; [1]Univ. of Maryland Sch. of Med., Baltimore, MD, [2]Johns Hopkins Univ., Baltimore, MD, [3]NIH, Bethesda, MD, [4]Univ. Peruana Cayetano Heredia, Lima, Peru, [5]Inst. Natl. de Salud, Lima, Peru, [6]INBIOMEDIC, Lima 14, Peru, [7]Univ.e Federal de Minas Gerais, Belo Horizonte, Brazil

The harsh environments, including the high-altitude Andean mountains and the Amazon rainforest, posed significant challenges to the early settlers of South America. Physiological adaptations to these environmental challenges have been well characterized. However, our understanding of the genetic basis of these adaptive processes, explored mainly through single nucleotide variants (SNVs), remains limited. It has been hypothesized that structural variants (SVs, ≥ 30 bp) could be pivotal in genome evolution and adaptation. We investigate the impact of SVs on the adaptation of Indigenous Americans to Andean and Amazon environments. For this purpose, we generated high-coverage long-read sequence data on two Peruvian individuals with predominant Indigenous American ancestry (>99%) using PacBio HiFi sequencing. By integrating our data with 34 publicly available PacBio HiFi from Human Genome Structural Variation Consortium, we created an SV panel with 200,341 SVs. Subsequently, we used a graph-genotyping method and our SV panel to genotype SVs in short-read data from the Peruvian Genome Project (n=150), HGDP, and 1KGP, including two Andean and three Amazonian populations.
Our population structure analyses revealed the clear continental separation with a contrasting structure between Andean and Amazonian groups, with higher differentiation observed among Amazonian groups. We used the population branch statistic (PBS) in the form ((Andes; Amazon); East Asia) to identify highly differentiated alleles as candidates for adaptation to Andes and Amazon rainforest. We identified two candidates with strong allele frequency (AF) differences between populations of these two geographical regions. First, a 30 bp deletion (derived allele) in intron 23 of the *JAK2*, a gene that plays a crucial role in tolerance to hypoxia. This locus is highly differentiated in Andean populations (PBS value above 99.5th percentile), with the ancestral AF exceeding 0.7 while the Amazonian ancestral AF was below 0.3. Second, a 3305bp insertion in an intron region of the *CASP8* gene was nearly fixed in Amazonian populations (PBS value above 98th percentile; Amazonian AF > 0.91; Andean AF <0.55). *CASP8* plays an essential role in innate inflammatory cytokine production during bacterial infection. Our approach to exploring structural variation provides new directions to discover genetic variants under natural selection to harsh environments during the peopling of the Americas.

Session Title: Evolutionary and Population Genetics Poster Session I

PB3118 Structure and function of mammalian chitinases: Insights from chimeric analysis.

Authors:

K. Takahashi, E. Tabata, F. Oyama; Kogakuin Univ, Tokyo, Japan

Chitin, a predominant polysaccharide in nature, serves as a vital structural component in various organisms, including crustaceans, insects, nematodes, and fungi. Despite lacking chitin and its synthase, mammals produce two chitinases: chitotriosidase (Chit1) and acidic chitinase (Chia). Chit1, the first purified mammalian chitinase, shows elevated levels in the plasma of individuals with Gaucher disease, an autosomal recessive lysosomal storage disorder. Chia, highly expressed in mouse stomach tissues, exhibits optimal activity at pH 2.0. These mammalian chitinases are thought to act as defense mechanisms against chitin-containing pathogens and as digestive enzymes breaking down chitin in ingested material. However, the specific structural characteristics underlying their distinct properties remain elusive. In this study, we generated chimeric constructs of Chit1 and Chia to explore the structural elements contributing to the unique features of these mammalian chitinases. Our analysis of these chimeras identified multiple regions involved in the enhanced enzymatic activity of Chia under strong acidic conditions. To further refine our understanding, we are currently preparing a new chimera and conducting additional investigations.

Session Title: Evolutionary and Population Genetics Poster Session II

PB3119 The Emirati Genome Program: 1 million genomes for a better healthcare in the United Arab Emirates.

Authors:

T. Magalhaes, F. Al Marzooqi, The Emirati Genome Program; G42 Hlth.care, Abu Dhabi, United Arab Emirates

The Emirati Genome Program (EGP) aims to whole-genome sequence at ~30X ~1 million Emiratis in the United Arab Emirates (UAE) by 2025. EGP will be amongst the largest national genome sequencing program worldwide and likely the largest in terms of proportion of the population sequenced. The primary objective of EGP is to translate unprecedented genomics datasets, integrated with phenotypic data, into clinical practice.
As of June 2023, we have collected DNA samples from 440,000 EGP voluntary participants covering a representative geographic, age and gender range. We have sequenced 422,000 participants at >90 gigabases (Gb) across three high-throughput sequencing (HTS) platforms. A unique feature of EGP is the combination of both short-read, Illumina and MGI, as well as long-read Oxford Nanopore Technologies (ONT) sequencing. Through the latter we are conducting an unprecedented characterization of structural variation (SV) and population in a human population.
We have matched these multi-platform HTS capacity with computing, storage and analytics infrastructure optimized for each sequencing technology. For the raw-to-VCF processing of Illumina data, we use the DRAGEN system fully on premise. For MGI, we combine both on-premise and cloud processing leveraging MGI-proprietary, open-source and commercial software. We find ONT analysis workflow the most complex due to the balancing of GPU (base-calling) and CPU (alignment and variant calling) resources, the high computational cost of base-calling (especially when including methylation) as well as the sequencing and computational changes in a fast-evolving technology such as ONT. In all three analysis workflows, we are trying to call not only small variants (SNP and INDEL) but also additional genomic features (SV, CNV, HLA, pharmacogenomics, regions of homozygosity) and epigenetics. To minimize the storage footprint, we have adopted CRAM and decided to only keep alignments, variant calls and summary statistics alternating archival and fast access storage. To efficiently mine millions of genetic variants across tens of thousands of genomes, we are implementing an infrastructure based on Glow coupled with commonly used bioinformatics software for downstream analyses, e.g. genome-wide association analyses.
We have started translating the resulting EGP datasets into pilot clinical applications - from the screening of the EGP cohort for actionable genes recommended by the ACMG and rare disease-causing variants to the characterization of large likely deleterious SV and the HLA - as well as the implementation of nation-wide programs (pre-marital, pre-implantation and newborn screening).

Session Title: Evolutionary and Population Genetics Poster Session III

PB3120 The evolutionary fate of Neanderthal introgression in recently admixed African-American genomes.

Authors:

A. Pfennig, J. Lachance; Georgia Inst. of Technology, Atlanta, GA

A recurring theme in studies of ancient introgression in Eurasian populations is that Neanderthal DNA was generally deleterious to modern humans, leaving 1-2% of Neanderthal introgressed DNA in present-day non-Africans. Although simulation-based studies suggest that negative selection alone can explain observed introgression patterns in Eurasians, the exact evolutionary mechanisms of this purging have not been definitively established. For instance, another initially hypothesized mechanism involves hybrid incompatibilities, i.e., synthetic lethality or sterility. Furthermore, the purging of introgressed DNA likely occurred quickly as Neanderthal ancestry remained constant in Europe during the last 45,000 years, suggesting Neanderthal DNA is evolutionary neutral nowadays. Genomes of recently admixed individuals with combinations of African-like and European-like ancestry allow the testing of evolutionary mechanisms behind the purging of Neanderthal DNA and its evolutionary fate in extant genomes. Because African genomes have very little Neanderthal ancestry, most Neanderthal-derived variants in admixed genomes were contributed by European populations. Thus, these Neanderthal alleles have recently been tested against African DNA as they were introduced into novel genetic backgrounds. Here, we leverage whole-genome sequences of ~50,000 recently admixed African-American individuals from the 1000 Genomes Project and *All of Us* databases to test for secondary selection of Neanderthal DNA. We identify putatively selected Neanderthal introgressed segments and novel introgression deserts in admixed African-American individuals relative to expectations from the source populations. Preliminary results suggest that there is a slight enrichment of Neanderthal ancestry in African Americans. Furthermore, selected introgressed segments coincide with GWAS signals for anthropometric traits like height and BMI.

Session Title: Evolutionary and Population Genetics Poster Session I

PB3121 The evolutionary history of 17q21.31 structural haplotypes in ancient and modern humans

Authors:

S. Sridharan[1], J. Rocha[1], P. Sudmant[2]; [1]Univ. of California, Berkeley, Berkeley, CA, [2]Univ. of California Berkeley, Berkeley, CA

Structural variation accounts for a large proportion of genetic diversity within human populations, and between humans and other primates. One type of structural variation, inversions, are segments of the genome that are flipped end-to-end. Among the 150 large inversions segregating in human populations, one extraordinary example is at the 17q21.31 locus. The 17q21.31 locus spans greater than 900 kilobases and comprises ten genes, including *MAPT* and *KANSL1*. Variations in this locus have been implicated in human diseases such as Alzheimer's, Koolen de Vries microdeletion syndrome, and various autism spectrum disorders. There are at least seven major inversion haplotypes worldwide, each exhibiting distinct frequencies across geographic regions and subpopulations. Four haplotypes exist in the direct orientation (H1), and three exist in the inverted orientation (H2). Each of these haplotypes maintain a complex duplication architecture in addition to its inversion status. Some of these haplotypes have also been identified as being under selection, with studies evidencing women who carry these haplotypes presenting higher fecundity rates. Consequently, this locus provides a unique system to study the impact of copy number variation on recombination suppression and extended linkage disequilibrium in human populations.
Here, we report extensive and surprising patterns of temporal and geographic variation at the 17q21.31 locus in several primates and in different modern and ancient human populations. We find that while the inverted haplotype (H2) is the ancestral haplotype in humans, it is the less common haplotype in both archaic and modern-day human samples. We observe an exceptional amount of copy number diversity in direct orientation (H1) haplotypes globally, particularly in understudied modern populations including South Asians. Using Stone Age Eurasian genomes, we model how the distribution and frequencies of different haplotypes have changed over recent human history. As a result of these distribution patterns, we hypothesize that recombination suppression in large and structurally complex regions of the genome contributes to the differential accumulation of deleterious mutations. Finally, we observe several unique instances of ancient double recombination in African populations and date the timing of these events. Together, these results provide valuable insights into the role of chromosomal rearrangements in shaping evolution, diversity, disease susceptibility, and fitness.

Session Title: Evolutionary and Population Genetics Poster Session II

PB3122 The genetic legacy of African Americans from Catoctin Furnace.

Authors:

É. Harney[1,2], S. Micheletti[1], K. Bruwelheide[3], W. Freyman[1], K. Bryc[4], A. Akbari[2], E. Jewett[1], E. Comer[5], H. Gates, Jr.[2], L. Heywood[6], J. Thornton[6], R. Curry[2], S. Ancona Esselmann[1], K. Barca[3], J. Sedig[2], K. Sirak[2], I. Olalde[7], 23andMe Research Team, N. Rohland[8], J. Mountain[4], D. Owsley[3], D. Reich[2]; [1]23andMe, Sunnyvale, CA, [2]Harvard Univ., Cambridge, MA, [3]Natl. Museum of Natural History, Smithsonian Inst., Washington DC, DC, [4]23andMe, Inc., Sunnyvale, CA, [5]Catoctin Furnace Historical Society, Thurmont, MD, [6]Boston Univ., Boston, MA, [7]Univ. of the Basque Country UPV/EHU, Vitoria-Gasteiz, Spain, [8]Harvard Med. Sch., Boston, MA

Few African Americans can trace their family histories to their earliest enslaved African ancestors in the United States. Genetic analysis of historical individuals and identification of present-day close relatives has the potential to restore this knowledge. We analyze 27 individuals buried in Maryland's Catoctin Furnace African American Cemetery during the 18th-19th centuries, and identify identical by descent DNA segments documenting 41,799 relatives among consenting research participants in 23andMe, Inc.'s genetic database of ~12.8 million people. One of the largest geographic concentrations of close relatives is in Maryland, revealing that some descendants of the enslaved and free African Americans at Catoctin Furnace stayed in the area after records indicate their removal from the ironworks, a finding not demonstrable through genealogical analysis alone. We show that many Catoctin individuals likely derived their ancestry from a small number of African groups, specifically the Wolof and Kongo, and that some carried risk factors for Sickle Cell anemia and G6PD deficiency, genetic diseases common in African Americans today. These analyses provide the first direct look at the genome-wide genetic ancestry of enslaved people in the United States, and demonstrate the power of joint analysis of DNA from historical individuals and the extremely large datasets generated through direct-to-consumer ancestry testing.

Session Title: Evolutionary and Population Genetics Poster Session III

PB3123 The in-depth analysis of the Transylvanian Seklers' genetic profile using high-resolution autosomal markers

Authors:

B. Melegh[1,2], V. Ádám[1], K. Sümegi[1,2,3], A. Szabó[1,2], G. Büki[1,2], L. Magyari[1,2], K. Hadzsiev[1,2], M. Kásler[4], Z. Bánfai[1,2]; [1]Univ. of Pécs, Med. Sch., Dept. of Med. Genetics, Pécs, Hungary, [2]Univ. of Pécs, Szentágothai Res. Ctr., Pécs, Hungary, [3]Univ. of Pécs, Med. Sch., Dept. of Biochemistry and Med. Chemistry, Pécs, Hungary, [4]Natl. Inst. of Oncology, Budapest, Hungary

The Seklers, also known as Szekelys in Hungarian, are a Hungarian-speaking Catholic ethnic group residing primarily in Transylvania, Romania. The first written evidence of the Seklers dates back to the 12th century, but their exact origin remains unknown, and their internal population structure has yet to be investigated. In this study, our aim was to characterize their genetic composition by examining subpopulations from various regions of Transylvania. We analyzed a total of 319 Sekler samples collected from 10 different regions within Transylvania. Using Illumina 720K genotype data, we conducted a thorough investigation encompassing allele frequency and haplotype analyses. Our methods included principal component analysis, ancestry analysis based on maximum likelihood estimation, and various DNA segment analysis techniques such as identity-by-descent segment analysis. Our findings reveal that the Seklers of Transylvania exhibit similar ancestry components to a similar degree. This observation was strengthened by both allele frequency and haplotype-based analyses. However, our allele frequency-based investigations highlighted that individuals from the Szentegyhaza and Szovata regions show a significantly larger proportion of one specific ancestry component shared with East Europeans. Based on our investigations, the Seklers residing in Transylvania are generally genetically homogeneous, with the exception of two regions. These areas exhibit a noticeably higher proportion of the East European-West Asian ancestry component, which provides valuable insight into the presumed homeland and origin of the Sekler people. This discovery opens up further opportunities for studying the ancestral roots of the Seklers. Grant References: Supported by the NKFIH 138669.

Session Title: Evolutionary and Population Genetics Poster Session I

PB3124 The longitudinal study of accumulation and inheritance of germline mutations is a way to understand microevolution.

Authors:

H-R. Park[1,2], H-M. Cho[1,2], S-H. Choe[1], M-K. Ko[1,2], Y-J. Lee[1,2], S-J. Park[1], Y-H. Kim[1], J-W. Huh[1,2]; [1]Korea Res. Inst. of BioSci. and Biotechnology, Cheongju, Korea, Republic of, [2]Dept. of Functional Genomics, KRIBB Sch. of BioSci., Korea Univ. of Sci. and Technology (UST), Cheongju, Korea, Republic of

Mutations are caused by errors in the DNA replication process or environmental factors such as radiation or chemicals and occur in germ cells can be transmitted to the next generation. Whether the process of transferring unrepaired mutations to the next generation is repeated, it can proceed to microevolution that changes population gene pool at the species level. In the case of sexual reproduction, compared to female, male germ cells have more opportunities to mutate due to spermatogenesis during whole lifetime. In the present study, we perform whole genome sequencing (WGS) to analysis of somatic/germline mutations accumulated within individuals and patterns of mutations transmitted to offspring by parents' age in crab-eating monkeys (*Macaca fascicularis*). We aim to analyze accumulation patterns of somatic/germline mutation throughout lifetime in crab-eating monkeys, which have yet not to be reported. And based on them, we identify tendency in which germline mutations are transmitted to the next generation and ultimately find out the effect of parental age on the next generation. We selected 4 families (total 16 monkeys; 10 males, 6 females) around various age to get data at all age point and considered that they can produce 3rd generation for detecting transmission to next generation. Also, we considered families with males as 2nd generation offspring in order to directly confirm the effect of transmitted mutations on the reproductive cells of the child, especially sperms. Both somatic cells (blood cells, buccal cells) and germ cells (sperms) were sampled annually to identify total mutations in an individual and sort out germline mutations, from comparative analysis with somatic mutations. As a result of 3 years process, WGS data of monkey families was built up, and based on that, diverse pipelines were devised to analyze mutations accumulated at each sampling time. The envisioned two pipelines are one that considers cell lineage developmental process of three samples and the other that classify *de novo* mutation according to family-related between monkeys. From these analysis, we demonstrate the effect of germline mutations accumulating with age in crab-eating monkeys throughout whole life on next generation. We plan to build a model for germline mutations during the course of aging in parents transferred to offspring, and this could be used as a way to observe microevolution by molecular perspective in crab-eating monkey species.

Session Title: Evolutionary and Population Genetics Poster Session II

PB3125 The nature of selection on common diseases and complex traits

Authors:

N. Connally[1], E. Koch[2], N. Baya[3], B. Neale[4], E. Lander[3], A. Bloemendal[3], S. Sunyaev[2]; [1]Harvard Med. Sch., Brigham & Women s Hosp., Cambridge, MA, [2]Harvard Med. Sch., Brigham & Womens Hosp., Boston, MA, [3]Broad Inst., Cambridge, MA, [4]Broad Inst., Massachusetts Gen. Hosp., Cambridge, MA

The question of how various evolutionary forces maintain variation in complex traits is a mainstay of evolutionary quantitative genetics. Modern human genetics has contributed by demonstrating that effect sizes estimated using GWAS are negatively correlated with allele frequencies. This evidences negative selection on trait-affecting alleles which scales with the magnitude of those effects. For quantitative traits there is good theoretical justification for supposing this reflects the action of stabilizing selection. However, for disease traits one might intuitively expect that selection acts primarily against risk-increasing alleles in order to reduce the prevalence of a deleterious condition. We developed a framework to explicitly compare different models of selection using top GWAS loci from a set of disease, anthropometric, and metabolic traits. All traits under investigation showed a characteristic U-shape when plotting the effect-sizes and frequencies of risk/trait-increasing alleles. When distinguishable from neutrality, we found support (log10 BF>1) for models of stabilizing selection for both disease and quantitative traits. Schizophrenia and Type-2 Diabetes had strongest evidence for stabilizing selection among disease traits. The evidence against directional selection is due to the symmetry of frequencies when conditioning on the risk allele. Long-term directional selection should yield a bias towards risk alleles, yet high-frequency protective alleles are common. Within the realm of stabilizing selection, we also compared a model where the focal trait represents the only relevant fitness dimension (1D) to one where fitness is highly pleiotropic (HD). While we found consistent support for the HD model, we note that it is weaker for disease traits, sensitive to the inclusion of top loci, and not necessarily expected from simulations. Further investigation into the pleiotropic nature of selection on human complex traits is needed.

Session Title: Evolutionary and Population Genetics Poster Session III

PB3126 The Origin and Global Distribution of the 17q21 Inversion.

Authors:

D. Shriner[1], C. Pedicone[2], S. Weitzman[2], A. Renton[2], B. Fulton-Howard[2], A. Adeyemo[3], A. Goate[2], C. Rotimi[4]; [1]Natl. Human Genome Res. Inst., Bethesda, MD, [2]Icahn Sch. of Med. at Mount Sinai, New York, NY, [3]NIH, Bethesda, MD, [4]NIH, Woodstock, MD

A common inversion polymorphism exists at chromosome 17q21.31, with the highest frequencies of the inversion occurring in Italy and Turkey (~30%). The inverted region covers 19 genes, including microtubule-associated protein tau (*MAPT*), which is associated with Alzheimer disease and other tauopathies. The primary assembly of the human reference sequence represents the noninverted orientation and contains a haplotype called H1. The human reference sequence also contains an alternate sequence, chr17_GL000258v2_alt, that represents the inverted orientation and contains a haplotype called H2. Previous studies have identified 20 SNPs that tag H1 and H2 haplotypes. We investigated the genetic diversity of the inversion in 3,569 present-day individuals and 9,990 ancient individuals. Using tag SNP genotypes, we inferred that the present-day data included 6,489 H1 haplotypes, 647 H2 haplotypes, and two H1-H2 recombinants. Using read depth from short-read sequence data, we inferred the presence of zero inverted H1 haplotypes and three noninverted H2 haplotypes, suggesting that the H2 haplotype predated the inversion. Phylogenetic analysis revealed a monophyletic origin of all extant H2 haplotypes. All 37 H2 haplotypes from African individuals occupied derived positions in the phylogeny, consistent with an origin after Out-of-Africa followed by gene flow back to Africa and subsequent introgression. From the ancient DNA collection, the earliest H2 haplotype was in an individual from Taforalt (Morocco) from 12,900-12,100 BCE. No H2 haplotype was detected in 51 individuals from Ice Age Eurasia. The earliest H2 haplotype in Europe was observed in Serbia in an individual from 6,223-6,064 BCE, with genetic similarity to an Anatolian individual from 6,435-6,257 BCE, suggesting a Neolithic introduction of the H2 haplotype from a Middle Eastern source. Among 23,629 African Americans, we detected 2,180 H2 haplotypes, with 99.4% mapping to European ancestry at the locus, likely reflecting recent admixture and introgression from European sources. These results support origins of both the H2 haplotype and the inversion after Out-of-Africa, possibly in the Middle East, and highlight the importance of gene flow in introducing the inversion to new populations over the last 8,000 years.

Session Title: Evolutionary and Population Genetics Poster Session I

PB3127 Threading new data into reconstructed genealogies

Authors:

A. Ignatieva[1,2], S. R. Myers[1]; [1]Univ. of Oxford, Oxford, United Kingdom, [2]Univ. of Glasgow, Glasgow, United Kingdom

The evolutionary history of a sample of genomes can be fully captured by their genealogy in the form of an ancestral recombination graph (ARG): a graph connecting the genomes through shared ancestors in the past, annotated with the mutation events that have shaped the present-day genetic diversity. Inference methods that directly use ARGs as inputs thus offer a powerful way of estimating evolutionary parameters of interest from sequencing data, such as past population sizes or selection coefficients. However, the true genealogy is generally unobserved, and must be inferred from the data at hand: this is a notoriously difficult problem that has seen impressive recent progress, with several powerful and scalable ARG reconstruction tools now available.

The related problem of adding new data to a fixed reference ARG (and avoiding the computational cost of redoing the reconstruction from scratch) has also drawn significant interest. We present a new approximate but principled method for accurately and quickly threading a sequence into an existing ARG. We use the timing of the mutations that are shared by the reference ARG and the new sequence to construct a reduced state space of possible edges with which the new sequence can coalesce at each genomic location, from which the threading position is chosen. Simulation studies with human-like parameters demonstrate that the method maintains excellent accuracy while scaling sub-linearly in the sample size, for both simulated and reconstructed reference ARGs, and running in a fraction of the time compared to *de novo* genealogy reconstruction; the method is also exact in certain cases. Crucially, our method naturally handles issues such as the presence of recurrent mutation, missing data and sequencing errors, and allows for types of data which most ARG reconstruction methods do not accommodate, such as unphased genotypes and non-contemporary samples. This has myriad applications, for instance in adding low-coverage ancient DNA data into previously inferred large-scale human genealogies, allowing for integrated analysis of modern and ancient samples, and improved understanding of selection, trait evolution and demographic history.

Session Title: Evolutionary and Population Genetics Poster Session II

PB3128 Time series data reveal selection on blood group phenotypes over the past 10,000 years in West Eurasia

Authors:

A. Barton[1,2], A. Akbari[3,1,2], D. Reich[4,3,5,2]; [1]Dept. of Human Evolutionary Biology, Harvard Univ., Cambridge, MA, [2]Broad Inst. of MIT and Harvard, Cambridge, MA, [3]Dept. of Genetics, Harvard Med. Sch., Boston, MA, [4]Dept. of Evolutionary Biology, Harvard Univ., Cambridge, MA, [5]Howard Hughes Med. Inst., Harvard Med. Sch., Boston, MA

The discovery that selection on the *HBB* gene conferred resistance to malaria first indicated that blood traits may be important targets of selection in humans; however, identifying adaptation at other blood-related loci has proven difficult. Using a generalized linear mixed model approach and leveraging time series data from ancient DNA (method detailed in separate abstract), we detected evidence of selection on several blood-group related genes over the past 10,000 years in West Eurasia, including *ABO*, *FUT2*, and *GYPA*.

At *ABO*, variant rs8176746 (G>T) defines the B blood-type allele and is under selection (s=3.3%, $P=6x10^{-26}$). ABO blood types have previously been associated with differential resistance to infectious diseases. Our results showed selection on the B allele led to a decrease in the A allele frequency, but a stable frequency of O allele over the past 4,000 years in West Eurasia. Prior studies have supported the hypothesis that the O blood type provides some resistance to infection by malaria. The A and B phenotypes have been suggested to provide protection against cholera, plague and tuberculosis, and blood type B specifically has been proposed to protect against smallpox and norovirus. Similarly, rs601338 (G>A) in *FUT2*, a variant responsible for secretor status of blood type antigens, is also associated with resistance to norovirus and shows evidence for selection (s=1.0%, $P=2x10^{-14}$) during the same period.

Furthermore, genome-wide association analysis demonstrates that variants that define the A and B alleles have opposite effects on some polygenic phenotypes. One such trait is alkaline phosphatase level, which has been associated with differences in chylomicron absorption in the intestine for the A and B type alleles. *ABO* and *FUT2* have been linked to microbiome composition that affects both infection and autoimmune disorders of the bowels. These results along with the observation of selection on linked missense variants rs7658293 and rs7687256 in the *GYPA* gene (s=1.0%, $P=6x10^{-36}$), which contribute to the MNS blood group, support the hypothesis that blood types have been important candidates of selection, and further study of their frequencies in the past might lend insight into their role in human health today.

Session Title: Evolutionary and Population Genetics Poster Session III

PB3129 Tracing the genetic link of Mizo people of India: Genetic & archaeological evidence

Authors:

B. Bankura[1], P. Singh[2], G. Chaubey[2], M. Das[3]; [1]Med. Coll. Kolkata, Kolkata, India, [2]Banaras Hindu Univ., Varanasi, India, [3]Univ. of Calcutta, Kolkata, India

The Himalayan region played a crucial role in the human dispersal and colonization of South Asia. It forms a unique and narrow passage that connects the Indian sub-continent to East and South East Asia. Archaeological and anthropological studies suggest migration and cultural diffusion in this region with the Tibetan plateau in the north. Mizoram is the southernmost state in the northeast region, sharing its borders with Tripura, Assam, and Manipur, along with Myanmar and Bangladesh. The Mizo people are believed to be a part of the Mongoloid feature and Mizoram has the highest number of tribal people among all states of India. Therefore, our study is an attempt to clarify the origins of the Mizo people and their migration roads to their present settlements. Thus, issues of ethnicity and migration are complex but highly relevant to the modern state. A more concentrated effort is required to track these population movements and dispersals linking archaeology and genetics. Plausible migration histories of these groups have been addressed based on linguistic, ethnographical, historical, and folkloristic data, however; less attention has been given to genetics or modern genome mapping. Given the vastness of the region and considerable diversity in the topography, a specific region has been selected for the study: the region adjoining the Indo-Myanmar border in the Champhai district of Mizoram. Therefore, our study is an attempt to clarify the origins of the Mizo people and their migration roads to their present settlements. In the present study, we have investigated 110 individuals belonging to different clans of Mizo people for a genome-wide autosomal marker study. Our results suggest that they are a unique and isolated tribal group and have a very distinct population history. Haplotype and allelic frequency-based analyses indicate that they substantially shared ancestry primarily with East Asian populations and Tibeto-Burman speakers of India but also differentiated from them and became a distinct population with time.

Session Title: Evolutionary and Population Genetics Poster Session I

PB3130 Uncovering the genetic roots of androgenic alopecia in African men: a novel dataset reveals ancestry-specific loci and pathways.

Authors:

R. Janivara[1], Men of African Descent and Carcinoma of the Prostate (MADCaP) Network, T. Rebbeck[2,3], A. O. Adebiyi[4], I. Agalliu[5], J. Lachance[1]; [1]Georgia Inst. of Technology, Atlanta, GA, [2]Harvard TH Chan Sch. of Publ. Hlth., Boston, MA, [3]Dana Farber Cancer Inst., Boston, MA, [4]Univ. of Ibadan, Ibadan, Nigeria, [5]Albert Einstein Coll. of Med., New York, NY

Androgenic alopecia is a highly heritable trait, with a varying prevalence across the globe. Unfortunately, much of our understanding regarding the genetics of male pattern baldness comes from individuals of European descent, and it is unknown whether the genetic architecture of this trait varies by ancestry. Therefore, we examined a novel dataset comprising 2,136 men of African descent recruited from Ghana, Nigeria, Senegal, and South Africa that were genotyped on a custom array that was optimized for sub-Saharan African populations. To remedy existing knowledge gaps, we tested how well recent polygenic risk scores for male pattern baldness generalize to African men, conducted the first African GWAS of androgenic alopecia, and examined functional and evolutionary genetic causes of continental differences in the genetics of this complex trait. Applying polygenic risk scores for male-pattern baldness that were ascertained in men of European descent to African individuals yielded AUC statistics that ranged from 0.514 to 0.546, i.e., genetic predictions only slightly surpassed random chance. Subsequently, we conducted an African GWAS of androgenic alopecia on 19 million genotyped and imputed variants, focusing on baldness at age 45. After correcting for present age, population structure, and study site, we identified 266 moderately significant associations (p-value < 10^{-5}), 54 of which were independent. Marginally significant associations were found at 47 loci. These GWAS loci included previously identified baldness-associated genes such as HIPK1 and IRF4, as well as novel Africa-specific hits. Loci associated with baldness in African men were enriched for cellular processes such as keratinization, olfactory signaling, cell death in skin, and GPCR signaling. Although most of the loci associated with baldness were autosomal, X-linked variants contributed significantly to the SNP-based heritability of this trait in African populations, including intronic variants in the androgen receptor (AR) gene. To understand the source of continental variation in the risk, we inferred the evolutionary ages of baldness-associated alleles and tested whether this trait has been a target of recent natural selection. Further, we analyzed the impact of ancient human introgression on continental differences in the genetic architecture of male pattern baldness. Collectively, our findings underscore the limited transferability of predictions across different ancestries and emphasize the importance of conducting genetic studies encompassing a wide range of populations.

Session Title: Evolutionary and Population Genetics Poster Session II

PB3131 Unearthing the overlooked founder effect of Beauce (Quebec, Canada)

Authors:

M. Gagnon[1], C. Moreau[1], J. Ricard[2], A. Bureau[2], H. Vézina[1], M. Maziade[3], S. Girard[1]; [1]Université du Québec à Chicoutimi, Chicoutimi, QC, Canada, [2]Université Laval, Québec, QC, Canada, [3]Ctr. de recherche IUSMQ, Quebec, QC, Canada

Quebec is known for its founder effects and unique regional characteristics shaped by historical events such as settlement patterns and high natality rate. Sub-regions of Quebec such as Saguenay have been studied for decades for the obvious impact of the past on the genetics of the current population and its consequences on human health. However, our recent research on schizophrenia (SZ) and bipolar disorders (BD) has shed a light on an understudied founder effect that is particularly prominent in the Beauce region despite its proximity to major cities and lack of physical barriers limiting immigration. It is essential to understand the genetic structure of populations for unbiased research on genetic disorders. Therefore, we intend to investigate the founder effect in Beauce and its impact on the population before proceeding with our genomic sequence research on SZ and BD.

Our study focuses on Beauce, its founder effect, and how it impacts the population's genetic structure. We analyzed Illumina OmniExpress or Global Screening Array SNP chip imputed genotypes of 648 Beaucerons from 28 extended families, including both SZ and BD patients, as well as their healthy relatives. They were matched with regional population controls. Additionally, we reconstructed their genealogies spanning back 15 generations using the BALSAC population database. To ensure the independence of our findings from the phenotypes, we also obtained genealogical data for 32,169 supplementary Beaucerons from BALSAC, and performed a comparative analysis. The kinship and inbreeding coefficients showed no significant difference between our patients' families and the BALSAC dataset. With our genetic dataset, we compared Beauce to other well-studied regions using principal component analysis, runs of homozygosity, and identical-by-descent segments. With the genealogies, we characterized them with various measurements such as kinship and inbreeding coefficients.

Our analyses revealed that Beauce displayed higher values than the surrounding populations for the kinship and inbreeding coefficients, and that the genetics of this population substantially differ from the regions of Quebec which encountered no strong regional founder effect.

By investigating the understudied founder effect in Beauce, our results raise the need of considering population structure in genetic research. Our findings advocate for a shift towards investigating geographically and historically distinct subpopulations with unique genetic profiles.

Session Title: Evolutionary and Population Genetics Poster Session III

PB3132 Unraveling Genetic Associations and Ancestral Influences on Fibroproliferative Diseases

Authors:

G. Hampton[1], J. Hellwege[2], M. Puckelwartz[3], Y. Luo[4], D. Velez Edwards[5], T. Edwards[6]; [1]Dept. of Human Genetics, Vanderbilt Univ., Nashville, TN, [2]Div. of Genetic Med., Dept. of Med., and VGI, Nashville, TN, [3]Ctr. of Genetic Med., Northwestern Univ., Chicago, IL, [4]Inst. for Augmented Intelligence in Med., Inst. for Innovations in Dev.al Sci., (IPHAM): Ctr. for Hlth.Information Partnership, Northwestern Univ., Chicago, IL, [5]DRVE affiliation: Div. of Quantitative Sci., Dept. of Obstetrics and Gynecology, Dept. of BioMed. Informatics TLE: Div. of Epidemiology, Dept. of Med., Vanderbilt Genetics Inst., Vanderbilt Univ. Med. Ctr., Nashville, TN, [6]Vanderbilt Genetics Inst., Div. of Epidemiology; Dept. of Med., Vanderbilt Univ. Med., Nashville, TN

Fibroproliferative diseases (FPD), such as asthma and lupus, share a common etiology and often co-occur, suggesting the presence of shared genetic risk factors. Racial disparities reveal higher prevalence among individuals of African ancestry, possibly due to a Th2-favored genome resulting from generational exposure to helminth parasites in Africa. To investigate this hypothesis further, we analyzed global and local ancestry for associations with FPDs. Data on FPDs and genotypes were obtained from electronic health record (EHR) biobanks BioVU from Vanderbilt University Medical Center and the Electronic Medical Record and Genomics (eMERGE) Network. Global and local ancestry analysis employed logistic regressions in BioVU and eMERGE with ancestry proportions based on 1000 Genomes Phase 3 populations, specifically West African (WAFR), East African (EAFR), Southern European (SEUR), and Northern European (NEUR) populations. Regression analyses were conducted separately for self-identified non-Hispanic blacks (NHB) and non-Hispanic whites (NHW), adjusted for sex, age, and body mass index (BMI). Effect estimates are reported per 10% increase in genetically inferred ancestry proportions. Significant positive associations were observed between sarcoidosis with EAFR (odds ratio [OR]=1.17 [95% confidence interval [CI]: 1.12, 1.22]) and WAFR (OR=1.21 [95% CI: 1.16, 1.27]). Conversely, European ancestry demonstrated significant negative associations, with the most pronounced effects seen in sarcoidosis (SEUR: OR=0.87 [95% CI: 0.85, 0.90], NEUR: OR=0.88 [95% CI: 0.84, 0.91]). Local ancestry analysis in BioVU NHB revealed significant peaks at chr2p23.3 for sarcoidosis, chr19p12, chr4p15.33-32 for lupus, and chr10p13-15.1 for asthma. Targeted associations using BioVU NHB GWAS summary statistics identified potential causal SNPs in these regions. Notably, *CELF2* (rs543494131: OR=5.33[95% CI: 2.67, 10.64) was implicated in asthma, *TWIST2* (rs73104883: OR=6.38[95% CI: 2.97, 13.69]) in sarcoidosis, and *ZNF724* (rs188752820: OR=7.35 [95% CI: 3.23, 16.70]) in lupus. These genes have been previously associated with these phenotypes and implicated in T-cell activation and the inflammatory immune response. Our findings illustrate a shared immune-mediated susceptibility to FPDs in NHB, potentially contributing to the increased prevalence of FPDs in individuals of African ancestry. Furthermore, our results indicate a possible protective relationship between European ancestry and the development of FPDs, potentially attributed to variations in T-cell regulation and inflammatory responses.

Session Title: Evolutionary and Population Genetics Poster Session I

PB3133 Unraveling the Haplotype Diversity of the Immunoglobulin Heavy Chain Constant Locus Using Targeted Long-Read Sequencing

Authors:

U. Jana[1], O. Rodriguez[1], A. Dorgham[1], W. Gibson[1], K. Shields[1], C. Silver[1], M. Emery[2], G. Deikus[2], R. Sebra[2], E. Eichler[3], M. Smith[1], C. Watson[1]; [1]Univ. of Louisville, Louisville, KY, [2]Icahn Sch. of Med. at Mount Sinai, New York, NY, [3]Univ of Washington, Seattle, WA

The heavy chain constant region of antibodies (Ab) plays a crucial role in immune function by facilitating antigen neutralization and elimination through various effector functions. These functions are mediated by different Ab isotypes, namely IgM, IgG, IgA, IgE, and IgD, encoded by genes within the immunoglobulin heavy chain constant (IGHC) locus. Despite previous studies highlighting extensive allelic and structural variation (SV) within IGHC, our understanding of haplotype diversity in this region remains limited. Technical barriers have impeded the use of standard next-generation sequencing to resolve these complex loci, underscoring the need for improved genomic resources and tools to effectively characterize IGHC diversity.To address these limitations, we employed targeted long-read sequencing (Pacific Biosciences) of genomic DNA and fosmid clones, combined with the bioinformatics pipeline IGenotyper. This approach enabled us to generate high-quality haplotype-resolved assemblies spanning the IGHC region for six individuals of African, East Asian, and European ancestry, providing comprehensive coverage of IGHC diversity. In our analysis, we identified an elevated number of single nucleotide variants (SNVs; ranging from 2.01 to 5.24 per kilobase) compared to the genome-wide average. Additionally, we observed the presence of large SV including segmental duplications involving the IGHG4 gene, deletions, and inversions. Notably, our study resolved these SVs at the nucleotide level, leading to the discovery of novel IG genes and alleles.Furthermore, we employed these assemblies and two sets of trios to benchmark our novel high-throughput targeted long-read sequencing protocol and bioinformatics pipeline. Our results demonstrated that our approach generates highly accurate assemblies and genotype callsets, outperforming alternative methods. This enables large-scale studies of IGHC genetic variation and their functional effects.By unraveling the haplotype diversity of the IGHC locus, our findings provide an initial understanding of IGHC haplotype diversity. These assemblies serve as valuable resources to accurately interrogate IGHC polymorphisms. This advancement enhances the characterization of IGHC gene diversity at the genome and population levels, crucial for elucidating the roles of germline variation in antibody effector functions in disease and clinical phenotypes.In conclusion, our study offers new insights into the complex IGHC locus and its diversity, enabling further advancements in understanding the impact of genetic variation on antibody-mediated immune responses.

Session Title: Evolutionary and Population Genetics Poster Session II

PB3134 Unravelling the impact of admixture and adaptation on patterns of genetic diversity at the *TAS2R* bitter taste receptor genes in distinct African populations from Cameroon

Authors:

A. Lisi[1,2], X. Liu[2], C. Battaggia[1], R. Johnson[2], P. Anagnostou[1,3], G. Destro Bisol[1,3], M. C. Campbell[2]; [1]La Sapienza Università di Roma, Rome, Italy, [2]Univ. of Southern California, Los Angeles, CA, [3]Istituto Italiano di Antropologia, Rome, Italy

Bitter taste perception is a highly variable trait in humans, and the ability to detect bitter compounds has been largely attributed to genetic variants in 25 bitter taste receptor (*TAS2R*) genes located on chromosomes 5, 7, and 12. Moreover, bitter taste perception has been hypothesized to be a dietary adaptation in humans. While prior analyses have examined patterns of genetic diversity at many of the *TAS2R* genes mainly in non-Africans, relatively few studies have characterized patterns of diversity and signatures of selection at these bitter taste genes in ethnically and culturally distinct African populations. Given the extensive genetic substructure and evidence for local adaptation in response to diet in Africa, we hypothesized that African populations practicing diverse subsistence strategies (such as hunting/gathering and agriculture) could have distinct patterns of selection resulting from divergent diets. To test this hypothesis, we examined sequence variation at 22 bitter taste receptor genes in 120 African hunter-gatherers (Baka, Bakola, and Mbenzele) and Bantu-speaking agriculturalists (Bassa, Ewondo, and Ngoumba) from Cameroon. Here, we identified 353 single nucleotide polymorphisms (SNPs) across the *TAS2R* genes, and among these SNPs 35 of them have never been previously described. We also observed striking departures from neutral evolution at several *TAS2R* genes based on summary statistics—indicative of positive selection—in both hunter-gatherer and agriculturalist populations. To further investigate this evidence for selection, we characterized long-range linkage disequilibrium across chromosomes 5, 7, and 12 in our populations using *i*HS and EHH statistics. These analyses uncovered unusually long haplotype structure around alleles at the *TAS2R* genes (e.g., *TASR2R14*, *TAS2R19*, and *TAS2R20*). In addition, some of these signals of positive selection were shared between hunter-gatherer and Bantu-speaking populations. To better understand these results, we performed genomic genotyping in our hunter-gatherer and Bantu-speaking populations. Interestingly, we observed varying proportions of ancestry originating from Bantu-speakers in hunter-gatherers with the highest levels of admixture occurring in the Bakola population over the last 1,300 years. Indeed, these findings demonstrate that gene flow from neighboring Bantu-speaking agriculturalists has contributed to patterns of diversity in hunter-gatherers, including genetic variation at the *TAS2R* genes. Overall, this study provides new insights into the evolution of biologically relevant bitter taste genes in diverse African populations.

Session Title: Evolutionary and Population Genetics Poster Session III

PB3135 Unveiling Global Disparities in Idiopathic Infantile Hypercalcemia: A Genetic and Population-Based Study

Authors:

A. Tomaszewski[1,2,3], K. Lovrenert[2,3], D. Bodner[3,1], F. Schumacher[4], C-H. Wu[2,3]; [1]Case Western Reserve Univ. Sch. of Med., Cleveland, OH, [2]Dept. of Genetics and Genome Sci., Case Western Reserve Univ. and Univ. Hosp., Cleveland, OH, [3]Dept. of Urology, Case Western Reserve Univ. and Univ. Hosp., Cleveland, OH, [4]Dept. of Population and Quantitative Hlth.Sci., Case Western Reserve Univ. Sch. of Med., Cleveland, OH

Introduction/Objective: Idiopathic infantile hypercalcemia (IIH) is a rare autosomal recessive disorder caused by an impairment in the 24-hydroxylase enzyme encoded by the *CYP24A1* gene. The most common clinical presentation is related to symptoms of hypercalcemia due to increased osteoclast activity and decreased calcium excretion, eventually resulting in nephrolithiasis and nephrocalcinosis. Children receiving vitamin D supplementation are especially at an increased risk of developing these complications due to their inability to appropriately regulate calcium levels. The clinical prevalence of IIH is not well characterized. This study seeks to identify global superpopulations in which IIH may be more common to implement improved screening protocols and prevent complications from delayed treatment or inappropriate vitamin D supplementation.
Methods: Known pathogenic variants of *CYP24A1* causing IIH were procured from the Human GeneMutation Database (HGMD). The 1000 Genomes (1KG) database was used to identify healthyindividuals with *CYP24A1* variants. HGMD and 1KG databases were intersected to find sharedvariants, and prevalence was calculated using the Hardy-Weinberg equation. Individual allele frequencies were calculated using the same method divided into the following superpopulations: Africans (AFR), Admixed Americans (AMR), East Asians (EAS), Europeans (EUR), and South Asians (SAS).
Results: After intersecting the HGMD and 1KG databases, three disease-causing variants were identified in 8 heterozygote individuals. This yields a carrier rate of 1 in 314 and an affected rate of 1 in 391,876. Importantly, these variants were identified only among the AMR and EUR superpopulations at carrier rates of 1 in 174 and 1 in 168 and affected rates of 1 in 120,409 and 1 in 28,112, respectively. None of the pathogenic variants were found in the AFR, EAS, or SAS populations.
Conclusion: There exist marked disparities in the affected rate: 1 in 120,409 and 1 in 28,112 in the AMR and EUR populations respectively, and zero in AFR, EAS, or SAS populations. The probable causes for these variations could stem from the genetics of IIH and CYP24A1 being less researched in non-white populations. Additionally, there could be a founder effect present in the EUR/AMR populations, and it is plausible that pathogenic variants possess higher frequencies within EUR and AMR. Further research is imperative for the identification of other pathogenic variants within diverse populations in addition to more comprehensive evaluation of potential differences in carrier and affected rates for IIH across different populations.

Session Title: Evolutionary and Population Genetics Poster Session I

PB3136 Using 858,635 individuals' haplotype-sharing between British and Danish populations to infer the North Sea migration history

Authors:

X. Zhang[1], T. Fitzgerald[1], K. Banasik[2], S. Brunak[2], E. Birney[1]; [1]European Bioinformatics Inst., Hinxton, Cambridge, United Kingdom, [2]Univ. of Copenhagen, Copenhagen, Denmark

The North Sea has been a strategically important area with an extensive history of people and goods flowing through and around. However, it remains unclear how the historical migration events across the North Sea have affected the genetic structure of the modern human population in its coastal nations. Here we use the pairwise haplotype sharing from 858,635 British and Danes, to assemble population genetic evidence and infer the migration history between British and Danish populations within the past 1,000 years.

Using nation-wide healthcare registries in Denmark, we analyze the haplotype sharing from 370,259 Danes. Compared with previous studies that reveal the Danish population as highly genetically homogeneous, we identify discernible population structure based on haplotype sharing, which reconstructs the geography of Denmark. We use the haplotype sharing between 370,259 Danes and 488,376 British from UK Biobank to detect regions with excess haplotype sharing across the two populations. We find individuals from South Jutland share the most of their genome with British while English from Eastern coast spanning from Yorkshire to the South East share the most of their genome with Danes. Finally, we characterize the change of the haplotype-sharing patterns over time and estimate the time periods when the patterns emerge.

Our work provides a novel fine-scale view on haplotype sharing patterns within and across Denmark and Britain, identifying regions of interests to population history, and demonstrating the use of multiple nationwide biobanks as a powerful tool to study history across populations.

Session Title: Evolutionary and Population Genetics Poster Session II

PB3137 Using networks of short IBD to identify ethnolinguistic-related groups in African population

Authors:

J. Zhang[1], A. Sedghifar[1], N. Swinford[2], R. Curtis[2], A. Kim[1], J. Lang[1]; [1]Ancestry.com DNA, LLC, San Francisco, CA, [2]Ancestry.com DNA, LLC, Lehi, UT

Long chromosome segments of identity by descent (IBD) which are suggestive of recent common ancestry can be used to estimate how people are related to one another while distant matches mostly reflect population or demographic history instead of family history and can be difficult to distinguish from shared haplotypes. Here we look at networks of short IBD tracts in samples from Africa to learn about population structure. We used public and non-public samples and the samples cover 19 countries, 129 tribes and 44 language families from Africa. We detected ancestral groups by subdividing the IBD network into unsupervised clusters using the Louvain Method and taking country, tribe, and language information into consideration. We found that most of the samples in the same group came from a single country or adjacent countries and people from the same population or language family are mostly likely to be assigned to the same group. In addition, we observed some of the groups have samples from many different countries and samples from the same country could be assigned to different groups. There were also cases where samples from the same population or language family had been assigned to different groups. In conclusion, groups that we have detected were showing geography and population/language family aggregation while capturing extra relatedness that is different from what country or which population people are coming from.

Session Title: Evolutionary and Population Genetics Poster Session III

PB3138 Variation and novelty: Evolutionarily new genes enable protein structural innovation and function in the brain

Authors:

V. Luria[1,2,3], A. Karger[2], J. Cain[4], K. Nam[2], D. Marten[5,6], M. Kirschner[2], A. O'Donnell-Luria[5,6], N. Sestan[1]; [1]Yale Univ., New Haven, CT, [2]Harvard Med. Sch., Boston, MA, [3]Boston Children's Hosp., Boston, MA, [4]Harvard Univ., Cambridge, MA, [5]Broad Inst. of MIT and Harvard, Cambridge, MA, [6]Boston Children's Hosp., Boston, MA

How genuinely new protein-coding genes originate is a central question in biology. Many novel genes arising de novo from intergenic genomic DNA long considered to be "junk", or from long non-coding RNAs, were recently found in Eukaryotes. Novel genes are taxon-restricted, may encode structurally novel proteins with new protein domains, and thus illuminate the emergence of protein structure. To understand how novel genes arise, we built a mathematical model based on gene and genome parameters, and dynamic factors such as mutation rates. Combining phylostratigraphy and proteogenomics, we identified novel genes in over 100 eukaryotic genomes ranging from human to paramecium and evaluated their predicted biophysical properties. Compared to ancient proteins, novel proteins are shorter, more fragile, disordered and promiscuous, yet less prone to forming toxic aggregates. We experimentally measured structure content and protease resistance of novel human proteins and showed novel genes function in vivo in zebrafish brains. We showed novel human genes have fewer regulatory elements than ancient genes but more than control intergenic open reading frames. Several novel genes show depletion for loss of function variation in reference population data such as gnomAD. In a GTEx RNA expression analysis, we found that novel human genes are expressed in many tissues. Using single-cell RNA-Seq and proteomics, we found novel genes are expressed in human brains at multiple ages. Thus, genomic sequence turnover generates many novel genes encoding short proteins with distinct structural features and function in the brain. Variation in large eukaryotic genomes having large intergenic "dark matter" regions continuously generates new protein structures and new functions.

Session Title: Evolutionary and Population Genetics Poster Session I

PB3139 Visualization of geographic variation in linkage disequilibrium across human populations

Authors:

A. Zagal Norman[1], L. Anderson-Trocmé[1], A. Biddanda[2], J. Novembre[1]; [1]Univ. of Chicago, Chicago, IL, [2]Johns Hopkins Univ., Baltimore, MD

Regions of the human genome are co-inherited in large blocks due to meiotic recombination and as a result, mutations exhibit non-random association along the genome -- or linkage disequilibrium (LD). Large cohorts of human genomic data have revealed a rich diversity of genetic variation and patterns of LD that reflect the impact of evolutionary and demographic processes over time. LD patterns can substantially impact the interpretation of analyses of human genetic variation, playing a critical role in determining the statistical power of genome-wide association studies, fine mapping causal variants for complex traits, and highlighting the differential impact of natural selection across populations. Current methods for visualizing LD are designed to be used for single populations, often by representing the covariance between markers around a single locus in the form of a LD matrix. However, there are currently no tools that can visualize LD patterns at a single locus across multiple populations. To address this, we developed geoLD, a user-friendly, web-based tool capable of displaying LD matrices across multiple populations at the same time. geoLD enables rapid queries of LD matrices centered on a specific genomic coordinate across multiple populations within seconds. The visualizations rendered are initially interactive but can be exported to publication-ready graphics for loci of interest. We additionally provide an implementation of geoLD to run as a local offline instance, for smaller-scale exploration in education or field settings or when public sharing of genomic data may be restricted.

Using data from the 1000 Genomes Project, we provide visualizations of LD at loci related to complex trait associations and signatures of natural selection. Specifically, our visualizations show patterns of LD patterns at random regions of the genome, and the outlier patterns generated by strong instances of positive and negative selection respectively.

Beyond human genetics, we envision that our tool may be of use for visualizing LD patterns across any organism with genetic polymorphism data and geographically sampled populations (e.g. Arabidopsis, Drosophila, etc.). geoLD is also well-suited to support educational curricula on human genetic diversity, particularly for geographic differences in haplotypic patterns relative to single allelic variants. We anticipate that geoLD will provide intuitive visualizations for researchers and educators in human genetics.

Session Title: Evolutionary and Population Genetics Poster Session II

PB3140 Whole genome sequencing of 1,000 individuals clarifies the ancestral admixture of the Ugandan population and identifies variants associated with neonatal hydrocephalus.

Authors:

M. Movassagh[1], C. Hehnley[2], E. E. Jessica[3], M. Peterson[3], L. Hadley[4], M. Sasanami[4], A. Whalen[1], J. Smith[3], S. Kumar[5], J. Mugamba[6], P. Ssenyonga[6], J. Onen[6], K. Burgoine[7], E. Kumbakumba[8], E. Wegoye[6], D. Natukwatsa[6], R. Mulondo[6], E. Mbabazi[6], C. Fronterre[4], J. Broach[3], J. Paulson[9], S. J. Schiff[1], S. Morton[10]; [1]Yale Med. Sch., New Haven, CT, [2]Boston Children's Hosp., Boston, MA, [3]Penn State, Hershey, PA, [4]Lancaster Univ., Lancaster, United Kingdom, [5]Dana Farber Cancer Inst., Boston, MA, [6]Cure Intl., Mbale, Uganda, [7]Mbale Regional Referral Hosp., Mbale, Uganda, [8]Mbarara Univ. of Sci. & Technology (MUST), Mbarara, Uganda, [9]N-Power Med., Redwood, CA, [10]Boston Children`s Hosp. / Harvard Med. Sch., Boston, MA

Hydrocephalus is a complex neurological disorder marked by an abnormal accumulation of cerebrospinal fluid. Infant hydrocephalus (IH) is the most frequent indication for early neurosurgery in sub-Saharan Africa. Genetic risk contributes to 10% of congenital hydrocephalus in prior cohorts composed of mainly European ancestry. To determine the genetic contribution to IH in Uganda, we performed whole genome sequencing (WGS) of blood samples from 400 infants with IH and 600 infants without IH (non-IH). Using Ward or Classification and Regression Tree clustering, four populations were identified in the IH cohort and three in the non-IH cohort. In the overall cohort, Admixture analysis identified four populations. Moreover, we observed a discernible genetic distance between these populations, as evidenced by a calculated Fst index of 0.004. After filtering for variants with a minor allele frequency >0.1, 988 samples were included in an genome-wide association study (GWAS) to identify variants associated with IH. After accounting for batch effects detected through principal component analysis, we conducted a GWAS using Plink2.0. Our analysis revealed 11 specific genomic regions, some of which were either overlapping or in close proximity to CDC42, CRB2, YBX2, and CDR2L, that were associated with IH. To facilitate our analysis and shed light on their intricate relationship with the presence of pathogens, we developed a computational method called mbQTL, which allowed us to report on the association between the identified genetic variants and the presence of pathogens through various novel statistical approaches. Through the integration of metagenomic and WGS approaches, our study provides a comprehensive understanding of the genetic landscape underlying hydrocephalus susceptibility in the Ugandan population. These findings also offer potential avenues for the development of targeted interventions and personalized treatments.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session I

PB3141 6q25.2q25.3 duplication and 2q14.1q14.2 deletion due to a three-generation familial chromosomal insertion is identified in a proband with the unbalanced form presenting with multiple congenital anomalies

Authors:

N. DeMetrick[1], J. Biscoe[1], S. Morsey[1], V. Stinnett[1], J. Murry[2]; [1]Johns Hopkins Hosp., Baltimore, MD, [2]Johns Hopkins Sch. of Med., Baltimore, MD

Background: Chromosomal insertions are uncommon structural rearrangements involving two or more chromosomes. Most balanced insertions lack any phenotypic consequence; however, unbalanced rearrangements can confer a phenotype associated with multiple congenital anomalies.
Methods: Conventional cytogenetic analysis was performed on peripheral blood in a child presenting with multiple congenital anomalies (developmental delay, cognitive delay, and growth abnormalities). Fluorescent in situ hybridization (FISH) studies using standard DiGeorge and subtelomeric-specific probes proved unremarkable. Given the high clinical suspicion for a chromosomal abnormality, Single Nucleotide Polymorphism (SNP) array using the Illumina® 850k BeadChip was performed.
Results: SNP array showed an interstitial deletion on chromosome 2 from cytobands q14.1 to q14.2 that was 7.6 Megabases (Mb) in size, along with an interstitial duplication on chromosome 6 between cytobands q25.2 and q25.3 (9.4 Mb in size). To clarify the orientation for these loci, FISH was performed using custom probes mapping within both copy number changes. FISH confirmed an unbalanced insertion involving chromosomes 2 and 6, whereby the gain of chromosome 6 material was inserted into the deleted loci on chromosome 2. A high-resolution karyotype was repeated on this patient, revealing a subtle change within chromosome 2 at the same bands identified by array. As the region involved did not differ in its banding pattern relative to the normal chromosome 2 homolog, visualization of the insertion was difficult to discern by karyotype alone. However, parental testing using the same informative FISH probes revealed that the unbalanced insertion was maternally derived and that the proband's maternal grandmother also carried the balanced form.
Conclusion: We confirmed the suspected chromosome abnormality in the proband using complementary cytogenomic methods. In this case, the genomic imbalance was identified upon array testing and would have been missed if using conventional cytogenetics alone. Other family members should be tested to see if they are carriers of this insertion for genetic counseling and estimation of recurrence risk. Emerging cytogenomic technologies may be considered to clarify the rearrangement further.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session II

PB3142 A case of 17p13.3 microdeletion with breakpoint in the gene for *CRK*

Authors:

I. Ohashi[1,2], J. Matsuda[3], M. Nagata[4], Y. Ishihara[4], Y. Miyashita[4], Y. Asano[4], Y. Yamanouchi[5,2], K. Takao[2], K. Tawa[2], T. Otomo[2,6], K. Kurosawa[7], M. Masuno[8]; [1]Mitoyo Gen. Hosp., Kanonji, Japan, [2]Dept. of Med. Genetics, Kawasaki Med. Sch. Hosp., Kurashiki, Japan, [3]Dept. of Pathophysiology and Metabolism, Kawasaki Med. Sch., Kurashiki, Japan, [4]Dept. of Cardiovascular Med. (IRUD Analysis Ctr.), Osaka Univ. Graduate Sch. of Med., Suita, Japan, [5]Kawasaki Univ. of Med. w, Kurashiki, Japan, [6]Dept. of Molecular and Genetic Med., Kawasaki Med. Sch., Kurashiki, Japan, [7]Kanagawa Children s Med Ctr, Yokohama, Japan, [8]Genetic Counseling Program, Graduate Sch. of Hlth.and Welfare, Kawasaki Univ. of Med. Welfare, Kurashiki, Japan

The Miller-Dieker syndrome caused by deletion of the 17p13.3 region, including PAFAH1B1, is well known, and case reports of 17p13.3 microdeletion syndrome, characterized by facial dysmorphic manifestations, developmental delay, growth disturbance, and abnormal brain structure, caused by deletion of two genes, YWHAE and CRK, have become more common. We have now seen scattered case reports of 17p13.3 microdeletion syndrome. In this presentation, we report a case of 17p13.3 microdeletion with a breakpoint in the CRK gene and the deleted region is not contained with PAFAH1B1 or YWHAE. The patient, a 1 year and 2 months old boy, consulted with our clinic with the chief complaint of small for gestation age, significant failure to thrive, and relative macrocephaly. There was no family history of growth disturbance. Physical examination revealed height (-2.8 SD), weight (-4.0 SD), head circumference (+0.4 SD), prominent forehead, hypertelorism, and inverted triangular face. Enjohji developmental test revealed his DQ was 97.5 at 1 years old. Echocardiography and head MRI showed no significant abnormal findings. Endocrinological examination revealed thyroid function within normal range, but IGF-I was markedly low. Chromosome analysis (GTG) showed normal male karyotype, and methylation analysis performed to differentiate Silver-Russell syndrome also showed no abnormalities. In whole exome sequencing, it was not detected disease-specific variant. Array CGH showed an interstitial deletion of approximately 830 kb in the 17p13.3 region, and the region contained 19 protein-coding genes and breakpoint at p-terminal was within the CRK gene. CRK gene is conjugated a proto-oncogene. It is not only involved in carcinogenesis, but is also thought to be related to growth regulation. Overgrowth in duplicated cases and growth restriction in deleted cases have already been reported. It was found that growth retardation can occur partial deletion of the CRK gene in the present case. Growth hormone administration has been shown to improve growth retardation in complete deletions of CRK, and is expected to be effective in this case as well. Array CGH analysis is useful for the diagnosis of unexplained short stature and growth disturbance and can predict the effect of treatment.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session III

PB3143 A case with critical risk for X chromosome aneuploidy in non-invasive prenatal testing (NIPT); confirmed to deletion of Xq27.3-q28 and duplication of Xq28

Authors:

J. Kim[1], H. Kim[1], J. Park[1], K. Kang[1], H. Jang[2], M. Go[3], S. Yang[1], D. Cha[4], S. Shim[5]; [1]CHA Biotech Inc., Seoul, Korea, Republic of, [2]CHA Univ., Seoul, Korea, Republic of, [3]CHA Univ., Seoul, Korea, Republic of, [4]Gangnam CHA Hosp, Seoul, Korea, Republic of, [5]Cha Univ., Seongnam, Korea, Republic of

Background: Non-invasive prenatal testing (NIPT) has been widely used for prenatal screening of common fetal aneuploidy such as trisomy 13, 18, 21. According to several large-scale cohort studies, the positive predictive value (PPV) of NIPT varies depending on the chromosomes, with ranges of 50%~73% for trisomy 21, 20%~58% for trisomy 18, and 4%~47% for trisomy 13. NIPT has been expanded to include sex chromosome aneuploidies (SCAs), rare chromosomal aneuploidies (RCAs), and copy number variants (CNVs), although the PPV for SCAs is low. **Case:** A 38-year-old pregnant woman was referred for NIPT at 12+1 weeks of gestation because of advanced maternal age. No ultrasound abnormalities were identified. There was no specificity in her anthropometric characteristics such as weight, height, and body mass index (BMI). NIPT result showed a critical risk of sex chromosome [fetal fraction (FF) = 6.89%]. For accurate diagnosis, amniocentesis was recommended, and conventional karyotyping and chromosome microarray (CMA) were performed. Fetal karyotype was 46, XX but, CMA results showed a 3.6 Mb deletion of the Xq27.3q28 region [arr Xq27.3q28(146,806,191_150,386,543)x1] including *FMR1* and immediately followed by a 4.8 Mb duplication of the Xq28 region [arr Xq28(150,408,165_155,233,731)x3] including *MECP2.* Parents showed normal karyotypes. At first, it was assumed that the abnormal X was inherited from the mother, but maternal CMA result showed no deletion or duplication at these sites. Deletion of *FMR1* gene and duplication of MECP2 in the fetus were confirmed by PCR/Southern blot and multiple ligation-dependent probe amplification (MLPA) methods, respectively. And X chromosome inactivation (XCI) assay showed highly skewed X-inactivation pattern of the paternal X chromosome. From the above results, it can be assumed that the fetus inherited a normal X chromosome from the mother and an abnormal X chromosome from the father. The paternal abnormal X chromosome might be occurred *de novo* in spermatogenesis. Paternal CMA and SNP-array to identify the origin of the abnormal X chromosome was not performed and the follow-up observation related to fetal development could not be performed due to abortion. **Conclusion:** In this study, we report a case of critical risk of X chromosome in NIPT results, and a case in which genetic counseling based on the result of an invasive diagnostic test and additional molecular genetic test was provided as an important reference for prenatal decision-making.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session I

PB3144 A combination of cytogenetic and molecular genetic tools, including long-read sequencing, resolves multiple breakpoints in a patient with complex structural variation.

Authors:

J. Knijnenburg, S. Kloet, Y. M. C. Hendriks, T. T. Koopmann, W. van den Ende, M. E. Y. Laurense-Bik, K. Heijboer, F. de Vries, G. W. E. Santen; Leiden Univ. Med. Ctr., Leiden, Netherlands

A combination of different cytogenetic and molecular genetic tools is often needed to unravel complex translocations and structural variation to a base pair level. The current first-tier tests in patients with developmental delay and/or multiple congenital anomalies are CNV and SNV analysis using SNP-array and/or whole exome sequencing. These approaches are in general not capable of detecting structural variation such as (complex) balanced translocations or rearrangements.
In a patient with dysmorphisms, hypotonia and developmental delay, genome wide CNV analysis on whole exome captured short-read NGS data revealed two relatively large interstitial deletions in the long arms of chromosome 13 and chromosome 14, both around 5 Mb in size. SNV analysis did not reveal additional variants that could explain the phenotype. Overlapping deletions, especially of the 14q region, were previously published in patients with similar phenotypic features, making pathogenicity likely.
Both interstitial deletions appeared to be de novo and (balanced) carriership of the deleted regions in both parents was ruled out with CNV analysis and FISH. Subsequent karyotyping and FISH on metaphases of the proband confirmed the 13q and 14q deletions, as expected. Surprisingly, these classic cytogenetic tests additionally showed involvement of the terminal part of the long arm of chromosome 3, which was exchanged with the part of chromosome 13, distal to the deletion.
To resolve the complete nature of the complex structural variation found in the presented patient, we eventually performed genome-wide long-read sequencing using the Oxford Nanopore PromethION sequencing technology. This technology allowed us to determine the composition of the derivative chromosomes, clearly showing the diagnostic potential of long-read sequencing.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session II

PB3145 A distinct genotype and phenotypes in pediatric patients with biventricular noncompaction

Authors:

K. Hirono[1], Y. Hata[1], K. Tsuboi[1], S. Takarada[1], M. Okabe[1], H. Nakaoka[1], S. Ozawa[1], N. Nishida[1], F. Ichida[2]; [1]Univ. of Toyama, Toyama city, Japan, [2]Intl. Univ. of Hlth.and Welfare, Tokyo, Japan

Background: Ventricular noncompaction (VNC) is a hereditary type of cardiomyopathy characterised by prominent trabeculaitons. VNC mainly involve the left ventricle, however, isolated right ventricles (RV) and biventricular noncompaction (BiVNC) also occur. This study aimed to clarify the clinical characteristics and genetic landscape of BiVNC. **Methods:** We enrolled 234 children (127 male and 107 female; mean age 4 months [range, 0-6.6 years]) with VNC from 2013 to 2021 in Japan. VNC was classified as follows: BiVNC, congenital heart defects (CHD), arrhythmia, dilated phenotype (DCM), or normal function. In the enrolled patients, 203 genes associated with cardiomyopathy were screened using next-generation sequencing (NGS). Cardiac death, left ventricular (LV) assist device implantation, heart transplant (HT), and appropriate implantable cardioverter-defibrillator (ICD) shock were classified as major adverse cardiac events (MACE). **Results:** Twenty-five patients were enrolled in BiVNC group, 38 in CHD, 32 in arrhythmia, 84 in DCM, and 55 in normal function group. Patients with BiVNC was significantly younger at diagnosis than those with other VNC phenotype ($P < 0.0001$). Eight patients (32.0%) and nine patients (36.0%) in BiVNC group had perinatal abnormality. Poor LVEF was more common in the DCM and BiVNC group than in other groups ($P < 0.0001$). Dyskinesis of LV region was most frequently observed in BiVNC group. Noncompacted layer was most prominent in BiVNC. There were 140 pathogenic variants in 120 patients with VNC including all phenotypes (51.3%): 108 missense, 9 deletion, 6 insertion, 10 nonsense, and 7 splice-site variants. Overall, the *MYH7* gene was most commonly identified (22.1%), followed by the *TAZ* (7.4%) and *TPM1* (6.6%) genes. As for BiVNC patients, there were 11 pathogenic variants were in 11 patients (44.0%): 10 missense, 1 insertion, 10 nonsense. Variants in genes associated with sarcomere group, ion channel group, mitochondrial group, heart development group, and desomosomal group accounted for 45.5%, 9.1%, 18.2%, 18.2%, and 9.1%, respectively. Of note, novel variants in *NKX2-5* gene and *TBX5* gene were identified in BiVNC patients. In the BiVNC group, MACE was noted in 5 patients (20.0%). The BiVNC group had the worst survival rate among all groups in contrast to the CHD group ($P = 0.0009$). **Conclusions:** This is the first study that focused on genotype-phenotype correlations in a pediatric patients with BiVNC. Our study suggests that comprehensive screening of multiple disease-causing genes may help identify VNC patients with specific cardiac phenotypes who are at a high for MACE.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session III

PB3146 A novel DNA collection method for genomic sequencing in infants

Authors:

A. Nagy[1], L. M. Mahanta[1], L. Hutchinson[1], S. Adelson[2], L. M. Sack[1,2,3], S. T. Ryall[1,3,2], B. Zettler[2], I. A. Holm[4,3,5], R. C. Green[3,2,6,5], M. S. Lebo[1,3,2,5], H. Zouk[1,3,7]; [1]Mass Gen. Brigham Lab for Molecular Med., Cambridge, MA, [2]Brigham and Women's Hosp., Boston, MA, [3]Harvard Med. Sch., Boston, MA, [4]Boston Children's Hosp., Boston, MA, [5]The Broad Inst. of MIT and Harvard, Cambridge, MA, [6]Ariadne Labs, Boston, MA, [7]Massachusetts Gen. Hosp., Boston, MA

Introduction: Genomic sequencing (GS) is becoming more common in both diagnostic and screening settings and is being explored for use in newborn screening across many clinical studies. However, obtaining samples from infants can be challenging. Specifically, venipuncture is invasive and other non-invasive techniques such as saliva sampling yield inconsistent results. Additionally, while dried blood spot cards could be used for sequencing, they are not always readily available. Moreover, the workflow for extracting DNA from these two other sample types can be labor-intensive. We aimed to explore alternative DNA collection methods that are less invasive, yet yield consistent results. To this end, we investigated whether the use of volumetric absorptive microsamplers (VAMS) for blood collection in infants would reliably yield sufficient high-quality DNA for GS. Methods: Blood samples from 17 infants at Boston Children's Hospital's Primary Care Center (CHPCC) were collected through a heel stick into Neoteryx Mitra® VAMS devices (4x 30ul tips) for DNA yield optimization for GS. Extraction using membrane spin columns (QIAamp DNA Investigator Kit, QIAGEN) was determined to be the most reliable method and as such, blood microsampling from an additional five infants was used for its validation. Nine families of infants with a previous microsampling collection were surveyed about their preferences and comfort with this method as compared to venipuncture. Results: All 5 microsamples extracted on membrane columns yielded enough high-quality DNA for GS despite an incomplete tip collection for 3 of the samples. The mean yield was 1.7 ug (range 0.8-3.3 ug) for infants and was directly correlated with a qualitative estimate of the proportion of tips filled. The majority of families surveyed (8/9, 88.9%) indicated their preference for heel stick microsampling for potential future blood collection from their infant. This method has since been adopted by the BabySeq Project, which is exploring the clinical utility of GS for screening in presumably healthy infants. Since recruitment began, an additional 7 samples have been extracted in our laboratory, with an average yield of 2.6 ug (range 1.3-3.5 ug). Conclusions: High quality DNA for GS can be isolated from infant blood samples collected on VAMS devices. Additionally, this collection method was found to be acceptable and preferable by the infants' families. This noninvasive method has been used successfully in various clinical applications including immunological and pharmacokinetic studies, but now has the potential to be leveraged for GS in newborns and in other patient cohorts, expanding access to genomic medicine.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session I

PB3147 A Novel Method for the Detection of 40 Amino Acids Relevant to Inborn Errors of Metabolism, in Under 60 Minutes Using Reverse Phase High Performance Liquid Chromatography

Authors:

K. Wilson[1,2]; [1]Children's Natl. Hosp., Washington, DC, [2]Natl. Human Genome Res. Inst., Bethesda, MD

Assessment of amino acid concentrations is important for the diagnosis and management of a number of inborn errors of metabolism (IEMs). There have been many methods developed that attempt to measure these biologically important metabolites, however, they are often cost-prohibitive and/or excessively time consuming. Currently, most amino acid assessments are conducted using either ion-exchange chromatography, which is a multiple hours long process that uses expensive dedicated technologies for amino acid analysis, or more recently mass spectrometry detectors which greatly increase the cost. We have previously developed a methodology for the rapid, repeatable, and cost-efficient separation of approximately 20 amino acids and now, based on this initial methodology, we have expanded it to additional amino acids relevant to IEMs. We thus describe here our protocol of amino acid separation using reverse phase high performance liquid chromatography (RP-HPLC) combined with ultraviolet-visible spectrum absorbance (UV-VIS) paired with derivatization just prior to injection with the agent o-phthaladehyde (OPA). We show reproducibility via concentration assessments for all amino acids separated in the method, in triplicate. We assess amino acids not only in prepared standard solutions but also in biologic samples with pathologic amino acids added to confirm separation. Quantitation was established using a single internal standard. We achieved a less than 60min separation runtime with our lower cost systems that obtain good separation of 40 primary amino acids as well as quantitation of ammonia in a single run. As the need for analysis of amino acids grows, and thus the number of assessments conducted both in the US and globally, so does the need for improved methodologies that decrease both the cost and time associated with accurate measurement. This rapid and accurate methodology that only requires a small volume of sample could greatly increase efficiency in many settings and has broad applications for both clinical utility and research investigations.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session II

PB3148 A sequencing first approach improves access to a precise genetic diagnosis in kids newly diagnosed with atypical development

Authors:

K. Dipple[1,2,3], D. Doherty[1,2], K. Anderson[1], O. Sommers[1], T. Wenger[1,2], K. MacDuffie[1], A. Keefe[1,2], M. Sikes[2], A. Scott[2], L. Kruidenier[2], H. Gildersleeve[1], C. Davis[1], K. Buckingham[1], J. Chong[1], D. Veenstra[1], D. Miller[1,2,3], D. Copenheaver[4], J. Juusola[4], K. Retterer[4], K. McWalter[4], A. Snook[4], P. Kruszka[4], J-H. Yu[1], M. Bamshad[1,2,3]; [1]Univ. of Washington, Seattle, WA, [2]Seattle Children's Hosp., Seattle, WA, [3]Brotman Baty Inst., Seattle, WA, [4]GeneDx, Gaithersburg, MD

Background and Hypothesis: The Seq First Developmental Differences (DDi) project aims to improve early and equitable access to a precise genetic diagnosis (PrGD) for children with atypical development by providing whole genome sequencing (WGS) upon identification of abnormal development, rather than after a conventional staged approach to evaluation and testing. We hypothesized that a simplified clinical workflow and early access to WGS would lead to a PrGD for more participants than for those receiving conventional care. **Methods:** Participants were referred for the study from one of two Early Intervention Centers (EICs) or an academic Neurodevelopmental clinic from 4/2021 to 6/2023. Eligibility criteria included less than 3 years of age with moderate or severe delay alone, or mild developmental delay plus an abnormal growth parameter or organ system. Exclusion criteria included prior evaluation or genetic testing for atypical developmental or a non-genetic explanation. To date, 198 participants have been referred, 28 were ineligible, 76 declined, and 94 were enrolled. Each eligible participant was randomized to either an immediate WGS (test; n=49) or control (conventional care for two years followed by WGS; n=45) group. Results from WGS were returned to families in the test group by a clinical geneticist. **Interim Results:** Of the 49 participants who received WGS, 21 (49%) had a PrGD that explained or partially explained their atypical development. In the control group, who were enrolled for at least 6 months (n=31), 22 (71%) had access to genetic testing. Yet in only 5 (16%) was a PrGD made that explained or partially explained their atypical development. **Discussion/Conclusions/Future directions:** Compared to the control group, a simplified clinical workflow in the test group including broad inclusion criteria led to increased access to a PrGD. Moreover, despite a more inclusive approach to offering testing, the overall diagnostic rate remained high. These observations suggest that children who might not otherwise be offered testing, or offered testing later in childhood, could benefit substantially (earlier access to precise interventions, reduced morbidity, etc.) and that this strategy could reduce healthcare costs (fewer non-specific evaluations, tests, etc.). Whether these benefits are realized is under evaluation. Surprisingly, recruitment from EICs proved challenging and nearly 40% of eligible families declined to participate. Identifying the hurdles to referral and enrollment and developing alternative strategies (e.g., community or primary-care based) will be critical to ensuring we build equitable capacity for a PrGD.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session III

PB3149 A universal blood genotyping array tailored for research by transfusion services achieves high accuracy in a large, ethnically diverse cohort

Authors:

J. Schmidt[1], J. Gollub[1], **R. Varma**[1], E. van der Schoot[2], A. Harmer[3], C. Westhoff[4], W. H. Ouwehand[5,6], B. Veldhuisen[2,7], N. Gleadall[6], W. J. Lane[8,9]; [1]Thermo Fisher Scientific, Santa Clara, CA, [2]Dept. of Experimental Immunohematology, Academic Med. Ctr., Univ. of Amsterdam, Amsterdam, Netherlands, [3]Histocompatibility and Immunogenetics Lab., NHS Blood and Transplant, London, United Kingdom, [4]Immunohematology and Genomics Lab., New York Blood Ctr., New York, NY, [5]Dept. of Haematology, Univ. Coll. London Hosp., London, United Kingdom, [6]Dept. of Haematology, Univ. of Cambridge, and NHS Blood and Transplant, Cambridge, United Kingdom, [7]Sanquin Diagnostic Services, Amsterdam, Netherlands, [8]Dept. of Pathology, Brigham and Women's Hosp., Boston, MA, [9]Harvard Med. Sch., Boston, MA

Background: Extended blood group matching substantially reduces sensitization to non-self antigens as a side effect of transfusion. For broad application, affordable high-throughput typing of relevant antigens is required. Following an earlier proof-of-concept study, our consortium presents here the development of a tailored universal blood donor genotyping array, including the results of an international study comprising an ethnically highly diverse panel of samples. **Methods:** The custom-designed, research use only Applied Biosystems™ Axiom™ UBDT_PC1 array contains 39,000 probes tagging 20,000 variants relevant for investigation of blood typing and other blood-related phenotypes. It has a 384-sample format and runs on the GeneTitan-MC™ instrument, which can generate data for 3,000 samples/week. The array allows simultaneous typing of human erythroid (HEA), platelet (HPA), and leukocyte (HLA) antigens. For the study, DNA samples and antigen typing data from 13,908 donors provided by seven blood services were analyzed. Samples representing diverse ethnic groups (74% European, 11% African, 15% others) were genotyped at three blood services in the Netherlands, USA, and UK. Array-inferred antigen types were analyzed for concordance with provided antigen types for the first 6,953 samples. **Results:** Reproducibility between sample results from the testing laboratories was outstanding for the vast majority of the 20,000 variants. The overall concordance for HEA, HPA, and HLA antigens with previous testing was high. For ~100,000 comparisons between blood service determined HEA antigen types and array determined types, the concordance exceeded 99.8%. Over 80% of the discordances can be resolved by simple algorithmic modifications and over half of the remaining ones were caused by incorrect serology. The results of the study were highly accurate for non-European genotype configurations. The 778 samples of African ancestry showed a HEA concordance rate >99.7%. The HLA class I and II concordance level, which could be assessed for 1319 DNA samples, was excellent. **Conclusions:** An affordable and comprehensive DNA-based test for automated high-throughput typing of donors is reported here. The universal blood typing array is tailored for the research needs of transfusion services and evaluated in an international, diverse cohort. The array represents a promising new tool to facilitate broader research in extended blood matching and identification of rare blood types.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session I

PB3150 Advancing Precision Diagnosis through Simultaneous Detection of Diverse Variant Types: Insights from Clinical Whole Genome Sequencing

Authors:

H. Dai[1], R. Rigobello[2], C. Settler[3], M. Chau[4], W. Bi[5], C. Eng[5], P. Liu[5], L. Meng[6], N. Owen[1], L. Vossaert[6], C. Wu[5], F. Xia[6], X. Zhao[5]; [1]Baylor Coll. of Med. / Baylor Genetics, Houston, TX, [2]Baylor Genetics, Houston, TX, [3]Baylor Genetics, Hoston, TX, [4]Baylor Coll. of Med., Houston, TX, [5]Baylor Coll. of Med./Baylor Genetics, Houston, TX, [6]Baylor Coll. Med./Baylor Genetics, Houston, TX

Introduction: Genetic disorders arise from diverse types of molecular variants. Historically, clinical genetic testing assays could only assess for specific types of variants. This resulted in many patients with complex phenotypes remaining undiagnosed or required to submit to multiple rounds of testing with alternate assays to reach a diagnosis. Whole genome sequencing (WGS) now offers healthcare providers the ability to test for many types of variants with one assay, providing a powerful tool to pinpoint an etiology accurately and more rapidly for these patients, thus avoiding the diagnostic odyssey. Here, we summarize our experience performing clinical WGS at Baylor Genetics to highlight its utility as a first-tier diagnostic test capable of detecting a spectrum of variant types. **Methods:** This is a retrospective evaluation of WGS results at Baylor Genetics. For each patient, the type of variant(s) reported as well as demographic data, clinical history, diagnostic findings were investigated. **Results:** We reviewed 233 cases that had diagnostic findings by WGS related to reported phenotype. Single nucleotide variations (SNVs) and small indels were the most common variant types, observed in 180 cases (77.3%). Copy number variations (CNVs) were reported in over one-fifth of the total cases (n = 46, 19.7%). Other variant types, such as short tandem repeats (STRs) (n = 4), chromosomal abnormalities (n = 16), and variants within the mitochondrial genome (n = 3), comprised 9.9% of the total solved cases. 70.4% (n = 164) of total solved cases were explained by SNV and indel findings alone. An SNV/indel and a non-SNV/indel finding leading to at least one diagnosis were identified in 16 cases. Variant type did not correlate significantly with clinical history. **Conclusion:** This review demonstrates that up to 30% of patients with a diagnostic WGS finding related to phenotype would not have been captured by assays only focusing on SNV/indels. These data affirm the utility of WGS as a comprehensive test for detecting an assortment of molecular variants to make a diagnosis. As many patients with hard-to-diagnose phenotypes are critically ill, this data supports WGS as an expedient tool which can improve patient outcomes and avoid diagnostic odysseys.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session II

PB3151 Aligning timely diagnosis of Duchenne Muscular Dystrophy with return of results from molecular diagnostic testing

Authors:

M. Hegde[1], A. Martin[2]; [1]PerkinElmer, Waltham, MA, [2]Parent Project Muscular Dystrophy, Washington, DC

Molecular diagnosis for Duchenne and Becker muscular dystrophies (DMD/BMD) has evolved significantly in last[EO1] two decades with advances in molecular technologies. It is now possible to sequence the entire 2.2MB *DMD* gene sequence using next generation sequencing in a single assay to detect all copy number and sequence variants in both index males and carrier females. Full-sequence analysis of the entire gene allows for detection of deep intronic pathogenic variants and accurate breakpoint detection of CNVs involving similar exons, which could have an impact on the outcome of clinical trials. This comprehensive assay is highly sensitive for diagnostic testing for DMD and is also suitable for confirmatory testing for newborn screening for DMD. Despite these advances, the time to diagnosis for DMD is still stagnated at ~5yrs. We performed a review of the data generated from a clinical laboratory offering *DMD* testing using single gene testing, neuromuscular and autism panels, exome and genome sequencing. *DMD* testing performed for molecular confirmation of clinical diagnosis has a diagnostic yield of 90% (diagnostic yield from a sponsored testing program for DMD/BMD). Upon retrospective analysis of >1000 panels, exome and genome samples, a *DMD* pathogenic variant was identified in 1.5% of the cases not specifically referred for DMD. This detection rate signifies the importance inclusion of *DMD* in neuromuscular and autism panels, given that many individuals have reported to be on the autism spectrum in the early years without specific muscular signs[EO2] . A subsequent analysis of consent forms for reporting of secondary pathogenic gene findings revealed a wide variation in opt in and opt out selections for panels, exomes and genomes. There is heterogeneity in how labs approach inclusion of secondary/incidental genetic findings within their reports to clinicians. Of the 8 labs researched, 25% of labs default to not providing secondary findings unless the form completer opts in to receiving them within the report; 25% of labs will include secondary findings in reports by default unless the completer opts out; 25% of labs will not provide any results (assumed) unless the completer chooses between opting in or out of receiving secondary findings within the report; and 25% of labs make no mention of secondary findings within their requisition/consent forms, which makes it unclear whether or not these findings are included by default. Given the detection rate of 1.5% pathogenic variants in *DMD* when a targeted *DMD* test is not ordered prompts the importance of return of results of secondary findings, which may facilitate early diagnosis and intervention.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session III

PB3152 Allelic drop-out (ADO) caused by variability in primer-binding sites and beyond influences an accuracy of PCR-based sequencing methods

Authors:

E. Zaklyazminskaya, A. G. Shestak; Petrovsky Natl. Res. Ctr. of Surgery, Moscow, Russian Federation

Background: PCR-based sequencing methods have a risk of allelic dropout (ADO) phenomenon which leads to selective allele amplification in the PCR process. This phenomenon reduces the accuracy, zygosity counts, and diagnostic yield of the genetic testing.
Methods: We compared BAM, VCF files (NGS) from 232 patients with selective ABI (Sanger) chromatograms to identify the cases of ADO. The primer binding sites were analyzed using the gnomAD database to reveal variations causing ADO. All amplicons with suspected ADO cases were re-sequenced using the alternative oligoprimers pairs.**Results:** We have identified 8 cases of ADO in both targeted genes panels (NGS) and Sanger sequencing results. The fact of selective allele amplification was confirmed in all cases by re-sequencing using an alternative pair of primers. Most cases of ADO (6 cases, 75%) were caused by common or rare single nucleotide genetic variants (SNVs) within primer-binding sites. In two cases ADO was presumably mediated by the presence of six- or more nucleotides indels located beyond primer-binding sites. Size of amplicons may also play a role in some cases of the selective allele amplification. Additionally, we found the under-representation of several SNVs on the NGS reads or the presence of the SNV only on reads of one amplicon out of two. We hypothesize that selective allele amplification may affect up to 1% of amplicons what is not negligible considering the growing amount of the massively parallel sequencing.**Conclusion:** ADO may decrease a DNA-diagnostic yield in both NGS and targeted Sanger sequencing approaches. We have demonstrated the impact of "point" substitutions in primer- binding sites and indels in amplicons on the amplification efficiency.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session I

PB3153 An efficient design for whole genome trio sequencing identifies key variants in rare neurological disease cases.

Authors:

S. Kruglyak[1], K. Ramsey[2], M. Naymik[2], B. R. Lajoie[1], K. N. Blease[1], M. Sanchez-Castillo[2], S. Billings[1], W. Jepsen[2], M. Huentelman[2], V. Narayanan[2]; [1]Element BioSci.s, San Diego, CA, [2]TGen, Phoenix, AZ

We sequenced nine trios in which the probands in an underserved population from Sonora, Mexico were affected by a rare and undiagnosed disease with neurological features. Families were enrolled in the study, with proband phenotypes including developmental delay, intellectual disability, language and motor difficulties, cardiac anomalies, a blood disorder, and/or short stature. We sequenced one trio per sequencing run, leveraging the design of sequencing the proband at twice the coverage of the parents. The reduced coverage in the parents led to sequencing efficiencies while retaining the benefits of trio sequencing: the ability to discover de novo variants and the ability to trace inheritance patterns of rare variants. Once the sequencing data was generated, our two teams used independent informatics pipelines for variant calling and interpretation. This enabled us to determine whether the two labs would arrive at the same conclusions when starting with the same data sets. In five of the nine cases, both teams found a single SNV or small indel that was deemed causal pending clinical validation. The findings matched exactly despite independent informatics pipelines used for variant prioritization. In three of the nine cases, both teams failed to identify the causal variant. In the final case, an additional scan for large CNVs performed by one of the teams identified a de novo duplication in the proband which is the likely cause of the underlying disease. The sequencing of the parents with the proband was critical. In addition to establishing de novo calls and verifying the inheritance of homozygotes, we were able to disqualify potentially pathogenic heterozygous calls that were observed in one of the (healthy) parents. Pending confirmation via clinical sequencing, we anticipate a 6/9 or 67% diagnostic yield for the study. The results across cases with significant findings showed a variety of affected genes (CFAP52, DYNC1H1, FANCE, TCF4, and TOP3A), variant types, and inheritance patterns. In three cases, the presumed causal variant was de novo, including a frameshift insertion, a splice altering SNV, and a CNV. Intriguingly, the other three cases showed a homozygous SNV or indel that was extremely rare according to gnomAD yet carried by both (unrelated) parents. Following confirmation via clinical sequencing, families will be counseled about potential therapies, disease management, current research studies (i.e. gene therapy), and family planning. This study not only demonstrated a successful sequencing approach for rare disease but also has the potential to provide families with a means of moving forward toward finding treatment options.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session II

PB3154 An efficient variant phasing utilizing a replication cycle reaction system

Authors:

A. Mitsutake[1,2], H. Ishiura[1,3], T. Matsukawa[1], J. Mitsui[4], S. Tsuji[5], T. Toda[1]; [1]Dept. of Neurology, Graduate Sch. of Med., The Univ. of Tokyo, Tokyo, Japan, [2]Dept. of Neurology, Intl. Univ. of Hlth.and Welfare Mita Hosp., Tokyo, Japan, [3]Dept. of Neurology, Okayama Univ. Graduate Sch. of Med., Dentistry, and Pharmaceutical Sci., Okayama, Japan, [4]Dept. of Precision Med. Neurology, Graduate Sch. of Med., The Univ. of Tokyo, Tokyo, Japan, [5]Inst. of Med. Genomics, Intl. Univ. of Hlth.and Welfare, Chiba, Japan

[Introduction] When two heterozygous variants are detected in genes with autosomal recessive inheritance, it is required to determine whether the two variants are located in *cis* or in *trans*. Although the analysis of gDNAs of the parents is usually employed, we face difficulties when parental DNAs are unavailable. Subcloning of long-rage PCR products obtained from gDNA is useful, but long-range PCR fails to amplify segments usually exceeding 10kb. Subcloning of cDNA is another possibility, but its success depends on the availability of cells expressing mRNA. Droplet digital PCR is useful even when distance between variants is up to 100 kb, but designing specific probes remains challenging. To accomplish amplification of a large genomic DNA segment carrying two heterozygous variants, we employed a replication cycle reaction (RCR) (Su'etsugu M, et al. Nucleic Acids Res. 2017;45:l525-34). [Methods] gDNAs were cleaved by CRISPR/Cas9 in such a way that two variants were in cleaved fragments, which were then ligated to *oriC*-cassette containing *rop*, *ori*, and *AmpR* from pBR322. Various molar ratios of gDNA to *oriC*-cassette (1:10, 1:100, 1:1000, 1:10000) were evaluated to determine the optimal conditions for the ligation step. RCR amplification of each sample was conducted in 5-8 tubes. RCR products in each tube were then subjected to direct nucleotide sequence analysis. When both alleles were amplified in RCR products, RCR products were transformed into *E. coli* (HST08). The resultant colonies were picked, followed by direct nucleotide sequence analysis. We analyzed gDNAs from seven patients carrying two heterozygous pathogenic variants in *SYNE1*, *CYP27A1*, *ATP7B*, *COQ4*, or *CLCN2* with distances ranging from 4.3 to 152 kb. [Results] The molar ratio considerably influenced RCR success. At a high ratio (1:10), success rate was low (4/24), and only one allele was amplified (4/4). In contrast, at a low ratio (1:10000), success rate was high (20/24), and both alleles were amplified (18/20). Intriguingly, we observed clonal amplification in six of the seven cases, resulting in RCR products containing only one allele. In one case, both alleles were simultaneously amplified. Transformation of *E. coli* with the RCR products enabled us to separate the alleles and to determine the zygosity. The limit of amplification range was 84 kb. In the case where the distance was 152 kb, we were able to reconstruct the haplotype by utilizing a heterozygous SNP located 78 kb away from the variant. [Discussion] The study showed that RCR enables phase-resolved analysis for the region up to 84 kb, which is useful for determination of zygosity for two heterozygous variants.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session III

PB3155 An interstitial deletion of chromosome 6q16 in an infant with multiple congenital anomalies: A case report.

Authors:

N. Hara, K. K. Gulen, X. Tang, C. Kumar; Westchester Med. center, Valhalla, NY

Interstitial deletions of chromosome 6q16 have been associated with Prader-Willi Syndrome-like (PWS-like) phenotype in most patients, featuring obesity, hypotonia, short extremities, and developmental delay. Recent studies have shown that certain rare single-minded 1 (SIM1) loss-of-function variants were associated with a high intra-familial risk for obesity with or without features of PWS-like phenotype. Although SIM1 seems to have a key role in the phenotype of patients carrying 6q16 deletions, some data support a contribution of other genes, such as GRIK2, to explain associated behavioral problems. To date a study has confirmed that the PWS-like phenotype is strongly linked to 6q16.2q16.3 deletions and varies considerably in its clinical expression. Recently, an unusual finding of two de novo microdeletions in cis position on chromosome 6q16.1q16.2 and 6q16.3 have been reported in a 19-year-old patient with intellectual disability and autism.

Here we present a new case of a pre-term male baby born with intra-uterine growth restrictions at 31 weeks 2 days by C-section due to mother's severe preeclampsia. He was born small for gestational age with a birth weight of 1.59 kg (~45 percentile). He had respiratory failure requiring ventilating support, cardiomegaly, thrombus in atrium, germinal matrix hemorrhage and hyperbilirubinemia. The mother had an amniocentesis due to prenatal findings of bilateral clubfeet and hypoplastic corpus callosum that revealed an interstitial deletion of chromosome 6q16. The cytogenetic and chromosomal microarray investigations were indicated. He underwent chromosome analysis on GTG banded metaphases prepared from peripheral blood cultures revealed the karyotype as 46,XY,del(6)(q16.3q16.3). The chromosomal microarray performed on whole blood for further characterization of the deletion and detected 8.2 megabase interstitial deletion from 6q16.1 to 6q16.3. The deleted interval involves 25 known genes (18 protein coding), including SIM1. Overlapping deletions of this region that include SIM1 and the other genes. Clinical correlation and genetic counseling were recommended. Parental chromosome and microarray studies were recommended to determine whether this abnormality is inherited or de novo in origin.

In conclusion, this case demonstrates the effective use of cytogenetics and microarray studies to characterize the interstitial deletion of 6q16 for patient's diagnosis and clinical management. Further research will improve our understanding of how loss of other genes in the 6q16.1q16.3 region may contribute to the genesis of neurodevelopmental diseases and other phenotypic disorders.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session I

PB3156 Analysis *GNAS* inactivation disorders in a large Korean cohort with pseudohypoparathyroidism using whole exome sequencing (WES) and methylation-specific multiplex ligation-dependent probe amplification (MS-MLPA).

Authors:

H. LEE; seoul Natl. Univ. Hosp., seoul, Korea, Republic of

Background: Pseudohypoparathyroidism (PHP) is caused by pathogenic variants (PHP-Ia or PHP-Ic) or epigenetic changes (PHP-Ib) at the complex *GNAS* locus on chromosome 20q13.3. Here, we investigated 53 patients suspected with pseudohypoparathyroidism using whole exome sequencing (WES) and methylation-specific multiplex ligation-dependent probe amplification (MS-MLPA).
Methods: A total of 53 patients was enrolled. To identify pathogenic variants in the coding exons of *GNAS*, sequence analysis of exon 1 through 13 was performed using WES. MS-MLPA was performed to identify deletions in *STX16*, and epigenetic changes at the four differentially methylated regions within *GNAS*.
Results: Among 53 patients, forty-nine ones proceeded with both WES and MS-MLPA, three only with MS-MLPA, and the last one with only WES. A pathogenic *GNAS* variant was identified in 9.4% (5/53) and abnormal epigenetic changes were identified in 45.3% (24/53). Four patients with abnormal epigenetic changes were *STX16* deletion-related familial PHP-Ib and the remaining 20 patients could be classified as sporadic PHP-Ib. The most frequent pathogenic variant, c.565_568delGACT resulting in frameshift mutation was discovered in 3 of 5 patients. The other two mutations were c.348_349insC and c.197_198insA respectively.
Conclusion: This study demonstrates molecular mechanisms of *GNAS* inactivation disorders in a large Korean cohort with pseudohypoparathyroidism. In this cohort, sporadic PHP-Ib was the most common, and followed by inactivating *GNAS* variant and *STX16* deletion-related familial PHP-Ib. Overall, the distribution of PHP patients shown in this study is expected to serve as meaningful data for promptly and accurately diagnosing PHP patients in the future.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session III

PB3158 Array-CGH analysis in a cohort of Czech patients with intellectual disabilities and autism spectrum disorders

Authors:

M. Fiser, H. Kucerova, M. Schwarz, R. Michalovska; PRENET - Prenatal Diagnostics and Genetics, Pardubice, Czech Republic

Background In this study, we used array comparative genomic hybridization (aCGH) for identification of copy number variants (CNV) in a cohort of Czech patients with intellectual disabilities (ID) and/or developmental delay (DD) and/or autism spectrum disorders (ASD). **Methods** Array CGH is a high-resolution technique that can detect submicroscopic deletions/duplications. The genomic DNA extracted from blood samples of 223 patients was analyzed. All patients had normal karyotype and tested negative for fragile X syndrome. We performed aCGH using 8x60 ISCA format (Agilent, Santa Clara) with an average resolution of 240 kb (48 kb in ISCA regions). **Results** Aberrations of significance/potential significance were found in 43 (19,3%) of the 223 patients. *SHOX* gene duplications were found in 4 of the 43 patients. Recurrent microdeletion/microduplication syndromes were found in 15 of the 43 patients. Aberrations of candidate genes were found in 12 of the 43 patients. 6 of the 43 aberrations are linked to well-described genes. Large aberrations encompassing more genes were found in 6 of the 43 patients. **Conclusions** All data are consistent with literature. There seems to be a significance in frequency of *SHOX* gene duplications. Aberations of *SHOX* gene are well described as the cause of Léri-Weill Dyschondrosteosis, Langer Mesomelic Dysplasia and Idiopathic Short Stature. Only few papers describe association of *SHOX* gene duplications with developmental delay and/or autism spectrum disorders. Our study suggests potential significance of *SHOX* gene duplications in ID, DD and/or ASD, but more evidence is needed and further studies are necessary.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session I

PB3159 Artificial Intelligence Based Chromosome Analysis

Authors:

S. Pallavoor; CARAIO Technologies, Bangalore, India

Human chromosome analysis through **Karyotyping** involves identification of 23 pairs of Chromosomes in human cell, out of which the first 22 pairs are autosomes and the 23rd pair is sex chromosome. This process is typically performed by experts in conventional cytogenetics. However, expertise across karyotyping and cytogenetics in general are very less and are mostly limited to cities across world.
With ever increasing number of cases for chromosomal disorders being seen worldwide, consisting predominantly cases such as down syndrome, infertility and repeated pregnancy loss, there is pressing need to make karyotyping available much beyond the current levels to make early diagnosis possible. This calls for widespread medical diagnostic tests for chromosome analysis of prenatal babies, newborn babies and adults screened for karyotyping. The main problems being faced in making karyotyping available for widespread adoption are **accessibility** and **affordability**, with heavy dependency on limited availability of **skilled cytogeneticists**. Adoption of **Artificial Intelligence** and **Digital Transformation** can solve this problem to a great extent, by way of application of image processing and computer vision techniques through the implementation of deep learning and other machine learning algorithms. We can achieve automation of processes involving heavy cognitive skills towards identifying chromosomes from digital images of metaphase precisely within few seconds which would otherwise take an expert in cytogenetics to spend anywhere between ten to twenty minutes to do the job per image. Medical diagnosis report generation typically would need about twenty digital images of different metaphase for a sample to be analyzed normally, and in cases of any abnormalities it could be even higher. AI can provide steep reduction in time and effort towards this. Suitable combination of human decision making in conjunction with machine produced results can ensure reliable level of accuracy needed for such decision-making processes. Karyotyping is the Gold Standard for detecting structural anomalies and numerical anomalies in chromosome defects in humans. Even newer techniques like Microarray can handle only at DNA level, not at structural and numerical levels of chromosomes. Karyotyping is the most suitable form of chromosome test for testing Amniotic Fluids. Through digital transformation, we can achieve **tele-medicine** for Karyotyping by provision of the AI platform-based software through cloud computing. This would solve the two predominant issues namely **accessibility** and **affordability** to a great extent for making Karyotyping available for the wider population.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session II

PB3160 Artificial intelligence improves early diagnosis in delayed puberty

Authors:

y. zouaghi[1,2], M-C. Antoniou[1], X. Cheng[3], G. Papadakis[1], S. Stoppa-Vaucher[1,4], Z. Kolesinska[1,5], B. Therese[1], I. L. Ruiz Arana[1], E. Elowe-Gruau[1], F. Phan-Hug[1], J. Pasquier[6], M. Adamo[1,2], J. Deladoey[7,8], J-M. Vuissoz[7,8], B. Pierre-Alain[1], K. Busiah[1], M. Hauschild[1], j. Acierno[1], F. Santoni[1,2], N. Pitteloud[9]; [1]centre Hosp.ier universitaire vaudois, lausanne, Switzerland, [2]Univ. of Lausanne, Lausanne, Switzerland, [3]Univ. of Toronto, Toronto, ON, Canada, [4]Hôpitaux Neuchâtelois, Neuchâtel, Switzerland, [5]Poznan Univ. of Med. Sci., Poznań, Poland, [6]Univ. of Lausanne, lausanne, Switzerland, [7]Pediatric Inst. of Southern Switzerland, Bellinzona, Switzerland, [8]Univ. of Southern Switzerland, Lugano, Switzerland, [9]centre Hosp.ier universitaire vaudois, Lausanne, Switzerland

Background/Objectives: Delayed puberty (DP) in boys is defined by the absence of puberty by age 14. Most individuals with DP exhibit a transient form of GnRH deficiency and will eventually complete puberty (constitutional delay of growth and puberty [CDGP]). A minority of individuals have a permanent form of GnRH deficiency with absent or partial puberty known as congenital hypogonadotropic hypogonadism (CHH). The aim of this study is to use machine learning approaches to facilitate the early clinical distinction between these two conditions thus helping the clinicians to design the appropriate therapeutic strategy. **Methods:** 41 male patients aged 14 to 16.5 years with delayed puberty (DP) and hypogonadotropic hypogonadism were followed up until final diagnosis (18-19 years). We employed an anomaly detection algorithm, Isolation Forest to distinguish these two groups. We used 21 clinical and biochemical markers such as testicular volume, testosterone and gonadotropin levels from the first visit and a polygenic score from rare variants in known CHH causative genes to predict final diagnosis. **Results:** Out of 41 DP patients, 27 had CDGP, 6 complete CHH (cCHH) and 8 patients exhibited partial CHH (pCHH). Isolation Forest, trained with clinical parameters, enabled identification of pCHH from CDGP with a diagnostic accuracy of 79%, and reached 86% when including all CHH patients. The addition of a polygenic score based on exome variants further improved this performance by an additional 4%. **Conclusion:** Machine learning with polygenic scores improves diagnostic accuracy up to 89% to differentiate CHH from CDGP and can decrease greatly the difficulty for early diagnosis.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session III

PB3161 Automated decision support for the clinical interpretation of copy number variants

Authors:

J. Budiš[1,2,3], T. Sládeček[1,2], M. Gaziova[1], M. Kucharik[1,2], Z. Pos[1,4,5], O. Pos[1,2,4], W. Krampl[1,2,4], R. Hekel[1,4,3], T. Sedlackova[1,2], J. Radvanszky[1,2,5], T. Szemes[1,2,4]; [1]Geneton Ltd., Bratislava, Slovakia, [2]Sci. Park, Comenius Univ. in Bratislava, Bratislava, Slovakia, [3]Slovak Ctr. of Scientific and Technical Information, Bratislava, Slovakia, [4]Dept. of Molecular Biology, Faculty of Natural Sci., Comenius Univ., Bratislava, Slovakia, [5]Inst. of Clinical and Translational Res., BioMed. Res. Ctr., Slovak Academy of Sci., Bratislava, Slovakia

Introduction: Clinical interpretation of copy number variants (CNVs) is a complex process that often leads to different decisions between diagnostic centers. In order to alleviate this obstacle, technical standards for the interpretation and reporting of variants have been developed, called ACMG guidelines. Several semiautomatic computational methods have been proposed to recommend appropriate choices. Alternately, the emerging machine learning based tools showed promising ways of even fully automated predictions.
Materials and Methods: We evaluated state-of-the-art computational prediction tools on CNV records collected from the ClinVar database. We compared their accuracy and showed the superior accuracy of their combination. We also tweaked the key parameters and underlying data sources to demonstrate their impact on the overall prediction.
Results: We demonstrate that the choice of underlying data sources significantly affects the prediction of CNV pathogenicity. We also show how automated prediction tools, especially their combination, further improve the evaluation of ACMG guidelines and thus lead to more reliable clinical decision support for clinicians.
Conclusions: Modern clinical decision support tools for the clinical interpretation of CNVs are able to provide valuable guidance to clinicians, relieving them of a great share of tedious annotation and interpretation processes.
Grants: This work was supported by the PANGAIA project H2020-MSCA-RISE-2019 (Grant agreement ID: 872539) funded under H2020-EU.1.3.3. Programme; by the ALPACA project H2020-MSCA-ITN-2020 (Grant agreement ID: 956229) funded under H2020-EU.1.3.1. Programme; by the Operational Program Integrated Infrastructure within projects ITMS: 313011F988 and ITMS:313021BUZ3, co-financed by the ERDF.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session I

PB3162 Automated ultra-high molecular weight DNA isolation from various sources enables high-throughput optical genome mapping.

Authors:

A. Hoischen[1], T. Hofste[1], L. Vullings[1], K. Huijbers[1], R. van Beek[1], E. Kamping[1], B. van der Sanden[1], J. Donner[2], S. Nothjunge[2], D. Laubscher[2], B. Ottenwälder[2], A. Files[3], Y. Zhang[3], M. Yadav[3], K. Neveling[1]; [1]Dept. of Human Genetics, Radboud Univ. Med. Ctr., Nijmegen, Netherlands, [2]Hamilton Bonaduz AG, 7402 Bonaduz, Switzerland, [3]Bionano Genomics Inc., San Diego, CA

The power of long-read genomic technologies is becoming increasingly apparent in both human genetics research and in clinical applications. One emerging technology that utilizes Ultra-High Molecular Weight (UHMW) DNA is Optical Genome Mapping (OGM) via Bionano Genomics, Inc. (San Diego, USA). We also refer to OGM as 'next generation cytogenetics', as this may replace most cytogenetics tests including karyotyping, fluorescence in situ hybridization (FISH) and CNV-microarrays. While the analytical part of OGM is largely automated, the pre-analytical part and mainly the isolation of UHMW DNA would allow faster uptake of routine use. To date, automated UHMW DNA isolation has been very challenging, mainly due to the fact that one aims to prevent fragmentation and to preserve its megabase-long molecular size. Here, we demonstrate for the very first time the automated isolation of 12 UHMW DNAs each from three different cell types: cell pellets, EDTA blood and bone marrow aspirates using the Long String VANTAGE robotic solution from Hamilton (Bonaduz, Switzerland). This resulted in UHMW DNA that matched or exceeded DNA quality of manual procedures for all samples; resulting in molecule-N50 of at least 230kb and fulfilling standard labeling quality metrics. The subsequent processing of this UHMW DNA on Bionano's Saphyr enabled genome-wide structural variant (SV) detection. Here, we demonstrate its utility to detect >10,000 structural variants (SVs) >300bp in germline analysis; and show its sensitivity for acquired i.e. somatic SVs of all types with variant allele fractions (VAF) down to 5%. The final SV data of manual vs. automated procedure are indistinguishable. This approach has recently led to the discovery of hidden pathogenic SVs in previously undiagnosed rare disease patients. The demonstrated use of fully automated UHMW DNA isolation will allow higher throughput and robustness - and will therefore further improve clinical research and routine use of OGM. Ongoing developments focus on automation of the downstream labeling procedure to additionally facilitate the scaled adaptation of OGM workflows.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session II

PB3163 Automation and miniaturization of high-throughput qPCR for gene expression profiling

Authors:

S. Tomei, S. Raveendran, A. Saeed, R. Mathew, H. Manjunath, M. Kalikiri; Sidra Med., Doha, Qatar

Background. Quantitative PCR (qPCR) is a technique commonly employed in laboratories and core facilities. QPCR results are strictly dependent on pipetting accuracy. Automation solutions overcome the variability generated by manual pipetting. The Mosquito HV liquid handler is a new generation automation system that offers multichannel pipetting with additional capability of assay miniaturization. In our previous study, we had shown the possibility to automate steps in the Standard BioTools gene expression workflow by pairing Mosquito HV with BioMark HD (Manjunath HS *et al*., 2022). Here we aimed at automating the full workflow and at exploring miniaturization capabilities.**Methods.** Each step of the Standard BioTools gene expression workflow was programmed on Mosquito HV genomics software. Mosquito HV liquid handling system was utilized for all pipetting steps followed by qPCR on Biomark HD. We performed three different runs: 1. Reference RNA samples (pooled healthy individuals; n=10) were run on an immunology gene expression panel. We tested the full reaction (FR) and three miniaturization conditions: 1.5x, 2.5x and 5x. 2. Test RNA samples (isolated from n=45 individuals) were run as FR and 1.5x on the immunology gene expression panel. 3. Test RNA samples (isolated from n=45 individuals) were run as FR and 1.5x on a pregnancy gene expression panel. The expression of each gene was calculated by $2^{-\text{deltaCt}}$ method. **Results**. In the reference RNA samples, successful amplification was observed when using FR and 1.5x conditions with a success rate of 96.81% and 93.42% respectively. The 2.5x reaction exhibited suboptimal amplification with a success rate of 37.55% while 5x reaction showed no amplification. Therefore, 2.5x and 5x conditions were excluded from further runs. A significant positive correlation was observed between the manual and automated workflows (manual vs automated FR, $r=0.9542$, $p<0.0001$; and manual vs automated 1.5x reaction, $r=0.9477$, $p<0.0001$). The validation of the automation workflow using patient samples with the immunology panel resulted in a success rate of 97.7 % and 96.8 % for FR and 1.5x reactions respectively. A positive correlation was observed with FR and 1.5x ($r=0.9$; $p<0.0001$). Consistent with the immunology panel, a positive correlation was also observed when using the pregnancy gene expression panel between FR and 1.5x reactions ($r=0.9952$; $p<0.0001$). Our results show that adoption and miniaturization capabilities of Mosquito HV system for automating the gene expression workflow did not interfere with data quality and reproducibility.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session III

PB3164 Baby Detect, the first European genomic newborn screening initiative: project spreading and analytical validation.

Authors:

K. Hovhannesyan[1], L. Helou[1], F. Piazzon[1], T. Dangouloff[1], D. Mashhadizadeh[1], L. Palmeira[2], F. Boemer[2], L. Servais[3]; [1]Univ. of Liege, Liege, Belgium, [2]CHU Liege, Liege, Belgium, [3]Univ. of Oxford, Oxford, United Kingdom

The rapid development of new therapies for severe rare genetic conditions underlines the need to incorporate first-tier genetic testing into newborn screening (NBS) programs for treatable disorders that cannot be screened by biochemical assays. Next-generation sequencing (NGS) has an advantage over first-tier biochemical screening for diseases without accurate biomarkers.
For two years we have developed a workflow to screen 126 severe pediatric treatable disorders among newborns in Baby Detect pilot project in Liege area (www.babydetect.com). As of June 07th, 2022 - 1879 parents were proposed to participate in Baby Detect and 1694 agreed (90,1% acceptance rate).
The validated workflow includes extraction of DNA from the dried-blood spots (DBS), a targeted NGS (tNGS) by Twist Bioscience technology, Illumina sequencing, in-house developed pipeline for secondary analysis, variant interpretation using artificial intelligence-based Alissa Interpreter of Agilent. The methodology of the workflow was validated for manual operation and for the pilot screening considering the criteria acceptable for diagnostic settings.
The isolation of DNA from DBS was done by QIAamp DNA Investigator Kit (Qiagen) with sufficient yield and quality to perform tNGS. We have developed a custom-designed targeted-panel of 361 genes to perform deep sequencing of regions of interest. Twist Bioscience library preparation was optimized to be well suited for DBS-extracted DNA samples. Sequencing was performed using Illumina technologies on NovaSeq 6000 and NextSeq 550 systems. Raw sequencing data was analyzed by an in-house developed bioinformatic pipeline. Different performance parameters were calculated, and the quality criteria were established.
In the first year of the project, we are reporting disease-causative pathogenic and likely pathogenic SNV variants. For variant interpretation we developed a decision tree for smart filtering of variants using Alissa Interpret, which was validated on large number of positive and negative samples (600 samples). Validation of Baby Detect workflow allows demonstrating unbiased information on the technical feasibility and potential clinical utility of NGS-based NBS.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session I

PB3165 Beyond genomics: Using RNA-seq in filter cards to unlock the clinical relevance of non-coding variation in splicing.

Authors:

P. Bauer[1,2], R. Al-ali[2], M. Radefeldt[2], L. M. Pardo[3], J. Pinto Basto[3], A. Bertoli-Avella[2]; [1]CENTOGENE, Börgerende-Rethwisch, Germany, [2]Centogene GmbH, Rostock, Germany, [3]CENTOGENE, Rostock, Germany

Clinical exome and genome sequencing are first line diagnostic methods in rare diseases with diagnostic yields going up to 60%. Still variants of unknown significance (VUS) account for up to 5% of reported variations in ClinVar database. Here we assessed the impact of splicing variants reported in our laboratory using RNA Sequencing experiments (RNA-seq) to gain insight into their clinical relevance.**Methods** A total of 108 consecutive patients with 113 splicing variants reported after exome/genome sequencing were selected for RNA sequencing (RNAseq). A protocol was developed to perform RNAseq using filter card (dried blood spots - DBS), library preparation and bioinformatic pipeline analysis. Relative gene expression was calculated using house-keeping genes and compared against controls. Splicing patterns in cases were inspected using IGV interactive program and adding three independent controls. **Results and Discussion** Forty-nine variants were located at canonical intronic positions 1-2 (43%), 64 (57%) were affecting other intronic regions (up to 824 bps). Eighty-one variants were heterozygous (72%), 26 were homozygous (23%) and six were hemizygous variants (5%). Of the 113 variants, 65 (58%) were confirmed as leading to abnormal splicing (with a minimum of 5 supporting reads), 15 (13%) did not have any evident splicing effect and 33 (29%) were inconclusive. The main reason for the inconclusiveness was insufficient coverage of the area. **Conclusions** We propose a method for a systematic experimental evaluation of the splicing impact of intronic variants, integrated in diagnostic exome/genome sequencing, which impact the assessment of their clinical relevance. The approach can be implemented in the routine workflow by diagnostic laboratories, adding an additional -omics layer to the diagnosis of rare disorders.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session II

PB3166 † Beyond the exome: Finding answers for clinical exome-negative patients

Authors:

S. Meyn[1], A. Hall[2], H. Motiff[1], J. Heilmann[1], L. Frater-Rubsam[3], H. Nguyen[4], X. Shao[4], V. Horner[4], D. Pavelec[5], B. Webb[2]; [1]Ctr. for Human Genomics and Precision Med., Univ. of Wisconsin Sch. of Med. and Publ. Hlth., Madison, WI, [2]Dept. of Pediatrics, Univ. of Wisconsin Sch. of Med. and Publ. Hlth., Madison, WI, [3]State Lab. of Hygiene, Univ. of Wisconsin Sch. of Med., Madison, WI, [4]State Lab. of Hygiene, Univ. of Wisconsin - Madison, Madison, WI, [5]Biotechnology Ctr., Univ. of Wisconsin - Madison, Madison, WI

Background: Just over half of 9,000+ rare genetic disorders have a known cause and most patients who undergo clinical WES fail to obtain a diagnosis. To address these issues we created the University of Wisconsin Undiagnosed Disease Program (UW UDP), which takes a "beyond the exome" approach to evaluating genetics patients.
Objectives: 1) discover new disease genes; 2) improve our understanding of genetic disorders; 3) provide patients with actionable diagnoses; and 4) evaluate novel technologies. We select patients, in part, based on the complexity/rarity of phenotype and the number of affected family members. Our workflow begins with clinical WES reanalysis, followed by trio short read genome sequencing. Long read sequencing, RNA-Seq, optical mapping, and epigenomic profiling are utilized *ad hoc*.
Results: To date, the UW UDP has enrolled 59 families, including 67 affected individuals and 110 unaffected relatives. ~90% of probands had prior clinical WES. We identified candidate causal variants for 8 of the first 14 patients. Short and long read WGS, WES reanalysis, and RNA-Seq each played a role in finding i) a 15q11.2 deletion and an instance of chromoplexy involving chromosomes 2, 3, 8, and 21 missed by clinical testing; ii) two new candidate disease genes; and iii) a patient whose novel phenotype is the likely result of synergy between pathogenic variants in genes causing a glycosylation disorder and grey platelet syndrome. A clinical diagnosis of Aicardi Goutieres syndrome was ruled out by RNA-Seq. Additional analyses are on-going.
Conclusion: Our initial results suggest that a significant fraction of clinical WES-negative patients can be diagnosed using combinations of WGS, long-read WGS, optical mapping, and RNA-Seq, indicating that diagnostic rates of >65% are attainable. An undiagnosed genetic disease program can serve as an important component of a comprehensive center for rare diseases, as it offers patients access to emerging technologies, increases diagnostic rates, and facilitates the discovery of new disease genes while advancing our understanding of rare genetic disorders.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session III

PB3167 Beyond the exomeUtility of long-read whole genome sequencing in exome-negative autosomal recessive diseases

Authors:

F. Alkuraya; King Faisal Specialist Hosp. & Res. Ctr., Riyadh, Saudi Arabia

At least 50% of patients with suspected Mendelian diseases remain negative after exome sequencing. Several factors have been identified at the technical and interpretation levels to explain this diagnostic gap. Long-read whole genome sequencing (lrWGS) has the potential to address the technical limitations of exome sequencing e.g. structural variants (SV). However, its utility in autosomal recessive Mendelian diseases is largely unknown. In a selected cohort of 34 families in which the suspected autosomal recessive diseases remained undiagnosed despite careful consideration of interpretation challenges of their exomes, lrWGS was performed at an average depth of 10x on the Pacific Bioscience Sequel IIe platform. Likely causal variants were identified in 14 (41%) of the cohort. These include: 1) a homozygous splicing SV in *TYMS* as a novel candidate gene for lethal neonatal lactic acidosis, 2) a homozygous SV that we propose impacts *ANO7* and *STK25* and causes a unique combination of albinism and microcephaly, 3) a compound heterozygous SV in *RP1L1* with complex inheritance pattern in a family with inherited retinal disease, 4) a homozygous UTR SNV in *MZT1* and deep intronic variants in *LEMD2* and *SNAP91* as novel candidate genes for neurodevelopmental disorders in three families, and 5) a promoter SNV in *SLC4A4* causing non-syndromic band keratopathy. Surprisingly, we also encountered causal variants that could have been identified by short read exome sequencing in 7 families. The latter highlight scenarios that are especially challenging at the interpretation level including complex compound heterozygosity, high allele frequency, intrafamilial genetic heterogeneity, isoform confusion, and phenotyping challenges. Our data highlight the continued need to address the interpretation challenge in parallel with our effort to improve the sequencing technology itself. We propose a path forward for the implementation of lrWGS sequencing in the setting of autosomal recessive diseases in a way that maximizes its utility.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session I

PB3168 Bringing constitutional genome sequencing to a children's hospital.

Authors:

J. Hunter, A. Otero Bravo, B. Kelly, B. P. Chaudhari, D. Corsmeier, J. Garee, V. Jayaraman, H. Jenkins, H. Kuck, G. Lammi, M. Leung, A. McKinney, S. Ravagnani, J. Scholl, E. Stonerock, P. White, C. Cottrell; Nationwide Children's Hosp., Columbus, OH

Genome sequencing (GS) is making its way into mainstream clinical genetic testing paradigms and provides a lifelong resource for medical care. Many aspects of our health are rooted in our genetic sequence, especially in infants and children who have had less environmental exposures that modify their health. We describe the reasoning for, and process of, validating and implementing clinical GS and rapid genome sequencing (RGS) at a U.S. children’s hospital.GS provides important advantages over other testing modalities. GS dramatically reduces hands on time in the lab, requires fewer steps at which errors can occur, and uses fewer PCR cycles resulting in reduced artifacts. GS yields cleaner, more even sequencing coverage, and comprehensive variant calling. GS has the highest diagnostic yield of any clinical genetic test available, providing the best chances of successful diagnosis for critically ill infants and children. While GS requires more sequencing and thus more computing and storage, these obstacles are mitigated in many ways by reduced lab time, decreasing sequencing time and costs, and decreased computing and storage time and cost.To validate GS and RGS as clinical genetic tests, we used 185 GS data sets primarily from a translational research protocol. We validated saliva, peripheral blood, and cord blood as acceptable sample types. Variant concordance was determined for over 200 variants from more than 100 cases reported through clinical ES or GS sequencing by internal or external clinical sequencing labs. We determined the limit of detection for GS at multiple sequencing depths and performed downsampling experiments to establish accurate variant calling saturation levels.We validated RGS as a true STAT test with a verbal report turn-around-time (TAT) of 3-5 days with a final report in 5-8 days. To meet this aggressive TAT, we optimized many processes including wetlab procedures, bioinformatics pipelines, and use phenotype-driven variant prioritization. Furthermore, we worked to optimize and expedite hospital logistics prior to sample receipt.Future work includes evaluation of additional bioinformatics pipelines including DRAGEN and additional data analysis of mitochondrial DNA, structural variants, as well as infectious disease identification.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session II

PB3169 CardioSeq: A prospective study investigating the impact of clinical genome sequencing in a diverse patient population with cardiovascular disease

Authors:

D. Lanfear[1], L. Amendola[2], K. Robinson[2], T. Ebai[2], T. Kalista[2], W. Cabral[1], K. Collins[1], D. Hostin[2], A. Huertas-Vazquez[2], M. Longoni[2], R. Taft[2]; [1]Henry Ford Hosp., Detroit, MI, [2]Illumina Inc., San Diego, CA

Introduction: Cardiovascular disease (CVD) is responsible for 1:5 deaths in the United States and is one of the leading causes of mortality globally. Targeted genetic testing is recommended for multiple CVD indications; however, society guidelines are inconsistently followed. The role of comprehensive genomic testing in cardiology has not been fully investigated.
Methods: The CardioSeq study is a prospective, open label, single-center investigation conducted at Henry Ford Health cardiology clinics, testing the diagnostic efficacy and clinical utility of clinical genome sequencing (cGS) that includes a 200-gene CVD panel, 4 CVD risk alleles, 10 PGx genes, 35 non-CVD ACMG secondary finding genes, and a polygenic risk score for coronary artery disease (CAD). The planned enrollment is 1500 adult participants. Enrollment criteria include a diagnosis of aortopathy, arrhythmia, cardiomyopathy or heart failure, coronary or peripheral artery disease, or dyslipidemia, without a prior molecular genetic diagnosis. The primary outcome is the proportion of patients who receive a new molecular diagnosis. Secondary outcomes include the proportion of participants with a finding leading to a change in management (CoM) compared to an EHR-derived contemporaneous matched (2:1) usual care group.
Results: To date, 158 participants have been enrolled in CardioSeq, among whom self-reported race is 49% Black or African American, 48% White, 1% Middle Eastern, 1% Asian. Self-reported ethnicity is 1% Hispanic or Latino. The majority of participants (105/158, 66%) have at least two concurrent CVD diagnoses, with the most common being dyslipidemia (88%) and arrhythmia (81%). Enrollment is anticipated to complete by early 2024.
Conclusion: The integration of a comprehensive cGS test into the care of diverse cardiology patients increases access to genetic information and will generate evidence on the clinical utility of the genome as a platform for enhancing CVD care.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session III

PB3170 Case report: A 3 year old female with skin hamartoma, facial masses, trismus and mosaic chromosome 22 trisomy.

Authors:

J. Fu[1], P. D'Souza[1], E. Macnamara[1], M. Sabii[1], N. Hanchard[2], T. Toluwani[3], E. Banfield[2], L. Castelo-Soccio[4], H. Kim[5], B. Pusey[1], NIH Undiagnosed Diseases Network, D. Adams[1]; [1]Undiagnosed Diseases Program, Intramural Res. Program, NIH, Bethesda, MD, [2]Childhood Complex Disease Genomics Section, Ctr. for Precision Hlth.Res., Natl. Human Genome Res. Inst., NIH, Bethesda, MD, [3]UCSF Sch. of Med., San Francisco, CA, [4]Natl. Inst. of Arthritis and Musculoskeletal and Skin Diseases, NIH, Bethesda, MD, [5]Natl. Inst. on Deafness and Communication Disorders, NIH, Bethesda, MD

Introduction: The NIH Undiagnosed Diseases Program (UDP) evaluates participants with significant illnesses that remain undiagnosed after standard of care clinical evaluation. We present a case of mosaic trisomy 22 with progressive skin lesions, deforming facial masses and trismus.
Methods: Study participants are enrolled in the UDP using an IRB approved protocol.
Results: A 33 m/o female presented with a history of congenital, evolving skin lesions involving the scalp, face and upper thigh. Lesions were non-painful nodules on an erythematous base, some with sebaceous-appearing components. The largest was a right-sided 30 mm x 20 mm periauricular lesion. Congenital right-sided facial hemi-hypertrophy primarily involved the right supra-zygomatic and pre-auricular areas. Growth and development were normal. A face ultrasound at 3 m/o revealed a 17 mm x 14 mm ill-defined hypoechoic complex mass adjacent to the right parotid gland. At NIH evaluation, a new diagnosis of moderate trismus was made. Facial MRI revealed T2-bright fat-non-enhancing infiltration-like signal involving the of masseter, pterygoid, parotid gland, and facial adipose tissue. The TMJ appeared intact, suggesting trismus secondary to infiltrated muscles. Biopsy of superficial neck lesions revealed an abnormal mixture of tissue types (vessels, nerves, skeletal and smooth muscle, eccrine glands and ducts) normally found in the skin. Some vessels showed a perivascular lymphocytic infiltrate, prominent in the deeper reticular dermis. The findings were consistent with a cutaneous hamartoma. Clinical genome analysis did not reveal diagnostic variants; research analysis did not yield research-candidates. A high-density Illumina SNP array performed on blood DNA showed a B-allele signal consistent with mosaic trisomy 22.
Conclusions: Previous reports have described trisomy 22 mosaicism in association with dysmorphic features not found in our case. However, hemihyperplasia has been described, leading us to hypothesize that our participant is affected by more restricted distribution of mosaicism than described in prior reports. We present imaging and further laboratory evaluation in a case report detailing this rare presentation and potential diagnosis.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session I

PB3171 Challenge to elucidate complex structural variants using long-read sequencing

Authors:

Y. Enomoto, Y. Kuroda, Y. Saito, K. Kurosawa; Kanagawa Children's Med. Ctr., Yokohama, Japan

Complex structural variants involving multiple chromosomes are sometimes detected in patients with congenital anomalies. G-banding is one of the methods for analyzing complex structural variants, however the resolution is low and it is difficult to elucidate the details of variants. Long-read sequencing is a useful method for analysis of complex structural variants. In this study, we try to elucidate complex structural variants in a Japanese girl patient with developmental delay. She was the second child of healthy parents. She was born at 38 weeks gestation with a birth weight of 2296 g (-1.53 SD) and length of 44.3 cm (-1.79 SD), and head circumference 32 cm (-0.74 SD). Apgar scores were 9 at 1 min and 10 at 5 min. She was referred to the hospital for suspected multiple congenital anomalies syndrome at the age of two months. She had developmental delay and facial features including arched eyebrows, ptosis and a triangular nose. At the age of 3 years, her weight of 10.85 kg (-1.4 SD) and length of 84.2 cm (-2.4 SD). She could speak a single word. G-banding detected complex structural variants: der(5)(5pter→5q13.1::?), der(6)(5qter→5q15::6p23→6qter), der(9)(?::9p24→9q22.1::?), der(10)(9qter→9q22.3::?::10p11.2→10qter), der(18)(18pter→18q22.1::?). Long-read sequencing revealed four translocation breakpoints out of the seven breakpoints detected by standard karyotyping. Furthermore, long-read sequencing identified 1.28Mb inversion at the breakpoint of t(5;6) and uncovered the additional fragments of der(9) and der(18). These results indicated that the long-read sequencing can identify complex rearrangements of cytogenetic translocations at high resolution and reveal the underlying mechanisms of the structural variations. A detailed study of phenotypic associations with genes adjacent to breakpoints created by translocations remains a future challenge.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session II

PB3172 Challenges and breakthroughs in unified digital laboratory systems in diagnostics and research core laboratories.

Authors:

B. Dunn; L7 Informatics, Austin, TX

Recent advances in scientific procedures and best practices have been matched with a proliferation of specialized software solutions. This leads to most diagnostics and research labs using multiple software systems that each only track, manage, or process a subset of the lab's overall operations - thus effectively siloing each team member, group, or process from the rest of the lab. These silos then result in decreased data flow through the lab, decreased visibility into sample and data status, and decreased sample processing throughput. Here we demonstrate breakthroughs in digitizing laboratory processes in fully compliant systems from the point of sample collection through to data reporting. This leads to an increase in laboratory efficiency, increased data integrity, better decision making due to better data visibility, and reduced costs for diagnostics laboratories and research cores. As a proof of principle we highlight how two different academic research cores have implemented this new technology, L7|ESP, to reduce their laboratory silos. Laboratory 1 is a sequencing core at an academic institution and processes hundreds of samples per day. Utilization of L7|ESP provided them with the ability to send information and samples across various groups within their core and seamlessly output final reports to their customers. Laboratory 2 is a cutting-edge protein sciences group at a large nonprofit biomedical research institution. Their work utilizes multiple different instruments, sample types, and desired outputs and thus requires flexibility and adaptability without sacrificing throughput, data integrity, or user experience. Through modeling these processes in L7|ESP they are able to run specialized processes at scale while still running generalized processes to meet the changing needs of their clients. These case studies show that modern diagnostics and research core laboratories can leverage today’s latest breakthroughs in computer science and software to enable them to successfully scale and grow to tomorrow’s scientific needs.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session III

PB3173 Chronic mucocutaneous candidiasis due to*Candida auris*in patients with a mild*TP63*-associated ectodermal dysplasia

Authors:

H. Vahidnezhad[1], F. Palizban[2], A. Saeidian[3], L. Youssefian[4], H. Hakonarson[5], S. Khodavaisy[6]; [1]Univ. of Pennsylvania, Philadelphia, PA, [2]IBB, Tehran, Iran, Islamic Republic of, [3]Children s Hosp. of Philadelphia, Philadelphia, PA, [4]UCLA, Los Angeles, CA, [5]Children's Hosp. of Philadelphia, Philadelphia, PA, [6]Tehran Unuiversity of Med. Sci., Tehran, Iran, Islamic Republic of

Persistent fungal infections are associated with host immune status and the type of fungal species. Genetic variants that lead to immunological susceptibility to fungi have long been sought and recognized. Understanding these genetic underpinnings will be a powerful tool in the age of personalized medicine. In this study, unbiased, whole-transcriptome sequencing by RNA sequencing (RNA-Seq) was utilized as a first-tier strategy for the concomitant detection of the genetic determinant of abnormal host immunity and the causative fungal agent in a father and daughter of Iranian descent who presented with chronic mucocutaneous candidiasis (CMC), mild ectodermal dysplasia, and recurrent bacterial infections. Additionally, the patients' humoral and cellular immunity were evaluated, and the patients' peripheral blood DNA were analyzed with whole-exome sequencing (WES). Skin, hair, and nail biopsies were independently evaluated by conventional mycological tests and fungal DNA sequencing. WES and RNA-Seq disclosed a monoallelic mutation in *TP63* that matched with the clinical phenotypes of our patients. In addition, RNA-Seq confirmed the pathogenicity of detected sequence variants and expression profiling. Additionally, non-albicans *Candida* species, *Candida auris,* were detected in CMC lesions of both patients and *Candida parapsilosis* was isolated in the father's samples. Both patients had normal humoral and cellular immunity. Although a strong association between ED and *TP63* mutations exists, the mechanism of immune dysregulation remains unclear. This study represents the first case of CMC caused by the multidrug-resistant emerging pathogen *C. auris* in the context of *TP63*-associated ED and demonstrates the RNA-Seq utility for concomitant host mutation detection and accurate identification of rare and often misidentified pathogens.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session I

PB3174 Chronic recurrent pancreatitis: importance of genetic testing.

Authors:

S. Carlo[1,2], N. Arciniegas[2,1], L. Rodriguez[1], F. Benitez[1], A. Cornier[3], E. Albino[4]; [1]Ponce Hlth.Sci. Univ., Ponce, PR, [2]Mayaguez Med. Ctr., Mayaguez, PR, [3]San Jorge Children s & Women s Hosp., San Juan, PR, [4]Univ. of Puerto Rico - MSC, San Juan, PR

Pancreatitis occurs when there is inflammation in the pancreas because of autodigestion by enzymes secreted from the exocrine portion of this organ. Clinical diagnosis of the disease can be differentiated into the following types based on duration: acute (few days to six months), acute recurrent (more than one episode of acute pancreatitis), and chronic (more than six months). An earlier onset of pancreatitis tends to correlate with phenotypic presentations of mutations on pancreatic associated genes. Those that have been previously associated with pancreatitis include the *SPINK1*, *CTRC*, *PRSS1* and *CFTR* genes, which will be further discussed in this case report. We introduce the case of a 6-year-old Puerto Rican female who has suffered from seven episodes of acute pancreatitis in the past year. A diagnosis of acute recurrent pancreatitis with unknown etiology was initially made. As part of the diagnostic criteria for patients who present with unexplained episodes of pancreatitis, tests were performed in order to verify possible genetic etiology. Testing revealed that the patient was heterozygous for two variants that predispose to pancreatitis found on genes *CTRC* and on *SPINK1* . Both of these genes are essential in the control of trypsin, a proteolytic enzyme that regulates the exocrine enzymatic cascade in the pancreas.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session II

PB3175 Clinical and Genetic Features of 22 Neonatal Dubin-Johnson Syndrome in Japan

Authors:

T. Togawa[1], S. Ito[1], K. Imagawa[2], H. Hayashi[3], K. Ito[1], T. Endo[1], T. Sugiura[1], S. Saitoh[4]; [1]Nagoya City Univ., Nagoya, Japan, [2]Univ. of Tsukuba, Tsukuba, Japan, [3]The Univ. of Tokyo, Tokyo, Japan, [4]Nagoya City Univ, Nagoya, Japan

Dubin-Johnson syndrome (DJS, [MIM 237500]) phenotype results from biallelic pathogenic variants in *ABCC2* which encodes a member of the large ABC transporter family. DJS was rarely diagnosed in the neonatal period and thus few case reports with molecular genetically confirmed neonatal DJS have been reported. Our previous study reported 10 patients with neonatal DJS determined by molecular genetic diagnosis using targeted next-generation sequencing (Togawa T, et al. *J Pediatr 2018*). After the study, we further identified 12 more patients with neonatal DJS. Here, we aimed at clarifying clinical and genetic features in 22 East-Asian patients with neonatal DJS. We recruited 22 patients with neonatal DJS from pediatric centers in Japan between September 2013 and February 2021. We defined neonatal DJS as follows: onsets of clinical manifestations of cholestasis, such as prolonged jaundice and acholic stools, during the neonatal period; increased serum direct bilirubin (D Bil) value; homozygous or compound heterozygous pathogenic variants in *ABCC2*. Clinical course, laboratory findings, and variants concerning *ABCC2* were retrospectively and prospectively examined. Regarding genetic analysis, we previously developed targeted next-generation sequencing and bioinformatics methods. The median duration of gestation and birth weight was 39 weeks and 2956 g, respectively. All 22 patients presented persistent jaundice starting from the neonatal period. The median values of the highest serum total bilirubin (T Bil) and D Bil were 10.5 mg/dL and 5.9 mg/dL, respectively. Ten of the 22 patients who received hepatobiliary scintigraphy examinations showed abnormal results. For all patients, cholestasis which was determined by serum D Bil value improved rapidly and serum D Bil value decreased to <1.5mg/dL within 12 months. Regarding causative pathogenic variants, all patients harbored compound heterozygous or homozygous pathogenic variants in *ABCC2*: Six variants (R768W, c.1967+2T>C, R100Ter, c.2125T>C, c.2439+2T>C, and c.2882A>G) occurred in multiple alleles accounting for 9, 7, 4, 4, 4 and 3 of 44 affected alleles, respectively. We identified 5 novel pathogenic variants, A364NfsTer37, I719T, L984R, c.3614+1G>A, and D1361N. The clinical courses of the 22 patients with neonatal DJS were similar. Severe cholestasis developed from the neonatal period and disappeared rapidly during the infantile period. This clinical course is unique to neonatal DJS. Several hot spots of pathogenic variants in *ABCC2* might exist in patients with neonatal DJS among the East Asian population. Molecular genetic testing can uncover patients with neonatal DJS.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session III

PB3176 Clinical and genetic overlap between CHH and CLP

Authors:

I. Habibi[1,2], F. De-Azevedo-Correa[1,2], J. Zhai[1,2], A. Boizot[1], M. Adamo[1,2], K. Chachlaki[1], P. Luca[3], M. Rauch[4], V. Popovic[5], R. Quinton[6], M. Clemente Leon[7], A. Nemeth[8], B. Braumann9[9], g. papadakis[1], Y. zouaghi[1,2], N. Niederlander[1,2], N. Pitteloud[1,2]; [1]Lausanne Univ. Hosp., Service of Endocrinology, Diabetology and Metabolism, Lausanne, Switzerland, [2]Univ. of Lausanne, Lausanne, Switzerland, [3]Dept. Endocrinologia e Malattie del Metabolismo, IRCCS Istituto Auxologico Italiano, Milano, Italy, [4]Inst. of Med. Genetics, Univ. of Zurich, Zurich, Switzerland, [5]Inst. of Endocrinology Dept. of Neuroendocrinology, Belgrade, Serbia, [6]Dept of Endocrinology, Newcastle-upon-Tyne, United Kingdom, [7]Unidad de Endocrinología Pediátrica, Barcelona, Spain, [8]St. John's Hosp., Budapest, Hungary, [9]Dept. of Orthodontics, Univ. of Cologne, cologne, Germany

Purpose. Congenital hypogonadotropic hypogonadism (CHH) and Cleft lip palate (CL/P) are two complex genetic conditions. Here we report an overlapping syndrome. **Methods**. First, we identified patients with CHH and CL/P through international collaboration. Genetic and clinical analysis of 41 patients with CL/P and reproductive defects were completed. Second, we focused on 442 probands with CHH without CL/P and studied the implication of CL/P genes in this population. **Results**. We found pathogenic (P) and likely pathogenic (LP) variants in 11 patients (27%) with CL/P & CHH in CL/P genes and/or CHH genes (*FGFR1, SOX9, CHD7, KLB+PROKR2, CTNND1*, and *SOX10*). Given the broad roles of SOX9 during development, it is a putative candidate gene for CL/P. Interestingly, *SOX9* has been associated in one patient with CHH. As expected, the most frequent CL/P gene involved in the CHH cohort without CL/P was *FGFR1* (31 P and LP variants, 60%), a gene known to be involved in CHH. The second most prevalent gene was *CHD7*, with 9 CHH patients exhibiting mutations in this gene. Other CL/P genes appear in our CHH cohort including two variants in *DVL3* and one P or LP variant in *PLCB4, PIEZO2, TP63, TGFBR2, TCOF1, NIPBL, CHD1, KMT2D, INTS1* and *COL2A1*. We were surprised to identify several P and LP variants in CL/P genes among our CHH cohort. **Conclusion.** Taken together, this extensive clinical and genetic study revealed heterogeneity in the genetic landscape of patients exhibiting both CL/P and central reproductive defects.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session I

PB3177 Clinical whole genome sequencing after non-diagnostic genetic testing identifies distinct types of genetic variants

Authors:

S. Berger[1], E. Delot[2], G. Pitsava[1], A. Cohen[3], J. Lotempio[4], S. Marmolejos[3], A. Kahn-Kirby[4], S. Knoblach[1], A. Ko[5], R. Nussbaum[6], V. Fusaro[4], E. Vilain[4]; [1]Childrens Natl. Res. Inst., Washington, DC, [2]Children s Natl. Res. and Innovation Campus, Washington, DC, [3]Children's Natl. Res. Inst., Washington, DC, [4]Univ. of California - Irvine, Irvine, CA, [5]Children's Natl. Res. Inst., Washingtion,DC, DC, [6]Invitae, San Francisco, CA

Background: Exome sequencing has become the genetic test of choice for a wide range of clinical presentations but has limitations. Genome sequencing is a possible alternative as a first-tier test in the clinical environment as it promises to overcome many of these limitations. We analyze here the improvement of the diagnostic yield and the types of variants whose detection is facilitated by genome sequencing.Methods: Participants without diagnosis from clinical testing including CLIA exome sequencing were enrolled in the Pediatric Mendelian Genomics Research Center (PMGRC), a Genomics Research to Elucidate the Genetics of Rare disease (GREGoR) Site. Genome sequencing and analysis were performed utilizing clinical pipelines. Diagnostic results were categorized as "coding region", "near coding intronic", "deep intronic", and "structural variant". For variants affecting exons or within 20 bases of an exon, original clinical testing was reassessed to understand the reasons why variants were not initially reported.Results: For 122 probands enrolled in the PMGRC, 31 (25%) had variants considered likely diagnostic by whole genome sequencing. Diagnostic variants not detectable by exome included deep intronic or exon-copy neutral structural variants (e.g. deletion of promoter into 5'UTR or rearrangement and inversion of exons). RNA sequence analysis provided supporting evidence for pathogenicity for several deep intronic variants. Diagnostic variants affecting exonic regions were most often for genes that lacked published clinical validity at the time of the original exome but had since been published as associated with a novel syndrome. Additionally, several known pathogenic exonic variants were identified that had not been reported on the original exome due to clinical pipeline exclusion of regions with frequent artifactual calls.Conclusions: While some increased diagnostic yield from genome resulted from being an interval reanalysis after exome, diagnostic yield was also improved through detection of non-exonic variants, structural variants, and technically challenging exonic regions.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session II

PB3178 Clinically Significant Copy Number Variants (CNVs) are associated with Common Fragile Sites at their breakpoints - A novel approach in the study of CNV biogenesis

Authors:

F. Islam, S. Chandrashekar, K. Awayda, M. Iqbal; Univ Rochester Med Ctr, West Henrietta, NY

Intro: Copy number variants (CNVs) occur when parts of the genome are duplicated or deleted. CNVs are commonly caused by non-allelic homologous recombination (NAHR) but other mechanisms may exist. Common fragile sites (CFSs) are regions of the genome susceptible to breakage through chromosome condensation, leading to oncogene-induced DNA damage. CNV induction by CFSs remains unknown. In this study, we retrospectively reviewed clinically significant CNVs reported as per the ACMG guidelines in our microarray CGH laboratory database. We investigated the association of CNVs breakpoints overlap within 200+ CFSs. This is a novel approach being undertaken to understand CNV biogenesis.
Methods: We recruited and anonymized 523 CNVs from our DNA microarray CGH laboratory database, including pathogenic (P) (n=89), clinically uncertain-likely pathogenic (CULP) (n=51), and clinically uncertain (CU) (n=383) CNVs. ACMG guidelines and lab criteria were used to select CNVs; P, CULP, CU: >250 kb with >1 morbid gene. A list of over 200 CFSs was compiled from published studies into a BED file and uploaded into cytogenomic analysis software (4x180 CGH+SNP Microarray Kit, Agilent). A Python pipeline was developed to map CNV breakpoints to CFS coordinates, and subsequent analyses were undertaken to further examine the nature of this association.
Results: 37.7% (n=197) of analyzed CNVs overlapped CFSs. By category, these were CU (33.4%, n=128), P (44.9%, n=40), and CULP (56.9%, n=29). Most CNVs (P, CULP, U) analyzed ranged from 100Kb to 1Mb. No significant size distribution difference was observed between CNV groups positive and negative for CFS overlap, as CNVs of all sizes overlapped with CFSs. A high number of overlaps were found to occur on 15q, with the CFSs FRA15Bi and FRA15C having the most overrepresentation in the data set (17 and 10 events, respectively). Ongoing studies are being conducted to determine the 5' and 3' breakpoint regions of CNV/CSF junctions and will be presented.
Conclusion: In breakpoints of a substantial population of clinically relevant CNVs, nearly 40%, overlap with CFSs. Over 100 different CFSs located amongst almost every chromosome were found to overlap CNVs, but clusters of implicated CFSs exist. We hypothesize that the notable overlap between CNVs and CFSs could be the result of DNA instability which exists in CFSs during replication, which proffers a novel mechanism for CNV biogenesis. Furthermore, we speculate there may be more CFSs in the genome which have yet to be discovered and may be associated with the CNVs which are found not to associate in the current study. Future studies will seek to better define the association of these two genetic elements.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session III

PB3179 CNVviewer: A web-based interactive visualization and annotation tool for accurate clinical diagnosis and reporting of copy number variations.

Authors:

J. Zhou, K. Jiang, C. Li, C. Peng, J. Zhang, F. Chang, Q. Zhang, X. Cai, Y. Wang, P. Fang, S. Tang, W. Mu; KingMed Diagnostics, ShangHai, China

DNA copy number variations (CNVs) play a pivotal role in the development of inherited and acquired diseases. While optimized bioinformatics algorithms for CNV detection from next-generation sequencing (NGS) data are essential for accurate clinical diagnosis, it is equally important to have an interactive visualization tool with user-friendly interface that enables efficient assessment of technical validity and clinical significance of CNV calls.
To address this need, we present CNVviewer, a web-based interactive analysis tool designed for accurate clinical analysis and reporting of CNVs for both germline and somatic applications. CNVviewer accepts CNV results from widely used read depth-based calling tools such as CNVkit and CNVnator, with a particular optimization for whole-genome sequencing (WGS) data to enhance loading and response time through down-sampling. Moreover, CNVviewer allows for the importation of results from split read/read pair-based methods for the same sample, enabling additional supporting evidence for more precise breakpoints delineation. The tool provides both a global view and a local view, displaying copy ratio signals, B-allele frequency (BAF), cytobands, and gene annotations across the whole genome and for individual CNVs. In trio analysis and paired tumor-normal analysis, CNVs from multiple samples can be displayed together, aiding in the determination of CNV origin. In addition, CNVviewer performs key clinical gene annotations such as OMIM inheritance (MOI) and haploinsufficiency/triplosensitivity scores, presenting valuable clinical insights in a tabular view. The tool includes sorting capability for CNVs based on specific annotations (e.g. size, MOI) and dynamic link-outs to important public CNVs databases, further simplifying the prioritization and interpretation of potentially significant findings.
Our molecular geneticists have successfully utilized CNVviewer in our routine clinical practice for over 15 months. Our results demonstrate that CNVviewer is a versatile and comprehensive tool that can facilitate the identification of various CNVs at different resolutions, including intragenic exon-level duplications/deletions, focal gene-level alterations, and chromosome arm-level changes. Furthermore, we were able to use CNVviewer to review and in silico confirmed CNVs as small as 1.8kb in a WGS trio case and a 2-exon deletion in a targeted sequencing panel. We prove CNVviewer's strong capabilities in improving clinical yield and enhancing operational efficiency through diverse case studies in a clinical diagnostic setting.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session I

PB3180 Comparative medical genetics to facilitate the interpretation of rare missense variation

Authors:

B. Haque[1,2], G. Guirguis[2], M. Curtis[1], H. Mohsin[2,3], S. Walker[4,5], M. Morrow[6], G. Costain[1,2,7,8]; [1]Genetics and Genome Biology, SickKids Res. Inst., Toronto, ON, Canada, [2]Dept. of Molecular Genetics, Univ. of Toronto, Toronto, ON, Canada, [3]NeuroSci. and Mental Hlth., SickKids Res. Inst., Toronto, ON, Canada, [4]Genomics England, London, United Kingdom, [5]The Ctr. for Applied Genomics, The Hosp. for Sick Children, Toronto, ON, Canada, [6]GeneDx, Gaithersburg, MD, [7]Div. of Clinical and Metabolic Genetics, The Hosp. for Sick Children, Toronto, ON, Canada, [8]Dept. of Paediatrics, Univ. of Toronto, Toronto, ON, Canada

Background: Interpreting rare missense variants is a major challenge in human genetics. Existing in silico prediction tools often consider evolutionary conservation of a residue. However, the correlation between disease associations of specific variants across species is not well characterized. We hypothesized that for variants causing Mendelian disease in non-human livestock and domestic species, the orthologous variant would be similarly functionally significant in humans.Methods: We extracted missense variants from the Online Mendelian Inheritance in Animals (OMIA) database that were classified as likely causal for Mendelian phenotypes in one of nine species (cat, cattle, chicken, dog, goat, horse, pig, rabbit, sheep). We mapped the genomic coordinates to the human reference genome (GRCh38) using UCSC's LiftOver tool, and manually inspected each genomic region in the UCSC human genome browser to confirm that the reference and alternate amino acid residues were conserved. To assess clinical classifications in humans, we cross-referenced these variants with publicly available genetic databases, and from a large private commercial genetic laboratory database of genomic testing results from >400,000 individuals with suspected rare Mendelian disorders and their family members. Results: Of the 339 variants from the OMIA database were mappable to the human genome and impacted conserved residues, 57 had been previously seen in humans and classified with respect to pathogenicity: 32 (56.1%) as pathogenic/likely pathogenic, 24 (42.1%) as variants of uncertain significance, and 1 (1.8%) as benign/likely benign. The odds ratio for these variants having a pathogenic or likely pathogenic classification in ClinVar was 6.45 (95% CI, 3.81-10.94, $p<0.0001$), compared to all other germline missense variants in these same genes. Overall, 338 of the 339 variants were very rare in humans (minor allele frequency <0.001 in gnomAD). Most variants (n=228; 67.3%) had REVEL scores ≥0.644 (supporting evidence for pathogenicity) and only 13.6% variants had REVEL scores ≤0.290 (supporting evidence for benignness). Conclusion: Cross-species comparisons could facilitate the interpretation of rare missense variation. These results provide further support for comparative medical genomics approaches that connect big genomic data initiatives in human medicine and veterinary contexts.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session II

PB3181 Comparison of Cytoscan Xon array with other existing molecular methods for detecting exonic copy number variants.

Authors:

Y. Kim[1], J-S. Lee[2], M. Kim[1], M-W. Seong[3]; [1]Seoul Natl. Univ. Hosp., Seoul, Korea, Republic of, [2]Jee-Soo Lee, Seoul, Korea, Republic of, [3]Seoul Natl. Univ Hosp, Seoul, Korea, Republic of

Background: Recently, exonic copy number variants (CNVs) could be detected through next-generation sequencing-based (NGS-based) methods and multiple ligation-dependent probe amplification (MLPA). However, there are no fully established NGS-based methods, and MLPA has a limited number of probes. Cytoscan Xon was developed to enable detection of exonic CNVs with improved resolution by adding probes for each exon. In this study, based on the exonic CNVs verified by previous molecular methods, we analyzed the results of Cytoscan Xon. Using the summarized log2ratio values from Cytoscan Xon, we investigated optimal cut-off values to determine copy number state.Methods: In total, 59 patients' blood samples were used for testing. A total of 45 clinically-relevant exonic CNVs identified in MLPA or NGS-based methods were compared by Cytoscan Xon results. The copy ratio for each exon was determined by following two methods: 1) Xon region copy number-state (Xon Region CNS) data provided by Chromosome Analysis Suite (ChAS) software, 2) Summarized Log2ratio data, assigned to each region. In case of the latter, receiver operating characteristic (ROC) curve analysis was performed to assess the optimal cut-off value of Summarized Log2ratio for CNVs detection using MedCalc software v.22.3. Results: Compared to the MLPA and NGS-based methods, Xon Region CNS in ChAS yielded concordant results in 41 out of 45 (91.1%) samples. The Summarized Log2ratio, based on cut-offs of >0.427 for gain and <-0.351 for loss, yielded concordance results in 42 (93.3%) samples. There were a total of 4 discrepancies, 2 of which were single-exon level CNVs. For single-exon level CNVs, two out of 21 cases showed discrepant results. The remaining two other discrepant results were multiple exonic CNVs, including two independent exonic CNVs. In the Xon Region CNS data, these were considered single exonic CNVs. Conclusion: The Cytoscan Xon results showed relatively good concordance with MLPA and NGS-based methods. In addition, we demonstrated that higher concordance rate can be achieved by adjusting the cut-off value of the Summarized Log2ratio. However, it should be noted that there are some challenges for Cytoscan Xon to be utilized in clinical diagnosis like accurate identification of single-exon level CNVs or multiple adjacent exonic CNVs.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session III

PB3182 Comparison of two DNA library preparation methods for mitochondrial DNA sequencing

Authors:

Y. Seo[1], **K. Ryu**[1], Y. Park[1], M. Kim[1], S. Cho[1], H. Lee[2], J-S. Lee[1], M. Kim[1], M-W. Seong[1]; [1]Seoul Natl. Univ. Hosp., Seoul, Korea, Republic of, [2]Chosun Univ. Hosp., Gwangju, Korea, Republic of

Background: Long-range PCR is widely used for preparation of mitochondrial DNA (mtDNA) libraries, which uses two pairs of mtDNA-specific primers, enabling Next-Generation Sequencing (NGS) with a small amount of mtDNA. We performed the probe capture enrichment for library preparation for whole exome and mtDNA and compared the NGS results with the conventional methods. **Methods**: DNA samples from thirteen patients with negative or positive results of long-range PCR based mtDNA sequencing performed by Illumina MiSeq and three patients with large deletion of mtDNA detected by multiplex ligation-dependent probe amplification (MLPA) were included. Library preparation was performed using a custom 1:100 mixture of Agilent SureSelect Human All Exon v8 and probes for mtDNA, and NGS was performed by Illumina Nova Seq 6000. Variant calling was done by Nextgene V2.4.0.1 and only single nucleotide variants (SNVs) were included, and variants from known artifact-prone sites, with a common low heteroplasmy flag sign, and with no passing filters were excluded, referring to gnomAD v.3.1.2 and Mitomap. Structural variants (SVs) were confirmed by Integrative Genomics Viewer (IGV). An allele frequency between 10% and 95% was considered heteroplasmy, while a frequency greater than 95% was considered homoplasmy. **Results:** In terms of SNV, total of 249 variants including 12 variants classified as variant of unknown significance, likely pathogenic, or pathogenic by the conventional method were detected, and 214 SNVs (86.0%) were detected concordantly, 20 (8.0%) were detected only in long-range PCR based NGS, and 15 (6.0%) were detected only in capture-based NGS. Among concordantly detected 214 SNVs, in respective of qualitative detection of homoplasmy/heteroplasmy, the agreement of the two methods was 99.1% (212/214). Conventional mtDNA sequencing identified 207 SNVs are homoplasmy, two of which were identified as heteroplasmy (allele frequency: 94.19% and 89.36%, respectively) in the capture-based mtDNA sequencing. In the Bland-Altman analysis, the mean bias was 0.20(%) and limit of agreement was [-10.0(%), 9.5(%)]. As for SV, capture-based NGS detected all known large deletion of mtDNA in three patients. The heteroplasmic levels detected through capture-based NGS were within the range of results determined by MLPA ratios in two of the three patients. **Conclusion**: Preparation of mtDNA libraries using the probe capture enrichment produces similar results with long-range PCR based NGS and more detailed information than MLPA. Capture-based mtDNA sequencing might be more beneficial because it has less risk for allele dropout and also enables simultaneous exome sequencing.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session I

PB3183 Comprehensive variant analysis in clinical samples by cost effective whole genome sequencing.

Authors:

S. Pollock[1], P. Winer[1], D. Bogumil[1], T. Biniashvili[1], D. Shem-Tov[1], I. Soifer[1], S. Hussey[2], G. Krautz-Peterson[2], A. Kaplun[2], O. Barad[1], D. Lipson[1]; [1]Ultima Genomics, Newark, CA, [2]Variantyx Inc., Framingham, MA

We present a comprehensive clinical variant analysis study using the low-cost Ultima Genomics UG100 platform on a meticulously selected cohort of samples aimed at representing common pathogenic clinical variants. The dataset included 61 samples with annotated pathogenic variants, comprising of 18 cell-lines with clinical annotation and 45 patient saliva samples. We generated WGS data at an average coverage of 50X from these samples, with alignment percentage in the range of 62.2-99.9 due to microbiome contamination in the saliva samples. The primary objective of our study was to assess the recall on this clinical variant dataset. We achieved a detection rate of 100% for 32 expected pathogenic single nucleotide variants (SNVs), which included 20 SNPs and 12 indels of lengths 1-6 in genes such as MME, SYNE1, GALNS, PKHD1, OPA1. Additionally, we identified an average of 4.2 rare (population AF less than 0.01) candidate ClinVar pathogenic variants per sample (using the Ensembl Variant Effect Predictor on GENCODE v43 CDSs), demonstrating a low rate of false positives. Leveraging an optimized version of cn.mops we successfully detected all 20 expected pathogenic copy number variants (CNVs), encompassing 16 CNVs of length 134Kbs to 22.68Mbs and 4 CNVs spanning entire chromosomes (Trisomy 21, XXYY syndrome, XYY syndrome, and 90% mosaicism for Monosomy X). We additionally examined 5 pathogenic structural variations (SVs), including 2 inversions and 3 Alu insertions, and 14 cases with pathogenic short tandem repeat expansions (STRs). All 5 expected SVs were identified using GRIDSS and STRs were evident in the data through manual inspection, though optimized tools for classification of STR expansions samples and optimization of the SV detection in UG100 data are still currently in active development. By enabling variant detection across the entire genome, WGS allows for a more comprehensive understanding of an individual's genetic profile, offering valuable information for personalized healthcare. Cost-effective sequencing further enables WGS accessibility in clinical settings, enabling healthcare providers to harness the power of genomic information for improved disease diagnosis, treatment selection, and monitoring.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session II

PB3184 D-2-hydroxyglutaric aciduria with enchondromatosis, severe vascular lesions and multiple cancers/growths,

Authors:

C-H. Tsai[1], c. hollas[2]; [1]Univ. of Illinois Chicago, Chicago, IL, [2]Univ. of oklahoma, Oklahoma city, OK

D-2-hydroxyglutaric aciduria (D2HA) with enchondromatosis is a disorder characterized by developmental delay, epilepsy, hypotonia, dysmorphic features, and multiple benign masses of hyaline cartilage in the medulla of metaphyseal bone. IDH1, IDH2 (mosaic) and D2HGDH gene changes are strongly associated with this condition. We report an 8-year-old male ascertained at 1 year of age with D2HA and enchondromatosis, unilateral cataract, hyper-hypo pigmented, segmental whorls and streaks on his abdomen which evolved into angiokeratoma circumscriptum. He was diagnosed with AML at age 3, relapsed at age 5 and received a hematopoietic stem cell transplantation. At 6 years of age, he was found to have more than 100 polyps from stomach to rectum. The skin lesions worsened over time and developed a tree bark like verrucous hemangioma. He is short (32 inch, 0%tile) with disproportionally large hands. WES on buccal smear revealed mosaic VUS in CREBBP3 p.V2109M (14.9% of 207 reads), and a VUS p.Y58C in IGF1R. No pathogenic variants were found in IDH1/2, D2HGDH, or SLC25A1 genes. Previous WES of Blood and skin sample showed a 7-10 % mosaic pathogenic R132C variant in the IDH1 gene. Pathogenic variants of the isocitrate-dehydrogenase (IDH) enzymes are one of the central research topics regarding gliomagenesis and tumorigenesis. Our patient has a very high level of 2-HG and D-2-HG, which promotes the cell cycle in cartilage bone marrow, skin, vessel, intestinal mucosa and face. This child has been published twice: Preston et al., 2019 and Srinivasan et al., 2020; there are lessons to learn. Review of 12 cases showed that our patient has the most severe cancers/growths. His CREBBP3 variant is mosaic, which could be related to the early mosaic skin pattern or could also be due to AML relapse (known strong association). The IGF1R variant may contribute to his extremely small size and the unique vascular malformations: angiokeratoma circumscriptum (AC) vs Verrucous hemangioma (VH**).** IGF1R enhances in situ endothelium regeneration. With a regeneration defect during rapid endothelial proliferation from the stimulation by D2HA, we hypothesize the vascular lesion may die/program death and appear as angiokeratoma circumscriptum instead of classic Maffucci type of hemangioma**.** VH and AC can only be distinguished histologically: in AC the vascular alterations are limited to the papillary dermis, VH extends deep into the dermis. Our patient has dermal involvement over the years which suggests the two conditions be a continuum.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session III

PB3185 *De Novo* Germline Loss of the Distal End of the P-Arm of Chromosome 17 in an Individual with Intellectual Disability

Authors:

D. Sant, E. E. Pope, M. H. Huertas, A. A. Paul, B. R. Cunningham, M. R. Jennings, R. A. Perry, S. N. Chavez, J. A. Kriak, K. B. Bills; Noorda Coll. of Osteopathic Medic, Provo, UT

Purpose: In this report we present a case of a 20-year-old female with congenital intellectual disability, stunted growth, and hypothyroidism. Competitive genetic hybridization (CHG) revealed a loss of a portion of 17p13.3 at least 195 Kb in size, and the deletion was not present in either parent. This deletion has not previously been characterized, but mutations on the p-arm of chromosome 17 are responsible for Miller-Dieker Syndrome and Isolated Lissencephaly Sequence, both of which share symptoms in common with the patient.
Methods: Peripheral mononuclear cells (PBMCs) were used for karyotyping and competitive genetic hybridization (CGH) at Baylor College of Medicine. Further bioinformatic analysis was carried out using the Genome Data Viewer (ncbi.nlm.nih.gov/genome/gdv).
Results: Karyotype was found to be normal, but CGH revealed a deletion of the tail end of the p-arm of chromosome 17, 17p13.3. At least 134 genes are present in this genomic location, 35 of which are uncharacterized. Both Miller-Dieker Syndrome (MDS) and Isolated Lissencephaly Sequence (ILS) are characterized by a smooth cerebral cortex and intellectual disability, but the patient's symptoms more closely mirror MDS because muscle tone was normal. The patient was significantly shorter than peers, but growth hormone therapy over the course of several years allowed the patient to reach a normal height, albeit shorter than her siblings and parents. Genes in the deleted region that are possibly related to the condition include *HIC1, PAFAH1B1, YWHAE, DPH1, CCDC92B, RAP1GAP2, TAX1BP1,* and *SMG6*. Further work is needed to determine if one or more of these genes is likely contributing to the phenotype. Additionally, qPCR will be used to confirm deletion of genes determined to be of interest.
Conclusions: Here we present a patient with intellectual disability and a previously uncharacterized deletion on chromosome 17. Similar, though not identical conditions have been previously reported, but not well characterized indicating that the present patient could possibly have one of these conditions. Further directions include investigation of the deleted genes to determine a probable cause for the symptoms exhibited.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session I

PB3186 † Deciphering Developmental Disorders in Central African setting (DDD-Africa): clinical characterization and molecular study

Authors:

G. Mubungu[1], P. Makay[2], D. Tshika[1], N. Louw[3], N. Carstens[3], H. V. Firth[4], M. E. Hurles[5], A. Krause[6], Z. Lombard[7], P. Lukusa[1], A. Lumaka[1,8]; [1]Université de Kinshasa, Kinhasa, Congo, Democratic Republic of the, [2]Université de Kinshasa, Kinshasa, Congo, Democratic Republic of the, [3]Univ. of the Witwatersrand, Johannesburg, South Africa, [4]Cambridge Univ. Hosp. NHS Fndn. Trust, Cambridge, United Kingdom, [5]Wellcome Sanger Inst., Cambridge, United Kingdom, [6]Div. of Human Genetics, Wits/NHLS, Johannesburg, South Africa, [7]Natl. Hlth.Lab. Service, Johannesburg, South Africa, [8]Service de Génétique Humaine, CHU de Liège, Liège, Belgium

Developmental Disorders (DD) affect about 1-3% of children globally. Clinical and genetic data related to DD in Central Africa are scarce. Given the broad genetic diversity in Africa and the fact that this continent has been understudied, we anticipate that the study of DD in Africa might yield novel insights.The aim was to recruit and characterize 150 families clinically and genetically with at least 1 individual presenting unexplained DD in DR Congo, as part of the DDD-Africa project. Patients were recruited through the collaborative network of the Centre for Human Genetics of the University of Kinshasa. A total of 149 probands were recruited, including 97 males and 52 females, recruited as Trio (n=79), Duo (n=56), Extended Duo (n=9), and Extended Trio (n=5). The mean age was 7.84±5.84 years (range 0-38.6). We excluded patients with clinically recognizable syndromes or strong suspicion of acquired causes. The clinical phenotype was determined and whole exome sequencing performed. Clinical information was deposited in Decipher using HPO terms. 130 cases (87.2%) were sporadic and 19 (12.8%) familial. In familial cases, autosomal recessive and X-linked inheritance were suspected in six cases each, autosomal dominant inheritance seven cases. This well-phenotyped cohort of 149 families with DD patients illustrates the very broad DD phenotype in patients from Central Africa. The molecular analysis of the first 68 families returned a definite diagnostic, likely pathogenic or pathogenic, in 32,4% (22 patients). Out of the 22 patients, the inheritance dominant de novo was in 36%. Many diagnosed patients represent the first to be reported with these syndromes in this community such as KBG syndrome, Stickler syndrome, GRIN2B-related neurodevelopmental disorder, Phelan-McDermid syndrome, Wolf Hirschhorn syndrome, Schinzel-Giedion midface retraction syndrome, Kaufman oculocerebrofacial syndrome, Pontocerebellar hypoplasia, type 1D, Intellectual disability and overgrowth, PPP2R5C-related, Kabuki syndrome, Myoclonic-atonic epilepsy/Developmental disorder, SLC6A1-related, Rett syndrome.
Keywords: DDD-Africa, Developmental disorders, phenotype, genomic, DR Congo

Session Title: Molecular and Cytogenetic Diagnostics Poster Session II

PB3187 Deletion *16P12.2* and small supernumerary marker chromosome (sSMC19) detected by array CGH in an adult patient with Goldenhar Syndrome.

Authors:

b. dikmen, B. Zhang, A. Iqbal; Univ. of rochester, rochester, NY

Goldenhar Syndrome (GS), also known as -Oculo auriculo vertebral spectrum (OAVS)- is a rare congenital condition characterized by impaired development of ears, eyes, nose, palate, lip, mandible and maxilla. Etiopathogenesis is multifactorial and dependent on genetic and environmental factors. The *16P12.2* recurrent deletion is inherited in an autosomal dominant manner from a carrier parent in most cases. It is linked to clinical manifestations that are frequently seen in probands, including developmental delay, speech delay, cognitive impairment, epilepsy, cardiac/skeletal malformations, and psychiatric and/or behavioral issues. The incidence of supernumerary marker chromosome (SMC) in newborn infants is about 0.044%. It generally presents in mosaic form. The patient, 31-year-old male, diagnosed with GS as an infant, has acute gastric volvulus, thrombocytopenia, splenomegaly, testicular cancer, systemic lupus erythematosus, cough, anxiety, chronic kidney disease, long term use of plaquenil, latent hypermetropia of both eyes, amblyopia in left eye, heart defect, bicuspid valve, pericarditis, long term use of immunosuppressant medication. On clinical examination, facial asymmetry is present with the right being more dominant than the left, down slanting palpebral fissures and microtia on left side, and surgically repaired cleft palate. In his upper extremity, he was noted to have left radial anomaly with short forearm and absent thumb. Chromosome microarray analysis (CMA) (4x180K CGH+SNP, Agilent technologies, USA), revealed 471kb loss at *16P12.2* from 21959891 to 22430592 base pairs and a mosaic 7.3 Mb gain at centromeric *19P12-Q12* from 22470583 to 29795821 base pairs. Interphase and metaphase FISH with dual color probe RP11-359H18 (*19P12*) and RP11-714C4 (*19Q11*) probes confirmed sSMC19 as a small ring chromosome. The level of mosaicism is approximately 80%. Genes of interest in these regions: *16P12.2: UQCRC2, EEF2K, POLR3E, CDR2. 19P12Q12: ZNF98, ZNF99, ZNF91, ZNF254, UQCRFS1.* Although many chromosomal abnormalities associated with GS have been defined, there are no specific genetic tests for diagnosis of GS. To the best of our knowledge, this is the first case report of an adult patient with GS who has *16P12.2* deletion and sSMC19. It can be challenging for clinicians to identify GS as it has a large variety of abnormalities and different severity of symptoms. This case underscores the importance of performing CMA in GS for genetic counseling and patient management.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session III

PB3188 Detecting chromosomal aberrations with high sensitivity using an experimental model for mosaic copy number variation.

Authors:

J. McGirr[1], B. Bolstad[1], L. Wu[2], V. Huynh[2], S. Baral[2], N. Vegesna[2], D. Black[2], J. Wallace[2], A. Roter[1]; [1]Thermo Fisher Scientific, Emeryville, CA, [2]Thermo Fisher Scientific, Santa Clara, CA

Mosaicism is defined as the presence of two or more genetically distinct cell lineages within an individual derived from a single zygote. Post-zygotic mutations altering chromosome number or structure result in mosaic copy number variation. This type of mosaicism typically occurs due to errors in segregation during anaphase, causing the gain or loss of entire chromosomes, and non-allelic homologous recombination, which causes the loss, gain, or rearrangement of smaller genomic regions. Mosaic copy number variants have been implicated in a wide range of clinical phenotypes, including cancer, neurodegeneration, monogenic disease, and birth defects. Microarrays are routinely used to detect chromosomal aberrations and can also determine the relative abundance of an aberration that is present in a mosaic sample. However, it is challenging to detect the gain or loss of a small copy number variant in a mosaic sample, especially when it is present in only a small fraction of the total cell population sampled. We generated an experimental model for mosaic copy number variation by mixing genomic DNA from three pairs of cell-line samples, each harboring two known autosomal chromosome aberrations between 1.47 and 71.25 megabases. Sample pairs were titrated such that each mixture contained a copy number gain and a loss on different chromosomes at frequencies ranging from 15% to 85% of the total sample. These simulated mosaic samples were analyzed using the CytoScan HD Accel microarray to test whether structural variants as small as 1.47 MB could be detected in a sample mixture at a mosaic frequency as low as 15%. For each sample, probeset signal intensities were compared with a fixed set of reference intensities to calculate a log2 ratio. ROC curves showed that Smooth Signal (a probeset level metric resulting from a kernel smooth over a window of genomically contiguous log2 ratios) could reliably discriminate between integer copy number states at mosaic frequencies between 15% and 85% (AUC range 0.96-1.00). Furthermore, the Mosaic Segmentation Algorithm implemented in the Chromosome Analysis Suite software package correctly identified mosaic segments present between 15% and 30% with 95.8% sensitivity. This research provides a framework for using titration experiments to evaluate the sensitivity of automated detection of mosaic copy number variation. The CytoScan HD Accel microarray detected small structural variants (< 1.5 MB) at a low mosaic frequency (15%) with high sensitivity.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session I

PB3189 Detection of *CFTR* polyT/polyTG variations by an NGS-based method

Authors:

Q. Zeng[1], Y. Cao[2], N. Russell[1], J. Smith[1], P. Okamoto[1], S. Letovsky[1], H. Zhu[1], N. Leach[1], A. Kenyon[1]; [1]Labcorp, Westborough, MA, [2]Labcorp, San Diego, CA

Pathogenic variants in the cystic fibrosis transmembrane conductance regulator (*CFTR*) gene have significant impact on the structure and functions of CFTR, resulting in cystic fibrosis (CF) and a range of other conditions. *CFTR* is one of the most studied human genes with many well characterized variants. Most of these variants can be detected with high accuracy by Next Generation Sequencing (NGS) technologies.
However, detecting variation in the *CFTR* polyTG/polyT tract is challenging for an off-the-shelf NGS method. PolyTG/polyT tract is located in the *CFTR* intronic region just six bases upstream of exon 10 and contains a string of TG-repeats followed by a polyT tract with (TG)11(T)7 being the most commonly observed. Shorter polyT tracts (5T or less) have been linked to reduced levels of *CFTR* mRNA and protein, which can potentially lead to *CFTR*-related disorders. Also, polyT status is of special consideration for R117H variant, where the polyT size and phase affect its clinical consequence. Because the lengths of polyTG and polyT tracts can vary between the two alleles, this region presents challenges in aligning and interpreting NGS data. In fact, variants in this region account for over 70% of all difficult pathogenic or likely pathogenic variants among the low-complexity regions in the human genome.
We have developed a computational strategy with a digital read-out for detecting variants in the *CFTR* polyTG/polyT region. Specifically, we generated simulated samples for over 13,000 different polyT and polyTG combinations and analyzed them in our bioinformatics pipeline to construct a comprehensive look-up table containing variant fingerprint profiles. The profiles were then used to match variant profiles in clinical specimens to determine the lengths of polyT and polyTG in each sample. The accuracy of our approach was confirmed by Sanger sequencing, which is generally considered the gold standard for variant detection but because it utilizes an analog signal, can be time-consuming and challenging to disentangle two alleles with varying polyTG/polyT lengths. Application of our NGS-based method to the analysis of polyTG/polyT regions in clinical samples not only offers an accurate and high-throughput solution, but also provides valuable insights into their patterns in the general population.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session II

PB3190 Detection of putative dilated cardiomyopathy causing germline variants and DNA methylation by Nanopore sequencing

Authors:

Y-Y. Wang[1,2], S-H. Chen[1], D. Badescu[1], S. Oikonomopoulos[1], N. Giannetti[3], G. Thanassoulis[3], J. Engert, C.[4,3], R. Jiannis[1]; [1]McGill Genome Ctr., Dept. of Human Genetics, McGill Univ., Montreal, QC, Canada, [2]Kyoto-McGill Intl. Collaborative Program in Genomic Med., Kyoto, Japan, [3]Preventive and Genomic Cardiology, McGill Univ. Hlth.Ctr. and Res. Inst., Montreal, QC, Canada, [4]Dept. of Human Genetics and Div. of Experimental Med., McGill Univ., Montreal, QC, Canada

Dilated cardiomyopathy (DCM) is a disease of the heart muscle that is not due to an ischemic cause, valvular disease, or coronary artery disease. The inherited forms of DCM are most commonly autosomal dominant. Previous studies have reported causative mutations in more than 40 genes, including genes encoding structural components of the sarcomere and desmosome. Ten DCM patients had previously been subjected to genetic diagnosis by sequencing, but in only one of them was a causal variant identified. We produced Nanopore sequencing data and Illumina WGS data from the patients (nine with undetermined etiology based on the sequencing panel, and one control) and their relatives. We built a Nanopore analysis pipeline to investigate small variants, structural variants, and DNA methylation patterns at the haplotype level. Nanopore sequencing data was able to detect over 94% of SNPs called by Illumina WGS data. The known pathogenic variant (rs727505319 in TTN) in the control case was identified with Nanopore data alone, while likely-pathogenic variants (rs926741242 in TTN and rs727502851 in LAMA2) were identified in two additional cases by Nanopore data and Illumina WGS data independently. A 4.4Kb deletion across multiple exons in TTN could only be identified by Nanopore. Moreover, by phasing variants using the Nanopore data, the variants rs727505319 and rs926741242 in TTN were confirmed to be inherited from their affected parents. In addition, we observed different methylation patterns between haplotypes around DCM-associated genes and a deletion that may induce hypomethylation and confirmed their inheritance. Thus, Nanopore sequencing was successful in identifying and phasing haplotypes for both small and large variants, determining the methylation status of CpG islands, and integrating this information into haplotypes, with no need for additional methylation sequencing experiments. As a result, we have identified candidate mutations in TTN and LAMA2, as well as identified inherited, potentially deleterious variants in genes not associated with DCM so far.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session III

PB3191 Detection of Trinucleotide Repeat Expansions and Uniparental Disomy by Whole Genome Sequencing.

Authors:

M. Hammer, B. Cao, D. Averbuj, M. Wright, M. Tokita, K. Watkins, K. Kelly, Y. Kwon, L. Guidugli; Rady Children's Inst. for Genomic Med., San Diego, CA

Background To diagnose genetic diseases accurately and promptly, it is critical to clinically validate a comprehensive rapid genomic testing with optimized diagnostic yield. A subgroup of genetic disease affecting the NICU population is composed by trinucleotide repeat expansion (TRE) disorders and imprinting disorders due to uniparental disomy (UPD). Trinucleotide repeat expansion (TRE) disorders are caused by an increase in the number of trinucleotide repeats exceeding the normal, stable threshold. These include Myotonic Dystrophy 1 (DM1) due to TRE in *DMPK* gene and Congenital Central Hypoventilation Syndrome (CCHS) caused by TRE in *PHOX2B*. Both diseases manifest severely at birth in presence of full penetrance, neonatal onset-associated repeat expansions. On the other end, imprinting disorders due to UPD manifest when both homologous chromosomes are inherited from one parent, with lack of contribution from the other parent and the affected chromosomes include clinically relevant imprinted genes.
Objective To develop and validate an accurate TRE and UPD analysis workflow to integrate into a comprehensive rapid whole genome sequencing (WGS) for improving clinical diagnostic yield in the NICU.
Methods TRE calling for *DMPK* and *PHOX2B* was performed via Expansion Hunter (EH) v4.0.2 (Illumina) on DRAGEN v3.10.4 (Illumina). UPD events were detected using an in-house developed algorithm. Concordance with external orthogonal assay, sensitivity, specificity, and precision were assessed, and a reporting workflow was established accordingly.
Results TRE Detection: 100% concordance with results generated by previous orthogonal methods. There was 100% agreement of all expected genotypes from all samples tested as assessed in the accuracy, precision, and reproducibility studies. UPD Detection: 100% concordance with results generated by previous orthogonal methods when the UPD algorithms were not restricted to identify UPD within the critical regions, rather expanded to call any region of UPD within the whole clinically relevant imprinted chromosome (chr 6, 7, 11, 14, 15, 20). For samples with more than one replicate tested, all replicates were concordant. Due to the lack of a reference gold standard with hetero and isodisomy details and the restrictions of UPD algorithms-based detection, the validated method has some limitations.
Conclusion The addition of TRE and UPD on rapid WGS testing is expected to increase diagnostic yield and possibly lead to a more precise appreciation of the prevalence of such disorders in the neonatal population. Additional prospective work will further enhance the detection of TRE and UPD on WGS.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session I

PB3192 Development of a rapid genotyping method for screening of RNF213 p.Arg4810Lys variant in patients with intracranial artery stenosis (ICAS).

Authors:

P-s. Lai, G. Tan, C. P. Liu, M. B. J. Ng, L. L. L. Yeo, V. K. Sharma, B. Y. Q. Tan; Natl. Univ. of Singapore, Singapore, Singapore

The ring finger 213 gene (*RNF213*; NM_001256071.2) encodes a 591kDa cytosolic E3 ubiquitin ligase with RING finger and AAA+ ATPase domains. It has been identified as the major susceptibility gene for Moyamoya disease in East Asian populations. More recently, it has been reported to be linked to other vascular diseases including intracranial artery stenosis (ICAS). ICAS is the most common cause of ischemic stroke in Asia, and has a much higher prevalence in Asian compared to Caucasian populations. In several Asian cohorts, patients with ischemic stroke secondary to ICAS have been observed to carry the *RNF213* p.Arg4810Lys (c.14429G>A, rs112735431) variant. Presence of this rare variant has been reported to be associated with recurrent cerebrovascular events in ICAS patients. About 1-2% of healthy individuals in some East Asian populations have been reported to carry this variant. Therefore, a rapid and cheap genotyping method to screen for *RNF213* p.R4810K would be clinically useful for investigating the association of this variant with ICAS. In this case-control study, we developed a simple one-step tetra-ARMS assay to screen *RNF213* p.Arg4810Lys in a cohort of 100 symptomatic ICAS patients prospectively recruited from a tertiary stroke centre in Singapore. ICAS patients were diagnosed using neurovascular imaging with either computed tomography angiogram or magnetic resonance angiogram of the brain and had involvement of either the terminal internal carotid artery or middle cerebral artery with at least moderate severity (>50% stenosis). Genotyping of *RNF213* p.Arg4810Lys (c.14429G>A) showed 3 patients carrying this variant (3%), while a control cohort of 100 healthy individuals were homozygous for the wild-type G allele. The results were confirmed using both RFLP and sequencing assays. We demonstrate the involvement of *RNF213* p.Arg4810Lys in ICAS patients in Singapore. Further large scale studies to investigate ethnic variations can be performed using similar assays targeted towards other *RNF213* variants.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session II

PB3193 † Development of a region specific enrichment and long read sequencing strategy to phase de novo mutations in human genetic disease.

Authors:

J. Graham, J. Pitsch, M. Breuss; Univ. of Colorado Anschutz Med. Campus, Aurora, CO

Next-generation sequencing (NGS) panels have greatly accelerated the detection of disease causing de novo mutations (DNMs) in prenatal and pediatric cases. Clinically, the term de novo is applied when the causal variant is not detected in the peripheral blood or saliva of either parent. However, recent research has shown that recurrence of a specific mutation in subsequent pregnancies is possible depending on the method of inheritance of the variant. The parent of origin of a DNM modifies its recurrence risk, and known paternal variants can be directly assessed in sperm for accurate recurrence risk detection. As a result, there is clinical benefit in performing haplotype phasing of DNMs to understand which parent transmitted the variant. This is done by identifying informative single nucleotide polymorphisms (iSNPs) near the mutation that are present in the proband as well as either parent. Once iSNPs have been identified on the same allele as the DNM, the haplotype can be phased to the parent of origin. Previously, this has required costly whole genome sequencing for iSNP identification and in most cases, short read NGS is unable to capture a DNM and iSNPs within one read to resolve the haplotype. To overcome this challenge, we combine nanopore long-read sequencing with a protocol for region specific enrichment (RSE) of the DNM to increase our target sequencing depth and ability to detect iSNPs directly without generating unnecessary genomic data. To develop our protocol, we tested two RSE approaches: biotinylated 120 base pair capture probes and a biotinylated, enzymatically dead Cas9 (dCas9). Both protocols then rely on binding to streptavidin magnetic beads for isolation of the target sequence. Recovered fragments are subjected to whole genome amplification prior to determination of fold enrichment (FE) by qPCR analysis. Enriched DNA fragments are then sequenced on the Oxford Nanopore MinIon. Our results have shown that the most promising approach is the use of the dCas9 protocol, as we have seen approximately 30 to 50 target FE with three different genomic targets. While the 120 bp oligos showed a similar level of target FE, the cost of these probes is significantly higher than the sgRNAs and dCas9, making it a less flexible targeted assay. Additionally, our nanopore sequencing of RSE products has revealed that we can, indeed, phase known variants of a proband sample to either parent, thus, demonstrating a proof of concept. Our future goals include employing our assay to phase human de novo mutations to provide clinicians with a better understanding of recurrence risk, determine cis or trans relationship of pathogenic variants, or determine DNM status in incomplete trios.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session III

PB3194 Development of a targeted, low cost and high throughput single molecule molecular inversion probe based sequencing assay for rapid diagnosis of lysosomal storage disorders in low-middle income healthcare settings

Authors:

H. Sheth[1], A. Nair[1], C. Prabha[2], M. Joshi[3], C. Joshi[3], J. Sheth[1]; [1]FRIGE's Inst. of Human Genetics, Ahmedabad, India, [2]The Maharaja Sayajirao Univ., Vadodara, India, [3]Gujarat Biotechnology Res. Ctr., Gandhinagar, India

Lysosomal storage disorders (LSDs) are a collection of 70 monogenic disorders of lysosomal dysfunction with a combined incidence of 1:5000 in the general population. Current gold standard diagnostic strategy involves a sequential process of biochemical screening, followed by enzyme assays and gene sequencing. Due to the overlapping phenotypic features of some LSDs, the diagnostic pathway is expensive and time-consuming due to its iterative nature. Here, we describe a novel, low-cost and high-throughput targeted sequencing assay using single molecule molecular inversion probes (smMIPs) to carry out rapid genetic diagnosis of 29 common LSDs in India. 903 smMIPs were designed to target exon and exon-intron boundaries of 23 genes- *ARSA, ARSB, CLN6, GAA, GALC, GALNS, GBA, GLA, GLB1, GNPTAB, HEXA, HEXB, IDS, IDUA, NAGLU, NPC1, NPC2, PPT1, PSAP, SGSH, SLC17A5, SMPD1* and *TPP1*- using MIPgen and hg19 genome build. Pooled probes were used to generate sequencing libraries. Validation and diagnostic yield were assessed in a cohort of 49 patient samples with a known genetic diagnosis and 111 patient samples with clinical suspicion and/or biochemical diagnosis of LSD, respectively. The Illumina MiSeq platform was used for sequencing at a mean coverage of 200x. Data was analysed using a custom bioinformatics pipeline consisting of consensus calling using unique molecular barcodes. The average percentage of the coding region of the 23 genes covered by 903 smMIPs was 53.7kb (99.17%). A 98% (n=48/49) concordance was observed in the validation cohort with the discordance arising in 1 sample due to the lack of smMIP probes covering the exon 1 of the *GALNS* gene due to low sequence complexity. In the diagnostic yield cohort, 73.9% diagnostic yield (n=82/111) was observed. Crucially, 7 patients with Niemann-Pick type C and 1 patient with neuronal ceroid lipofucinosis-6 were identified, which can't be readily detected by enzyme testing. Additionally, the assay was able to detect p.L483P variant in the *GBA* gene in Gaucher disease patients with 100% accuracy. This is the most common variant observed in these patients but is currently difficult to detect by exome sequencing assays due to the presence of the pseudogene- *GBAP1*. Lastly, single and multi-exon heterozygous and homozygous CNVs were detected in 6 cases. Combined cost per sample including- reagents, sequencing on MiSeq platform and computational analysis is estimated to be US$60-85. Overall, the assay is robust in calling SNVs and CNVs together, thus obviating the need for MLPA. The proposed assay with its superior diagnostic yield and low cost makes it suitable for deployment in low-middle income healthcare settings.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session I

PB3195 † Diagnosing challenging cases: Eight years of experience with the Undiagnosed Diseases Network (UDN) at Baylor College of Medicine (BCM).

Authors:

J. Rosenfeld, H. Bellen, L. S. Blieden, L. C. Burrage, H-T. Chao, I. Chinn, G. D. Clark, W. Craigen, H. Dai, B. C. Dawson, L. Emrick, C. M. Eng, S. Ketkar, F. Jamal, S. Lalani, R. Lewis, P. Liu, K. Machol, R. Marom, P. Moretti, D. Murdock, S. K. Nicholas, J. P. Orengo, J. Posey, L. Potocki, D. Scott, A. Tarakad, T. Vogel, M. Wangler, M. Weisz-Hubshman, K. Worley, S. Yamamoto, Undiagnosed Diseases Network, C. Bacino, B. H. Lee; Baylor Coll. of Med., Houston, TX

The UDN seeks diagnoses with multidisciplinary, in-depth phenotyping and innovative genetic testing technologies. Eight years of evaluations at the BCM UDN clinical site have yielded a 38% (107/284) diagnostic rate. Nearly all diagnoses are genetic (103/107), highlighting the importance of genetic testing technologies. While almost half of diagnoses were a result of newer medical genetic knowledge, including novel disease genes (23%), recently identified disease genes (9%), and phenotypic expansions for known disease genes (13%), many diagnoses were for known genetic disorders that remained undiagnosed prior to study enrollment, frequently despite extensive genetic testing. In at least two-thirds of all genetic diagnoses, disease-causing variants were detectable by exome sequencing (ES). In 15 diagnoses of a known genetic disorder, previous ES had not reported diagnostic variants that were observable in extant ES data: 40% intronic variants, 53% coding variants, and 7% a copy number variant (CNV) *in trans* with a coding variant. Analytic methods used to identify diagnostic variants from ES data include addition of parental ES and loosening of filters relating to intronic location, depth of coverage, and/or phenotype matching. CNVs (4-977kb) were part of 15 diagnoses. While some larger CNVs were detected via chromosomal microarray (CMA), many smaller CNVs were not found on CMA (due to array design) and/or ES (due to limitations and/or absence of CNV calling on clinical ES analyses). In addition to small CNVs, other diagnostic variants not detectable by ES included noncoding variants, mosaic findings, repeat expansions, and a balanced translocation. Genome sequencing (GS) is a powerful tool used to detect many of these diagnostic variants, but other technologies, including long-read sequencing and optical genome mapping, detected pathogenic variants in some cases. Regardless of the testing technology, simply identifying a genomic variant is frequently insufficient to establish a diagnosis. Additional functional evidence may be required. RNAseq has proven a useful tool: RNAseq was performed in samples from 72 subjects with genetic diagnoses and provided supporting data in 50%, including 14% in which RNAseq gave the first evidence for the diagnosis. Additional supporting data came from candidate gene-specific functional studies. Finally, particularly in cases of novel discovery, model organisms and patient matching platforms facilitated final diagnoses. Therefore, while GS can improve detection of genomic variants, challenging cases will likely continue to require additional tools to achieve diagnoses.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session II

PB3196 Diagnostic Yield from Cardiogenomic Panel Testing for Inherited Cardiovascular Diseases

Authors:

R. Pyatt, J. Reiner, Z. Wolf, W. Xin, Z. Zhou, H. Zhu, N. Leach; Labcorp, Westborough, MA

Multiple organizations have issued practice guidelines and statements recommending genetic testing for heritable heart conditions. The American College of Medical Genetics and Genomics published a 2018 practice resource which recommended genetic testing using a 39 gene panel for patients with cardiomyopathy and at-risk relatives. The Heart Rhythm Society and European Heart Rhythm Association published a 2011 consensus statement which recommended a 9 gene panel for individuals with a suspected arrhythmic disorder. In a 2020 scientific statement on genetic testing, the American Heart Association reiterated the findings of the ClinGen Aortopathy Working Group which found definitive evidence associating 11 genes with heritable thoracic aortic aneurysms and dissections along with moderate/limited evidence for an additional 8 genes. Consequently, a retrospective analysis was conducted on the results from cardiogenomic testing in our laboratory to evaluate diagnostic yield. Panel based testing included 10 genes for aortopathy and the *MED12* c.3020A>G variant, 29 genes for arrhythmia, and 48 genes for cardiomyopathy. Disease-associated variants (Pathogenic and Likely Pathogenic) and variants of uncertain significance (VUS) were reported for all panels. Test results included 197 reportable variants across 1385 aortopathy panels, 333 variants across 596 arrhythmia panels, and 1,043 variants across 831 cardiomyopathy panels. Disease-associated variants were identified in 2% of aortopathy, 13% of arrhythmia, and 16% of cardiomyopathy panels. Only 1% of aortopathy panels identified more than one reportable variant, but significantly more of such scenarios were seen in arrhythmia (12%) and cardiomyopathy (35%) testing. One cardiomyopathy panel reported a single pathogenic *PKP2* gene variant along with seven VUS in the *DSP*, *MYH7*, *RYR2*, *SGCD*, and *TTN* genes. Only one panel had more than one disease-associated variant; a cardiomyopathy sample with a *FKTN* gene pathogenic variant and a likely pathogenic variant in the *DSP* gene. A majority of reportable variants for the aortopathy panel were identified in the *FBN1* (32%), *MYH11* (17%) and *COL3A1* (17%) genes, for the arrhythmia panel in the *KCNQ1* (13%), *SCN5A* (11%) *RYR2* (11%) and *KCNH2* (11%) genes, and for the cardiomyopathy panel in the *TTN* (37%), *ALMS1* (10%), *MYBPC3* (6%) and *MYH7* (6%) genes. Variants of uncertain significance were the most common group identified including 84% of aortopathy, 77% of arrhythmia, and 87% of cardiomyopathy reportable variants. These results reaffirm the diagnostic value of genetic testing for inherited cardiovascular diseases and illustrate challenges in result interpretation.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session III

PB3197 *Diagnostic yield of exome and genome sequencing after non-diagnostic multi-gene panels*

Authors:

M. Vernet Machado Bressan Wilke[1], E. Klee[2], R. Dhamija[3], F. Fervenza[4], Q. Qian[5], N. Leung[4], J. Kemppainen[2], K. Hanson[1], C. Vitek[2], K. Lazaridis[6], F. Vairo[7]; [1]Ctr. for Individualized Med., Mayo Clinic Rochester MN, Rochester, MN, [2]Ctr. for Individualized Med., Mayo Clinic, Rochester, MN, [3]Dept.s of Clinical Genomics, Mayo Clinic, Phoenix, AZ, [4]Dept. of Nephrology And Hypertension, Mayo Clinic Rochester MN, Rochester, MN, [5]Mayo Clinic Coll. of Med. and Sci., Mayo Clinic Rochester MN, Rochester, MN, [6]Div. of Gastroenterology & Hepatology, Mayo Clinic, Rochester, MN, [7]Dept.s of Clinical Genomics, Mayo Clinic, Rochester, MN

BACKGROUND: Next-generation sequencing (NGS) tests have become the primary diagnostic approach for rare diseases (RD) in the past decade. Common NGS approaches in clinical practice include targeted multi-gene panels (GP), exome sequencing (ES), genome sequencing (GS), and exome/genome-based panels (EGBP). While the value of NGS testing is well-established, determining the optimal implementation strategy remains debatable. Limited studies have reported on the additional diagnostic yield of reanalysis of ES and GS data in patients with unrevealing gene panels. Mayo Clinic's Program for Rare and Undiagnosed Diseases (PRaUD) offers genomic-based clinical services for RD, employing targeted GP or custom EGBP as the first-tier diagnostic strategy across various specialty divisions and departments. METHODS: ES/GS data of patients included in PRaUD was analyzed to determine if additional findings could enhance the diagnostic yield of previously performed genetic testing. The inclusion criteria consisted of patients evaluated at Mayo Clinic campuses in Minnesota, Florida, and Arizona between December 2018 and May 2023, who had non-diagnostic EGBP. RESULTS: ES data from 70 patients (53 adults, 42 females) and GS data from 13 patients (10 adults, 2 females) were analyzed. The mean age of patients is 41 years (range 3 to 81 years). Most patients were referred for suspicion of hereditary kidney disease (n=36), followed by auto-inflammatory syndromes (n=27), interstitial lung disease, and monogenic diabetes (n=10). After ES/GS data analysis, no additional relevant variants were identified in 80 cases (96.4%). In the remaining three cases (3.7%), a variant in a gene associated with a phenotype reported after the release of the original report (*RRAGD*), a variant in *COL4A3* not reported in the original panel testing, and a likely pathogenic variant in *HNF1A* related to part of patient's phenotype not targeted by the panel were found. Periodic automated re-analysis during the period flagged additional variants in 17 cases that were deemed not relevant for patient's phenotype after additional review. CONCLUSION: Based on our experience, ES and GS have similar diagnostic yield when compared to EGBP curated by a multidisciplinary team of experts. However, EGBP has limitations in identifying new gene-disease associations, which necessitates periodic updates to ensure their continued relevance.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session I

PB3198 Diagnostic yield of genome screening in the general adult population

Authors:

S. Casalino[1], E. Frangione[1], N. Aujla[1], D. Di Iorio[1], L. Jayachandran[1], J. Young[1], C. Fung[1], R. Mahajan[1,2], G. Morgan[1,2], M. Scott[1,2], D. Wolday[1], S. Arnoldo[2,3], E. Bearss[1], A. Binnie[2,3], B. Borgundvaag[1,2], H. Chertkow[4], S. Chowdhary[5], M. Clausen[6], M. Dagher[7], L. Devine[1,2], B. Dickson[1], S. M. Friedman[1,5], A-C. Gingras[1,2], L. W. Goneau[8], S. Haller[9], D. Kaushik[3], Z. Khan[9], E. Lapadula[1], G. MacDonald[1], T. Mazzulli[1,2], A. McGeer[1,2], S. McLeod[1,2], C. Mighton[1,2,6], R. No[9], T. J. Pugh[5,10], D. Richardson[3], J. Simpson[5,10], S. Stern[9], L. Strug[11], A. Taher[9,5,2], S. Thakur[1], I. Wong[9], N. Zarei[9], E. Greenfeld[1,2], A. Noor[1,2], Y. Bombard[2,6], H. Faghfoury[5], J. Taher[1,2], J. Lerner-Ellis[1]; [1]Sinai Hlth., Toronto, ON, Canada, [2]Univ. of Toronto, Toronto, ON, Canada, [3]William Osler Hlth.System, Brampton, ON, Canada, [4]Baycrest Hlth.Sci., Toronto, ON, Canada, [5]Univ. Hlth.Network, Toronto, ON, Canada, [6]Unity Hlth.Toronto, Toronto, ON, Canada, [7]Women's Coll. Hosp., Toronto, ON, Canada, [8]Dynacare Med. Lab., Brampton, ON, Canada, [9]Mackenzie Hlth., Richmond Hill, ON, Canada, [10]Ontario Inst. for Cancer Res., Toronto, ON, Canada, [11]The Hosp. for Sick Children, Toronto, ON, Canada

Background & Aim: Genome sequencing (GS) may be used in the clinical context to solve diagnostic odysseys. However, GS may also be used as a screening tool or reveal secondary findings (SF) related to inherited predisposition to disease. In some cases, SF may explain a previously undiagnosed medical or family history of disease. Advancements in GS warrants exploration into how this technology may be used at scale. We aim to evaluate the diagnostic yield of GS in symptomatic adults within the general population. Methods: Adults with a prior diagnosis of COVID-19 were recruited for GS through the GENCOV study in Ontario, Canada. Medical history data was collected through an online intake survey. GS data was filtered and analyzed using frequency and database filters and variants were classified based on ACMG guidelines. Results counseling and clinical intake were initiated for participants with clinically significant findings requiring follow-up. Survey, intake and GS data were consolidated to identify those with self-reported history of disease and a pathogenic or likely pathogenic (P/LP) variant associated with personal disease risk. Results: To date, 656 genomes have been analyzed. Of these individuals, 652 (99%) responded to the intake survey. Sixty percent (n=393/652) reported at least one underlying medical issue or disease comorbidity at the time of intake (e.g. cancer, cardiac disease, diabetes, high blood pressure). Of participants reporting at least one underlying medical issue, 84 (21%) had at least one P/LP variant associated with personal disease risk(s). A P/LP variant identified by GS was consistent with a previously reported medical or family history in 21 cases for a diagnostic yield of 5%. Six participants were homozygous or compound heterozygous for variants associated with autosomal recessive or semi-dominant conditions. Thirty-eight participants indicated that they completed genetic testing prior to enrollment in the study, and 14 had a known previously diagnosed hereditary condition. Conclusion: GS identified risk for hereditary disease in 21% of adults with underlying medical issues. GS findings explained a previously reported history in 5% of cases. These findings illustrate the utility of GS to diagnose previously undiagnosed disease in the general population. Early diagnosis is crucial for reducing disease risk in the context of actionable conditions with known medical treatment, prevention, or screening, as well as solving undiagnosed symptoms in adults that may otherwise present undue costs to the healthcare system by way of unnecessary testing and treatment.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session II

PB3199 Early use of whole exome sequencing (WES) informs lifelong medical management: case report.

Authors:

E. Dreikorn[1], C. Munro[1], N. Robin Berman[1], A. Kunovac[2], D. Bellissimo[3], M. B. Massart[1]; [1]Univ. of Pittsburgh, Dept. of Family Med., Pittsburgh, PA, [2]UPMC Genome Ctr., Pittsburgh, PA, [3]UPMC Clinical Genomics Lab., Pittsburgh, PA

Traditional medical management for patients with neurodevelopmental disorders (NDD) often involves the lengthy and expensive process of diagnostic testing, which puts immense burden on patients, their families, and the healthcare system. These barriers to diagnosis can often be overcome with the use of whole exome sequencing (WES), which provides the genetic resolution necessary to inform medical management now and throughout the patient's life. A 27-year-old patient was referred to the Primary Care Precision Medicine (PCPM) clinic with intellectual disability (ID), autism spectrum disorder (ASD), obsessive-compulsive disorder (OCD), tics, seizures, and PANDAS disorder. Previously, a microarray and single gene testing for *PTEN* were completed in 2009 and yielded no diagnostic results. The PCPM clinic ordered a WES duo, which revealed the presence of a missense SNV (c.2425-2A>G) in a splice acceptor site of the *HNRNPU* gene that was classified as pathogenic. Disruptions of *HNRNPU* can lead to HNRNPU-related NDD and developmental and epileptic encephalopathy-54 (DEE54), which suggests that this variant is causal for the patient's ID/ASD, seizure history, and encephalopathy. After WES and original diagnosis, two of the patient's physicians contacted the PCPM team with additional genetic questions. A UPMC physician requested reanalysis of WES to evaluate for hereditary hemorrhagic telangiectasia (HHT) as the patient's maternal aunt had recently been diagnosed. Reanalysis demonstrated no findings consistent with HHT and prevented the need for further diagnostic testing. Further, a UPMC hematologist contacted the PCPM team after suspecting the patient of having HFE-related hereditary hemochromatosis (HH). Indications were intermittent elevated iron levels and previous *HFE* sequencing in 2009 for HH that revealed a heterozygous missense variant in the *HFE* gene (p.C282Y), which leads to HH when the variant is homozygous. Reanalysis of the WES showed no compound heterozygous mutations that would support a full diagnosis of HH. WES is increasingly used to decrease the length of a patient's diagnostic journey for complex phenotypes; early use of WES can significantly decrease the number of tests and cost associated with identifying underlying causes for a patient's symptoms. As WES was used for the patient's initial presentation of ID/ASD, clinicians were able to reanalyze the sequencing data to inform two subsequent clinical indications for genetic data, shortening the length to diagnostic resolution. This case serves as an example of the clinical utility of WES across a patient's lifespan and the coordination of reanalysis through a primary care setting.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session III

PB3200 Enhancement of Trophoblast purity through removal of maternal cells using CD45 and CD56 cell surface markers.

Authors:

H. Jang[1], J. Park[2], K. Kang[2], M. Go[3], H. Kim[2], S. Yang[2], J. Kim[4], S. Shim[5], D. Cha[6]; [1]CHA Univ., Seoul, Korea, Republic of, [2]CHA biotech, Seoul, Korea, Republic of, [3]CHA Univ., Seoul, Korea, Republic of, [4]CHA Biotech Inc., Seoul, Korea, Republic of, [5]Cha Univ., Seongnam, Korea, Republic of, [6]Gangnam CHA Hosp, Seoul, Korea, Republic of

Introduction Trophoblast retrieval and isolation from the cervix (TRIC) is an alternative method to non-invasively obtain fetal origin trophoblast cells that can be used for prenatal diagnosis. The most important thing in TRIC is how to remove maternal cells, mainly immune cells such as natural killer (NK) cells and white blood cells (WBC) from cervical fluid and to isolate only fetal cells. In this study, we tried to isolate fetal cells using maternal-cell-specific surface markers. **Materials and methods** The endocervical cells were collected from pregnant women between 5-20 weeks of gestational age. To remove maternal immune cells from endocervical fluid, the cells are separated using a magnet labeled with leukocyte common 45 (CD45) or leukocyte common 56 (CD56), which are maternal-cell-specific surface markers for maternal immune cells. CD45 or CD56 positive cells are removed from the endocervical fluid using a magnet. Human leukocyte antigen G (HLA-G) positive cells were separated from the supernatant from which immune cells were removed for trophoblast separation, and the HLA-G positive cells were counted. Slides are prepared to identify trophoblasts expressing CD45, CD56, HLA-G, cytokeratin 7 (CK7), and β-human chorionic gonadotropin (β-hCG) by immunofluorescence (IF). **Result** The cell counting rate(%) was compared with the expression of specific genes in trophoblast cells between with and without maternal immune cell sorting. The trophoblast cells with removed immune cells were decreased number of cells. The expression of maternal immune cell specific genes CD45 or CD56 was decreased in trophoblast cells with immune cell isolation, but the expression of trophoblast cell specific genes β-hCG and CK7 was increased. However, double gene isolation labeled CD45 and CD56 were aggregated endocervical cells and did not differ in trophoblast cells. **conclusion** In previous study, the trophoblast was detected in endocervical cells, and maternal immune cells were also identified. To isolate maternal immune cells, cells were sorted from cervical cells using maternal cell-specific markers, and the number of immune cells was reduced. Also, the trophoblast specific gene was increased in trophoblast cells with removed maternal immune cells. In conclusion, maternal immune cell isolation was suggested as a method to separate trophoblasts more purely from maternal endocervical cells.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session I

PB3201 Enhancing Diagnostic Precision in Inherited Ocular Disorders: Insights from an iterative analytical approach.

Authors:

E. Ullah[1], B. Guan[1], A. Naik[1], C. Bender[1], R. Al Rawi[1], D. Blain[1], A. Agather[1], E. O'Heir[2], L. A. Huryn[1], W. M. Zein[3], C. A. Cukras[3], B. P. Brooks[1], A. O'Donnell-Luria[2], R. B. Hufnagel[1]; [1]Natl. Eye Inst., NIH, Bethesda, MD, [2]Ctr. for Mendelian Genomics, Broad Inst. of MIT and Harvard, Cambridge, MA, [3]Natl. Eye Inst., NIH, BETHESDA, MD

Introduction: Inherited ocular disorders are a diverse group of conditions affecting 1 in 1000 individuals worldwide, exhibiting clinical and genetic heterogeneity. These disorders can be isolated or syndromic, following any inheritance pattern. Over 500 genes have been implicated in causing these disorders. Genetic diagnosis is crucial for understanding disease mechanisms, improving patient counseling, potential therapies, and clinical trial eligibility. Current genetic testing has yielded diagnosis rates of 50-70% for inherited retinal dystrophies (IRDs) and 20-40% for developmental malformations (microphthalmia, anophthalmia, coloboma). Methods: This study used a systematic approach at the Ophthalmic Genomics Laboratory (OGL) to improve diagnostic yield in inherited ocular disorders. 479 patients underwent clinical evaluation and genetic testing on an exome/genome protocol. Gene panel-based testing with virtual Next Generation Sequencing (NGS) panels and exome-level Copy Number Variant (CNV) calling was performed. Exome data underwent re-analysis using an in-house pipeline. Unsolved cases underwent genome sequencing and analysis using the same pipeline. Results: The virtual panel-based approach identified pathogenic/likely pathogenic variants or high confidence Variants of Uncertain Significance (VOUSs) in 50% (242/479) of cases. Exome re-analysis resolved an additional ~2% of cases. Genome sequencing identified pathogenic/likely pathogenic variants or high-confidence VOUSs in 10 cases, contributing another 3% to the diagnostic yield, with around 100 cases pending genome data. ACMG secondary findings were found in ~5% of cases. Rare and damaging variants in over 30 previously unidentified genes may expand phenotypes for known disorders or establish new gene-disease relationships. Conclusion: Virtual exome-based panels are a powerful tool for diagnosing inherited ocular disorders, achieving a diagnostic yield of approximately 50%. Exome re-analysis captures additional disease-causing variants in lesser-known or new disease genes, as well as atypical variants in genes with a higher disease burden. Genome sequencing increases the diagnostic yield by identifying CNVs and deep intronic variants, primarily in known disease loci.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session II

PB3202 Enhancing Fabry Disease Screening and Diagnostic Efficiency: Integration of Enzyme, LysoGb3, and Next-Generation Sequencing Testing

Authors:

Y. Pan[1], L. King[2], E. Vengoechea[2], S. Smith[1], Y. Wang[1], P. Chakraborty[1], E. Kallu[3], B. Nallamilli[1], T. Donti[1], A. Walter[2], S. Kazemi[2], M. Hegde[1]; [1]Revvity Omics, Pittsburgh, PA, [2]Sanofi, Cambridge, MA, [3]Lambert High Sch., Suwanee, GA

Fabry disease is an X-linked lysosomal storage disorder caused by a deficiency of alpha galactosidase A (α-Gal A) due to pathogenic variant in *GLA* gene. This study reports findings from Fabry disease data collected through The Lantern Project, sponsored by Sanofi, including α-Gal A enzyme assay from dried blood spots (DBS), globotriaosylsphingosine (LysoGb3) biomarker assay from DBS and next-generation sequencing (NGS). A total of 513 enzyme, 284 LysoGb3 and 994 NGS tests were conducted on 1380 individuals (708 female, 661 male, 11 sex-unknown) who participated in the project from December 2018-April 2023. Among these, 21% (103 male and 2 sex-unknown) had a low enzyme level (range 0.054-1.069 μM/h, Normal ≥ 1.10 μM/h), and 70% (79 female, 115 male and 2 sex-unknown) had a high LysoGb3 level (range 1.12-130.56 ng/mL, Normal ≤ 1.11 ng/mL). 379 cases had reportable *GLA* variants. The most frequently identified *GLA* variants in this cohort included c.427G>A (56), c.679C>T (23), c.1088G>A (19), c.593T>C (18), c.644A>G (17), and c.1087C>T (16). 76% (104/136) of the variants have only been identified in 1 or 2 cases. Novel *GLA* variants identified in this cohort included c.[351T>G;361G>C], c.370-558_370-1del, c.370G>T, c.548del, and c.1165C>T. Within 188 possibly affected female cases (with abnormal LysoGb3 or reportable *GLA* variant or both), 40% (72) are reported to have Fabry-related symptoms (excluding 9 newborn screening confirmation cases). The symptomatic rate is not age-related, as the same rate is observed for childhood, adolescent, and adult females. Among the 30 reportedly symptomatic females with normal or unknown LysoGb3 levels, 100% were identified by NGS to have reportable variants. The symptomatic rate is related to the LysoGb3 levels of female patients. 22% of the 36 female cases with normal LysoGb3 are reported to have symptoms, while 33% of the 18 cases with LysoGb3 levels between 1.15-1.96 ng/mL and over 60% of the 58 cases with LysoGb3 levels ≥2 ng/mL are reportedly symptomatic. Female cases with *GLA* pathogenic/likely pathogenic variants c.427G>A (24), c.1087C>T (11), c.644A>G (10), c.1088G>A (9), and c.593T>C (7) mostly lack symptom information. Whereas 57% (8/14) of the females with c.679C>T and 100% (3/3) with c.337T>C have reported symptoms. We present the largest-to-date comprehensive multi-testing Fabry cohort with demographic information and phenotypes from a single clinical laboratory. Our dataset provides important genotype-phenotype correlation in Fabry disease and demonstrates that the integration of enzyme, LysoGb3, and NGS testing can enhance the screening/diagnostic efficiency for Fabry disease, especially for female patients.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session III

PB3203 Enhancing the power of NGS: Closing the diagnostic gap using short-read whole-exome sequencing.

Authors:

A. Haimovich[1,2], E. Javasky[1], N. Mor[1], Y. Khavkin[1], I. Bar-Joseph[1], O. Barel[3]; [1]Sheba Med. Ctr., Ramat-Gan, Israel, [2]Tel Aviv Univ., Tel Aviv-Yafo, Israel, [3]Sheba Med. Ctr., Ramat Gan, Israel

In recent years, whole exome sequencing (WES) has emerged as a groundbreaking tool for uncovering pathogenic variants and diagnosing genetic diseases. While WES has significantly enhanced our ability to provide accurate and efficient diagnoses, it does have certain limitations. One particular challenge arises when mutations occur in pseudogenes and homologous regions. These regions often pose difficulties during analysis, leading most Next-Generation Sequencing (NGS) computational tools to exclude reads with multiple alignments. Unfortunately, this exclusion can result in missed genetic diagnoses, leaving patients without proper answers. To overcome this hurdle, we have developed multiple modifications to our WES pipeline, focusing on detecting mutations in these intricate regions. By combining our understanding of these complex regions with the patients' clinical phenotype, we have successfully provided genetic diagnoses for individuals who would have otherwise received negative WES results. Notably, we have identified significant genes such as *HBA2*, *NSF*, *IKBKG*, and *CBS* using our tailored methodology. Our approach aims to surpass the limitations of standard short-read WES by ensuring the accurate detection of mutations in pseudogenes and homologous regions. By implementing these pipeline adjustments, we aspire to increase the identification of disease-causing variants, ultimately assisting clinicians in delivering precise genetic diagnoses. This novel methodology holds the potential to greatly enhance diagnostic outcomes, meeting a critical need in the field of clinical genetics, and improve patient care and outcomes.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session I

PB3204 Evaluation of diagnostic utility of custom array CGH for identification of copy number variants in patients with epilepsy and epileptic encephalopathies.

Authors:

A. Kucinska[1], W. Hawuła[1], M. Piotrowicz[1], L. Przesór[1], T. Chilarska[1], L. Kępczyński[1], K. Połatyńska[2], L. Przysło[2], A. Gach[1]; [1]Dept. of Genetics, Polish Mother's Mem. Hosp.-Res. Inst., Lodz, Poland, [2]Dept. of Dev.al Neurology and Epileptology, Polish Mother's Mem. Hosp.-Res. Inst., Lodz, Poland

Epilepsy is characterised by occurrence of recurrent, unprovoked epileptic seizures. This common neurological condition affects over 70 million people worldwide, accounting for about 1% of the total population. Epileptic encephalopathies encompass severe and drug-resistant form of epilepsy associated with cognitive and neurodevelopmental disorders. Genetic etiology of epileptic syndromes is complex and heterogeneous. With the development of genetic diagnostics, an increasing number of genes associated with epilepsy are being identified. Currently, several hundreds of such genes have already been reported. The contribution of genetic factors is estimated at approximately 70% of which about 5-12% are copy number variants (CNVs). The aim of this study is to demonstrate the clinical usefulness of dedicated array Comparative Genome Hybridization (aCGH) in patients with epilepsy and epileptic encephalopathies.

The tested group included 36 pediatric patients diagnosed with epilepsy, epileptic encephalopathy or epileptic syndrome and consisted of 24 males and 12 females. Prior to recruitment, patients tested negative using commercial targeted sequencing panels. DNA was extracted from peripheral blood. Array CGH was performed using 4x180k CytoSure Custom Array (OGT; UK). This microarray has specific coverage in 264 genes with both exonic and intronic sequences of proven and probable clinical significance in epilepsy and epileptic encephalopathies.

Overall, the strategy allowed for the identification of seven non-polymorphic relevant CNVs. Three of them were classified as pathogenic and four as variant of uncertain significance (VUS). Among these reported variants were six deletion and one duplication with the size range of 9.25 kb to 1.63 Mb. Importantly, four out of six variants would not be detected by standard microarray in routine testing. One of these cases is an intragenic deletion in CPA6 gene, which was not previously reported in the available databases or described in the literature.

The cases of patients with epilepsy presented here raise the question of whether routinely used standard microarrays are sufficient to continue genetic diagnostics in patients with normal sequencing-based test results.. Our results on a relatively small group encourage to study a much larger number of patients. We should consider the need to implement more targeted methods including dedicated chromosome microarrays for more efficient diagnosis in epilepsies and epileptic encephalopathies. This diagnostic tool may allow to expand the knowledge of genetic determinants and genotype-phenotype relationships in this group of neurological disorders.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session II

PB3205 Evaluation of missing disease-causing variation in autosomal recessive conditions using long-read sequencing

Authors:

K. Patterson, M. Galey, J. Goffena, University of Washington Center for Rare Disease Research, E. Eichler, M. Bamshad, J. Chong, D. Miller; Univ. of Washington, Seattle, WA

Conventional clinical genetic testing fails to provide a precise genetic diagnosis in ~50% of individuals with suspected Mendelian conditions. Individuals without a precise molecular diagnosis undergo more diagnostic tests than those with a diagnosis and are limited in their ability to participate in clinical trials or benefit from N-of-1 therapies. In individuals in which clinical testing identified only one variant in a gene associated with a suspected autosomal recessive (AR) condition, emerging technologies like long-read sequencing (LRS) can provide details about variant types that may have been missed.

We performed LRS of 23 individuals with a suspected AR condition who had either one variant (N=22) or no variants (N=1) identified after clinical evaluation. LRS identified a pathogenic (P) or likely pathogenic (LP) variant in 13/23 cases (53%), a variant of uncertain significance (VUS) in 5/23 cases (22%), and no second candidate variant in 5/23 cases (22%). Among the 18 cases with second variants classified as P/LP or VUS, we identified eight structural variants (SVs) (44% of cases), including single- and multi-exon deletions, transposable element (TE) insertions, and one large intronic deletion. Splice variants were identified in 7/18 (39%) cases, missense variants in two cases, and a promoter variant in one case.

While short-read genome sequencing (srGS) is expected to identify all splice, promoter, and missense variants identified by LRS, only 4/8 SVs would likely be detected due to inherent challenges posed by TE insertions or variant calling in repetitive genomic regions. Because it remains unclear what the next best test is after nondiagnostic clinical testing, we are performing in-silico simulations of the events to more concretely assess their detectability using srGS. Even if srGS can detect the variants found by LRS, the utility of LRS should not be underestimated as its additional benefits include better phasing of distant variants (e.g., when parental samples are unavailable or for de novo variants), the ability to more fully resolve complex SVs, simultaneous acquisition of methylation information, and higher sensitivity in complex genomic regions. These technical advantages can help increase confidence in excluding a candidate gene from consideration, evaluating pathogenicity, or in concluding that a heterozygous candidate variant is likely explanatory due to an unexpected mode of inheritance (e.g., dominant when AR inheritance is typical). Our findings suggest LRS as a potential next best test for recessive missing variant cases. Larger studies directly comparing the sensitivity and specificity of srGS and LRS are needed.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session III

PB3206 Exome reanalysis in a cohort of patients with inborn errors of immunity

Authors:

E. Vorsteveld, C. van der Made, A. Simons, S. Smeekens, Radboudumc multidisciplinary immune-exome team meeting, M. Netea, A. Hoischen; Radboud Univ. Med. Ctr., Nijmegen, Netherlands

Background/Objectives: Inborn errors of immunity (IEI) comprise a group of immune defects that display significant clinical and genetic heterogeneity. Exome sequencing has been fundamental in the understanding of these diseases and has led to a sharp increase in the number of genes associated with these disorders. The diagnostic yield of IEI remains low compared to other rare monogenic disorders. We aimed to improve the diagnostic yield of exome sequencing through a systematic reanalysis.**Methods:** We have collected exome data of 1300 patients with (suspected) IEI that were sequenced and analysed between 2013 and 2021 at the Radboud University Medical Center. We filtered for rare variants in 484 genes comprising the latest *in silico* gene panel for IEI. **Results:** Diagnostic exome sequencing provided a molecular diagnosis in 11.7% in our patient cohort. Through the systematic reanalysis of the exome data, we identified variants in genes added to the gene panel after the initial genetic analysis (n=36), or through addition of novel disease mechanisms to known IEI genes (n=4). Further diagnostic yield was attained through variant reinterpretation based on current insight (n=24), the addition of CNV analysis (n=4), diagnostic whole exome analysis (n=10) or through variant reannotation (n=1). Exome reanalysis yielded 79 additional variants in 74 patients (5.7%), leading to the diagnosis of an additional 24 patients.**Conclusion:** We have performed a systematic reanalysis of exome sequencing data from 1300 patients with IEI. We show additional disease causing variants in 5.7% of patients, increasing the diagnostic yield. There remains continued value in diagnosing individuals with rare (immune) diseases through systematic reanalysis of exome data.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session I

PB3207 Exome sequencing from urinary epithelial cells - exploring the potential role in rare disease discovery

Authors:

K. Ounap[1,2], K. T. Oja[1,2], K. Reinson[1,2], K. Muru[1,2], T. Reimand[1,2], U. Samarina[2], S. Pajusalu[1,2]; [1]Univ. of Tartu, Inst. of Clinical Med., Tartu, Estonia, [2]Tartu Univ. Hosp., Genetic and Personalized Med. Clinic, Dept. of Clinical Genetics, Tartu, Estonia

Increasing the diagnostic rate of rare diseases (RD) is extremely important. Current clinical practice will detect the exact etiology in 30-50% of RD cases. It has been previously shown that some neurodevelopmental disorders, like *SLC35A2* defect, can be caused by somatic mosaicism in brain tissue. Moreover, among mitochondrial disorders, the heteroplasmy of the pathogenic variant is variable in tissues, and urinary epithelial cell analysis corresponds in the best way to the impairment of the brain. In this study, we hypothesized that we could increase the detection rate by performing exome sequencing (ES) from different tissues and analyzing these data together. The initial study group consisted of 116 individuals; in all of them, a reanalysis of ES data was performed using Seqr platform. Among cases that remained negative after ES reanalysis, trio/quad genome sequencing was performed in 56 families, RNA sequencing in 53 individuals, and untargeted metabolome analysis in 51 individuals. We previously identified the disease cause in 70/116 (60%) individuals. In this study, we selected 25/116 (22%) unsolved individuals with neurodevelopmental phenotypes. In all of them, a fresh urinary sample was collected, and the DNA was extracted. The amount of DNA varied extensively (14-197 ng/µl). The new ES analysis was performed, and results were analyzed in quadruple (blood and urinary from the patient and blood from both parents). In the initial quadruple ES analysis, we tagged a *de novo* gene candidate in 16 cases in the urinary sample. Most of the tagged variants were located in low-covered regions of the exome. After thorough consideration, we determined four gene candidates worth validating in a new urinary sample. However, three of these variants were not confirmed in the new sample, and in one case, validation was not possible due to the region's complexity. Post-zygotic mosaicism is one etiological cause which is very difficult to confirm in the routine diagnostic clinical setting. We assumed to specify RD diagnosis by performing ES from different tissue sources. However, we could not prove our hypothesis due to the low DNA concentration extracted from urinary epithelial cells and not persistent findings in different samples. Funding: Estonian Research Council grants PRG471 and PSG774.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session II

PB3208 Exon-level copy number variants detected by CytoScan XON microarray in patients with genetic disorders

Authors:

c. nam, E. Seo, S. Min, B. Lee; Asan Med. Ctr., Dept. of Lab. Med., Univ. of Ulsan Coll. of Med., Seoul, Korea, Republic of, seoul, Korea, Republic of

Background Copy number variants (CNV) have been increasingly linked to various pathologies. Intragenic CNVs spanning one or two exons are more common in Mendelian disorders than previously thought. Although whole exome sequencing (WES), whole genome sequencing (WGS), or chromosomal microarray (CMA) can be used to detect CNVs, a confirmatory test is essential. Multiplex Ligation-dependent Probe Amplification (MLPA) is commonly used for confirmation. However, it requires a control group test and new probes need to be made for candidate genes. We performed CytoScan Xon array for 24 patients to evaluate its clinical performance. **Methods** The twenty four subjects who showed pathogenic CNV in high-resolution chromosomal microarray (CytoScan Dx) and MLPA were analyzed. The subjects had small CNVs which involved single to multiple exon deletions or duplications as follows; six *NF1* deletions, four homozygous *SMN1* deletions, three *DMD* deletions, two *H19-IGF2* duplications, gene deletions of *ANKRD11 and SPG7, COL4A3, CREBBP, GAA, MAOA and MAOB, SHANK3 and ACR, SNTG2, TSC2, VHL.* CytoScan XON array is designed to cover 21,844 genes, including 7,000 clinically relevant genes associated with genetic diseases and OMIM morbid genes. The array combines whole-genome amplified DNA with 6.85 million probes to integrate single-nucleotide polymorphism genotypes and copy number calls. We compared the results with CytoScan Dx or MLPA. **Results** The CytoScan Xon array was able to detect all chromosomal imbalances of 24 patients that were identified by CMA or MLPA. It was able to detect small CNVs as single exon deletion of *DMD* exon 44, *DMD* exon 45, and *MAOA, MAOB* 90 kb loss. Loss of heterozygosity of GAA gene in Pompe's disease was able to be detected. Especially, a muscular dystrophy case was difficult to determine *DMD* exon 51 deletion by CytoScan Dx while CytoScan Xon array clarified exon 51 deletion, as MLPA detected. **Conclusion** Our data shows that the CytoScan Xon array can detect exon level copy number variants more clearly. However, this array detected too many variants to be evaluated, as 100-300 variants per specimen. In conclusion, CytoScan XON array can serve as a complement to mutation analysis performed by WES/WGS, especially for inconclusive result or multiple gene analysis. It can be used in cases with only one heterozygous pathogenic mutation in highly suspicious autosomal recessive genes. Also, it can be used in cases with one homozygous mutation in gene to rule out uniparental disomy, homozygous mutation, and mutation with exon deletion.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session III

PB3209 † Exploring Blindspots in Genetic Diagnosis of Leukodystrophies: Integrating State-of-the-art Methodologies and HiFi Long-read Genome Sequencing for Undiagnosed Pediatric Patients

Authors:

I. Thiffault[1], E. Farrow[1], A. Cohen[1], J. Macintosh[2], T. Zion[1], E. Boillat[1], J. Johnston[1], S. Younger[1], J. Means[1], W. Cheung[1], C. Schwendinger-Schreck[1], S. Perrier[2], T. Truong[3], G. Evrony[3], J. Bonkowsky[4], G. Bernard[2], T. Pastinen[1]; [1]Children's Mercy, Kansas City, MO, [2]McGill Univ., Montreal, QC, Canada, [3]New York Univ. Sch. of Med., New York, NY, [4]Primary Children's Hosp. and the Univ. of Utah, Salt Lake City, UT

Leukodystrophies are a group of heterogenous genetic diseases leading to progressive disabilities and often associated with poor prognostic. Collectively, their incidence is ~ 1:5000 births. Although some subtypes of white matter disease (WM) have a high diagnostic rate by short-read exome (srES), other forms with more complex presentation (neuro+) remain mostly unsolved even by short-read genome sequencing (srGS). Beyond increased diagnostic yield by AI-driven reanalysis, we hypothesize that variations in "srES/srGS blindspots" (repeat expansions, structural variants (SV), intronic/regulatory variants) account for missed molecular diagnoses. Furthermore, HiFi-GS can be utilized for interrogation of non-coding/regulatory variation via methylation detection as well as RNA (HiFi-IsoSeq) and can overcome limitations of reference genomes by *de novo* assembly. We recently demonstrated that up to 3% of functional genomes can be missed by a reference genome as compared to *de novo* assembly. **Methods:** In the context of a collaborative genomic medicine program "Genomic Answers for Kids" with >5000 families with suspected genetic disorders; we have developed one of the largest HiFi-GS data resources by integrating an enhanced HiFi-GS workflow for unsolved pediatric cases. To determine the utility of HiFi-GS in WM diseases, we established a cohort of ~300 patients from three sites: McGill University Health Center, Children's Mercy Kansas City, and Primary Children's Hospital and the University of Utah. All patients entering the program have had srES/srGS performed. Prior to deploying HiFi-GS, existing srES/srGS data of unsolved leukodystrophy cases are systematically reanalyzed by AI-driven methods. **Results:** As previously reported, subtypes of WM (i.e., hypomyelinating, mitochondrial, or metabolic) have a high (70%) diagnostic rate versus neuro+ (22%). In neuro+, blended phenotypes among patients with dual diagnoses are observed, along with a higher rate of gene discovery. For non-trio cases, if VUS(s) detected by srES/srGS are confirmed *de novo* or *in trans* configuration, the genotype would be diagnostic (~20%). Using HiFi-GS, we identified diagnostic genotypes undetected by srES/srGS. We also identified a private hypermethylated "GCC" expansion in a non-OMIM gene. **Discussion:** We are pursuing full genome profiling of unsolved WM cases by leveraging HiFi-GS, *de novo* assembly, methylation, and IsoSeq. Moreover, the findings will be validated in patient-derived cells (iPSCs) to systematically study the impact of RNA or myelination (oligodendrocytes).

Session Title: Molecular and Cytogenetic Diagnostics Poster Session I

PB3210 Extended chromosomes after ICRF-193 treatment in human peripheral blood lymphocytes: An aid in the study of the mitotic chromosome.

Authors:

L. Prouty; Louisiana State Univ. Hlth.Sci. Ctr. Shreveport, Shreveport, LA

The study of the mitotic chromosome has been greatly enhanced by advances in electron and superresolution microscopy. The work presented here combines structured illumination microscopy with modified cytogenetic techniques which lengthen chromosomes in mitotic spreads, for the purpose of making FISH and immunofluorescence targets more accessible and clarifying spatial relationships between imaged structures. Normal peripheral blood lymphocytes are synchronized to select for long mitotic chromosomes with an established high-resolution technique. At harvest, a topoisomerase II alpha inhibitor, ICRF-193, is added to the cultures, markedly slowing chromosome compaction. Chromosomes captured in early prometaphase can be nearly twice the length of chromosomes in conventional high-resolution preparations. After hypotonic treatment of cells, two methods of fixation are used. The first, conventional acid/alcohol (Carnoy's) fixation, is a harsh, protein denaturing fixative, but its action enhances a segmented appearance in individual chromatids. In early prometaphase, rounded chromatin "beads" (mean dia. 260 nm) alternate with more diffuse segments, and, in regions free of stretching artifacts, the segmental structure of sister chromatids is nearly identical, giving rise to the hypothesis that individual segments contain the same DNA in succeeding mitoses. This hypothesis is testable with oligo-FISH probing. The second fixation method is an aqueous one, with detergent release of chromosomes, paraformaldehyde fixation, and "drying down" of spreads on slides. With the preservation of proteins, chromatin fiber structure is improved and segmentation is less pronounced. In addition to DNA probing, proteins are targeted with immunofluorescence technique, broadening the research potential to the many outstanding questions, such as the details of mitotic chromosome structure and compaction, including the positioning of condensins; histone modifications of the same loci in interphase and metaphase; with Click technology, transcription in mitosis; and others. Progress in the analysis of the mitotic chromosome is bound with microscopy, including imaging in live cells, an approach typically confined to relatively compacted chromosomes. Extended, fixed chromosomes in spreads offer a complementary approach and chromosomal configuration.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session II

PB3211 † Facing the challenges of the diagnostic odyssey: diagnostic genome sequencing in 501 patients with rare diseases from Colombia.

Authors:

H. VELASCO[1], A. Bertoli-Avella[2], J. Espinosa[3], D. Cardona[1], A. Gonzalez[1], M. Naranjo[1], J. Valencia[1], C. Buitrago[1], J. Gutierrez[4], J. Marcello[5], P. Bauer[6], C. Jaramillo[3]; [1]SURA, BIOSci. CENTER, MEDELLIN, Colombia, [2]Centogene AG, Rostock, Germany, [3]SURA, GENETIC UNIT, MEDELLIN, Colombia, [4]Centogene, Rostok, Germany, [5]CENTOGENE, Rostok, Germany, [6]CENTOGENE, Rostock, Germany

Exome and genome sequencing (ES GS) are routinely used for the diagnosis of genetic diseases in developed countries. However, their implementation is limited in countries from Latin America, affecting millions of patients living in the continent. In this work, we describe the results of GS in 501 patients from Colombia. GS was performed in the index cases using dried blood spots. Multiomic testing was performed in cases that needed further clarification on the clinical relevance of the identified variants (biomarker, enzymatic activity, RNA-seq). All tests were carried at an accredited genetic laboratory (CENTOGENE, Germany). Ethnicity prediction confirmed that 401 patients were mainly of Amerindian origin. A genetic diagnosis was established for 142 patients with a 28.3% diagnostic yield (101 pathogenic/likely pathogenic variants). The highest diagnostic yield was achieved for pathologies with a metabolic component, followed by syndromic disorders (p. minus 0.001). Patients with genetic syndromes have spent more than 75% of their life without a diagnosis, while for patients with neurologic and neuromuscular diseases, the time of the diagnostic odyssey tended to decrease with age. Previous testing, specifically karyotyping or chromosomal microarray (CMA) were significantly associated with a longer time to reach a definitive diagnosis (p minus 0.01). Furthermore, 1 out of 5 patients that had an ES before could be diagnosed by GS. In conclusion, we present the results of GS in a large cohort of patients from Colombia and demonstrate the value of GS as first tier test for the diagnosis of genetic diseases.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session III

PB3212 Furthering WGS/WES results: clinical implementation of RNA-based functional assessments of coding and intronic variants with potential loss-of-function

Authors:

J-L. Blouin[1,2], A. VANNIER[3], C. MARCONI[1], F. Masclaux[3], M. ABRAMOWICZ[1,2], T. Rio Frio[4]; [1]Univ. Hosp. of Geneva, GENEVA, Switzerland, [2]Univ. of Geneva-Sch. of Med., Geneva, Switzerland, [3]Univ. Hosp. of Geneva, Geneva, Switzerland, [4]Geneva Univ. Hosp., Geneva, Switzerland

Background: The introduction of NGS in clinical practice has improves dramatically the rate of elucidated cases, but has also created a paradoxal increase of variants of unknown significance (VUS). Current ACMG guidelines for variant interpretation focus on amino-acid level identified but filter out other non-coding or synonymous variants that may however have a clinical significance. **Methods**: After selection by an interdisciplinary team of variants pre-filtered on a compound of in silico predictors, fast and affordable molecular investigations on targeted transcript regions were performed on RNA extracted from patients' blood. **Results**: Besides confirmation of predicted impact on splicing for variants localized in critical regions of canonical splicing sites, missense variants of uncertain significance (relative to their impact on protein function) and rare deep intronic variants were shown to critically impact the splicing of genes they lie in. For example, missense VUS in *ATP8B1* and *MYBPC3* genes were found to create stronger donor or acceptor splice site than the canonical one, leading to abnormally spliced mRNAs degraded by nonsense-mediated mRNA decay (NMD). Also, a deep intronic variant localized at position 337 in an intron of 1212 nucleotides in *MYBPC3* gene and other intronic variants IVS-19 in *CHD7*, IVS-18 in *NUP93* and IVS+9 in *NEB* genes were found to create cryptic intronic splice sites leading to abnormally spliced mRNA degraded by NMD or leading to non-functional protein. We also demonstrated the utility of RNA analysis in prenatal diagnosis in the context of ultrasonography abnormalities using RNA from fetal amniocentesis cells for a IVS+5 variant of *FBN1* and a IVS-19 variant in *CHD7* that were upgraded to class 5 and 4, respectively. **Conclusion**: We have translated to diagnostics an analytical pipeline to assessing VUS predicted to impact gene transcript with the aim to decipher their pathogenicity. We demonstrated that combination of *in silico* predictors and such targeted RNA-based analytical approach is able to clarify a number of DNA-based interpretation of variants. It appears to be far-more cost-effective than total RNAseq that needs complex analytical process and would not be necessary in a number of such situatiosn where VUS are identified and could explain the phenotype. All variants, within several genes, analyzed through this pipeline so far had been re-classified, either upward (to pathogenic or to likely pathogenic) or downward (to benign) following the ACMG guidelines. We found also that nearly all assessed genes until now could be analyzed in RNA from blood even when a low expression in this tissue was reported according to GTEx database.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session I

PB3213 GenDiseak: Automated platform for human genetic variant interpretation based on ACMG guidelines.

Authors:

T-H. Yuan[1,2], Y-A. Chen[3], Y-B. Wang[1], J-H. Huang[1], C-Y. Chen[1,2,4], J-J. Hsu[3], P-L. Chen[2,3,5,6,7]; [1]TAIGenomics Co., Ltd., Taipei, Taiwan, [2]Genome and Systems Biology Degree Program, Academia Sinica and Natl. Taiwan Univ., Taipei, Taiwan, [3]Graduate Inst. of Med. Genomics and Proteomics, Coll. of Med., Natl. Taiwan Univ., Taipei, Taiwan, [4]Dept. of Biomechatronics Engineering, Natl. Taiwan Univ., Taipei, Taiwan, [5]Dept. of Med. Genetics, Natl. Taiwan Univ. Hosp., Taipei, Taiwan, [6]Div. of Endocrinology and Metabolism, Dept. of Internal Med., Natl. Taiwan Univ. Hosp., Taipei, Taiwan, [7]Graduate Inst. of Clinical Med., Coll. of Med., Natl. Taiwan Univ., Taipei, Taiwan

The application of Next-Generation Sequencing (NGS) technology in molecular genetic testing has greatly improved the research and diagnosis of genetic disorders in clinical practice. In 2015, the American College of Medical Genetics and Genomics (ACMG) published updated standards and guidelines for the clinical interpretation of sequencing variants associated with human diseases. However, there are challenges in the identification and interpretation of variants when translating high-throughput genetic sequencing into clinical practice. These complications arise due to variability in interpretation among individuals and the lack of standardized interpreting algorithms. To address these issues, our team has developed GenDiseak, an automatic interpretation tool. GenDiseak offers a user-friendly interface and provides users with custom bioinformatic workflows, coupled with automatic variant classification and prioritization based on the 2015 ACMG guidelines for clinical significance. We assessed the performance of GenDiseak using benchmark datasets from the DEAFNESS VARIATION DATABASE (DVD) and BRCA Exchange. Remarkably, GenDiseak achieved an F1 score of 89.98% for the classification of Pathogenic and Likely Pathogenic single nucleotide variants (SNVs) and InDels in the DVD dataset, and an impressive F1 score of 99.96% for the same Pathogenic categories of SNVs and InDels in the BRCA Exchange dataset. Furthermore, GenDiseak integrates pubmedKB, a novel literature understanding tool for extracting semantic relations between biomedical entities. This enables rapid and intelligent search through vast amounts of biomedical literature, providing valuable knowledge and insights. Given its features, GenDiseak has the capacity to assist professionals in variant identification and interpretation, leading to substantial enhancements in efficiency, precision, and the general standard of healthcare and genomic research.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session II

PB3214 Genetic causes of neurodevelopmental disorders.

Authors:

A. Mohamed[1,2], O. Eid[3], A. kamel[2], E. Ashaat[1], G. El Houseiny[1], M. Farid[4], R. Mahrous[1], P. Erian[1], M. Zaki[1], N. Meguid[1]; [1]Natl. Res. Ctr., Cairo, Egypt, [2]Natl. Res. Ctr., Cairo, Egypt, [3]Natl. research centre - Egypt, Giza, Egypt, [4]Natl. Res. Ctr., Cairo, Egypt, Cairo, Egypt

The aim of this research was to detect genetic causes in three groups with NDDs. Autism spectrum disorder (ASD), intellectual disabilities (ID) associated with multiple congenital abnormalities (MCA), and imprinted disorders. This paves the way for personalized medicine in some NDDs. Methods: chromosomal microarray (CMA), multiple ligation probe amplification (MLPA), and whole exome sequencing (WES) were performed. The 1st group included 80 child suffering from ASD, CMA applied on 40 patients showed pathological structural copy number variance (CNV) in 17.5%, this involved mainly chromosomes 15q13,16p11.2, 22q11.2 and 7q11.23.Whole exome sequencing (WES) was performed for 20 ASD patients, they showed frameshift and missense variants associated to ASD in 6 patients, there was one identical twin who had nonsynonymous SNV in AGTR2 gene, which is a strong candidate for ASD, the variant was inherited as X-linked trait from their mother. The 2nd group is the patients with ID associated with MCA (80 patients). The molecular karyotype revealed 44% structural CNV in these patients. The 3rd group is the patients with clinical manifestations of imprinting disorders (80 patients). This cohort revealed 2 patients with Beckwith- Wiedemann syndrome due to hypomethylation of KCNQ-CR, two patients with Silver-Russel syndrome, one due to uniparental disomy (UPD)in chromosome 7 with hypermethylation of MEST-1, the other one due to mutation in the imprinting centre, one patient with Prader-Willi syndrome due to UPD of chromosome 15, and with hypermethylation of SNPRN. This research indicates the great importance of conducting MLPA, CMA and WES for all neurodevelopmental delay patients. These tests must be carried out under the umbrella of health insurance, and this is the only way to give sound genetic advice, proper prenatal diagnosis, personalized medicine with Possibility of gene therapy for NNDs depending on accurate molecular diagnosis.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session III

PB3215 Genetic etiology analysis of 216 patients of infantile epileptic spasm symdrome with unknown cause

Authors:

Z. Pan[1,2], H. Tang[3], C. Chen[3], J. Peng[1,2]; [1]Dept. of Pediatrics, Xiangya Hosp. of Central South Univ., Changsha, China, [2]Clinical Res. Ctr. for Children Neurodevelopmental Disabilities of Hunan Province, Xiangya Hosp., Central South Univ., Changsha, China, [3]Ctr. for Med. Genetics & Hunan Key Lab. of Med. Genetics, Sch. of Life Sci., Central South Univ., Changsha, China

Background: The genetic etiology of Infantile epileptic spasm syndrome (IESS) is an important predictor of prognosis, however, nearly one-third to one-fifth of IESS patients have an unknown cause. Reanalysis of trio whole-exome sequencing (WES) data can improve the genetic diagnostic yield. Whole-genome sequencing (WGS) is believed to increase variant detection, especially non-coding region variants, and transcriptome sequencing (TS) can improve the interpretation of non-coding region variants. **Objectives:** To explore the genetic etiology of patients with IESS of unknown cause and to improve the diagnostic yield by WES reanalysis, and WGS combined TS. To investigate the utility of machine learning in the identification of candidate causative IESS genes. **Methods:** IESS patients meeting inclusion criteria were recruited from the Department of Pediatrics, Xiangya Hospital, Central South University from January 1, 2010 to June 1, 2021. Trio-WES data was pre-processed and re-analyzed according to the virtual epilepsy gene panel. Different epilepsy-causative variant prediction models were built base on GTEx and ClinVar database. An optimal prediction model was selected to identify candidate disease-causative genes from trio-WES data. WGS was performed on WES reanalysis-negative patients. TS were used to verify the functional impact of non-coding region variants from WGS to inform pathogenicity rating. **Results:** Reanalysis of WES data from 17 patients identified disease causative variants, with a positive diagnosis rate of 7.8%. The diagnostic yield of reanalysis after three years was higher than those analyzed at less three years. Forty-one candidate disease causative genes were detected by the random forest variant prediction model. 103 IESS patients with negative WES reanalysis were prospectively enrolled to perform WGS with TS analysis, which identified three patients with pathogenic disease-causative variants, including a *de novo* missense variant of *ATP6V0A1* and complex heterozygous intronic variants of *AFG3L2*, and a de novo missense variant of *MT-ATP6*, resulting in positive diagnosis rate of 2.9%. **Conclusions:** Periodic re-analysis of trio-WES, or trio-WGS with TS can improve genetic diagnostic yield of IESS with unknown cause and provide clues for further identifying candidate IESS-causative or risk gene variants. Further, the prediction model of random forests based on gene-tissue expression feature can be used to identify more candidate IESS-causative genes.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session I

PB3216 Genome sequencing uncovers a 8,8 kb deletion between L1CAM and AVPR2 in a boy with Hirschsprung disease and nephrogenic diabetes insipidus

Authors:

I. Bader, S. Beck-Wödl, K. Buder, P. Stöbe, M. Sturm, O. Riess, T. Haack; Univ.sklinikum Tübingen, Tübingen, Germany

Pathogenic loss-of function variants and missense-variants in AVPR2 (Arginine Vasopressin Receptor 2) are a known cause of X-linked nephrogenic diabetes insipidus. Pathogenic loss-of function variants and missense-variants in L1CAM (L1 Cell Adhesion Molecule) are a known cause of X-linked hydrocephalus with or without Hirschsprung disease. A contiguous-gene-deletion syndrome with a clinical picture that comprises nephrogenic diabetes insipidus and hydrocephalus with or without Hirschsprung disease, has been described in two male patients that harbor deletions (approx. 62 kb and 29 kb respectively) that affect coding regions of both adjacent genes (AVPR2 and L1CAM) (PMID: 18553546, 17318848). Here we describe a 1,5 year old boy with Hirschsprung disease and nephrogenic diabetes insipidus that harbors an 8,8 kb Deletion in Xq28 (NC_000023.11:g.[153883503_153892327del];[0]) located in between the two adjacent genes L1CAM and AVPR2. The deletion affects the 5' UTR exon 1 of L1CAM but no annotated protein coding exons of L1CAM. The deletion terminates 11 kb before the 5'-end of the nearest annotated AVPR2-trancript (NR_027419.2). The deletion does not affect any annotated coding or non-coding transcripts of AVPR2; that is why the deletion was uncovered only by whole genome sequencing and was undetectable in whole exome sequencing data of the patient. The unusual combination of the specific and rare symptoms of the patient raises the hypothesis that this 8,8 kb deletion affects chromosomal material that regulate the expression of both genes. Our data indicate that whole genome sequencing can uncover potentially disease causing CNVs in non protein-coding regions in patients with a specific clinical picture. Further analyses, i.e. RNA-expression of AVPR2 and L1CAM in blood and fibroblasts of the patient need to be performed to demonstrate dysregulation.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session II

PB3217 Going the extra mile for rare diseases: Feasibility and challenges of setting up a pilot Undiagnosed Disease Program (UDP) in Iowa

Authors:

J. Ginter[1], A. Paulson[1], B. Darbro[1], N. Niu[1], J. Nagy[1], H. Major[1], A. Wetzel[1], K. Coleman[2], B. Chandra[1]; [1]Stead Family Dept. of Pediatrics, Iowa City, IA, [2]Holden Comprehensive Cancer Ctr., Iowa City, IA

Undiagnosed rare diseases (URD) remain common despite comprehensive clinical evaluations and advances in next generation sequencing. The proportion of URD after trio exome sequencing (ES) is about 70-85% depending on parameters. As such, genetics services at University of Iowa Hospitals and Clinics (UIHC) have many patients with URD. UIHC houses the only comprehensive genetics center for a population of 3.19 million in Iowa, with 1.27 million living in rural areas. Travel to Undiagnosed Disease Network (UDN) sites has not been possible for some Iowa patients with URD. Through internal philanthropic funding, the Division of Medical Genetics and Genomics at UIHC initiated a pilot Undiagnosed Disease Program (UDP) in January 2023, led by a core team of a neurogeneticist and two genetic counselors (GCs). The UDP was modeled on the UDN framework. Participants with non-diagnostic ES or comprehensive genetic testing are eligible. GCs review the chart to abstract human phenotype ontology (HPO) terms and summarize past studies. These are reviewed by the physician and included in clinic notes as part of deep phenotyping. Appointments are a hybrid of clinical testing, such as global metabolomics, and research testing that insurance is unlikely to cover, like genome sequencing and RNA sequencing and interpretation. Participants provide specimens for the IRB approved UI rare disease biorepository (UIRDB) for research testing. The GCs or physician consent for the UIRDB. The physician conducts a detailed physical and neurological exam for participants. To remain cost effective, we perform research-based genome sequencing in-house through Shivanand R. Patil Cytogenetics and Molecular Laboratory. RNA extraction and sequencing with analysis are performed locally at our center through our collaboration with tissue procurement core at UIHC Holden Comprehensive Cancer Center and the Iowa Institute of Human Genetics. The Chandra laboratory performs induced pluripotent stem cell(iPSC)-based disease modeling. Challenges we experience include lack of a coordinator, no dedicated bioinformatician, and limited access to trio samples. Our UDP service recently hired a 0.5 FTE bioinformatician. Also, iPSC reprograming from peripheral blood had limited success and we are considering skin biopsy in the future. Our program has received 24 referrals through May 2023. Of these, 17 individuals have been evaluated. All have undergone deep phenotyping using HPO terms and genome sequencing. RNA sequencing is in process for 8/17 (47%). In conclusion, we have successfully initiated a UDP service with limited funding in the state of Iowa.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session III

PB3218 High fidelity copy number detection from difficult sample types using an accelerated chromosomal microarray assay.

Authors:

V. Huynh, V. Alford, A. Baker, S. Baral, C. Chen, A. Goyal, J. Law, K. Padamata, N. Vegesna, J. Wallace, B. Bolstad, L. Wu, J. McGirr, D. Black, A. Roter; Thermo Fisher Scientific, Santa Clara, CA

Chromosomal Microarray (CMA) has been widely used for studying genomic imbalances for over a decade and is currently recognized as a first-tier research tool for CNV detection. The application of molecular karyotyping to blood samples is often considered the “gold standard” test, but there is increasing interest and clinical research applications to enable accurate detection of copy number changes in alternative sample types such as oral and prenatal samples. These sample types present multiple challenges around adequate yields of high quality gDNA as well as contamination with extraneous sources of DNA, such as from the oral microbiome for saliva and buccal samples or maternal cell contamination often associated with detection of chromosomal aberrations in prenatal samples. To support robust detection of copy number changes, we built a diverse Reference Model by incorporating over five hundred samples of multiple tissue types (blood, buccal, saliva, amniocentesis, chorionic villi, and POC) to enable genome-wide detection of copy number variation on the CytoScan HD Accel microarray. Principal Component Analysis (PCA) illustrates that most, but not all, genomic DNAs are equal. Specifically, buccal and saliva sample types cluster separately from the other sample types. These sample batch effects can be corrected by incorporating buccal and saliva samples into the copy number reference model. The copy number reference model contains the expected log2 signal intensity for each probe at the diploid state and enables detection of copy number variants. A large reference model built from a variety of sample types and under a variety of processing conditions captures normal variation of process and sample type and is robust to these variations. Furthermore, we demonstrate passing array QC metrics with up to 25% maternal cell contamination and ability to detect copy number changes associated with a trisomy with up to 30% maternal cell contamination using a sample mixing cell-line model.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session I

PB3219 High-resolution karyotype reveals a maternally inherited complex chromosome rearrangement involving chromosomes 4p, 5p, and 12q with an inverted insertion and a 12q21.31q21.32 loss identified by SNP array.

Authors:

V. Stinnett[1], J. Biscoe[1], N. DeMetrick[1], J. Murry[2]; [1]Johns Hopkins Hosp., Baltimore, MD, [2]Johns Hopkins Sch. of Med., Baltimore, MD

Background: Complex chromosome rearrangements (CCRs) are defined as rearrangements involving three or more chromosomes or as having more than two breakpoints. CCRs can be further classified into subgroups based on their complexity. Unbalanced CCRs are often associated with phenotypic findings. Here we present lab findings for a 19-year-old patient referred for autoimmune hemolytic anemia. An outside karyotype performed on a bone marrow specimen identified an apparently balanced novel translocation between chromosomes 4 and 12 present in all cells examined, implicating a constitutional aberration.
Results: A karyotype performed on cultured fibroblast cells confirmed a germline origin and revealed chromosomal aberrations for 4p14, 5p15.1, and 12q14, raising a high suspicion for a rearrangement involving an insertion of 12q14q21 material into the 5p15.1 locus. Metaphase FISH performed using subtelomeric probes for all three chromosomes revealed a reciprocal translocation between 4p and 12q. A normal hybridization pattern for chromosome 5 supported an interstitial insertion of 12q material into 5p, as suggested by karyotype. The only genomic imbalance identified by SNP microarray (Illumina® 850k BeadChip) was a 234 kilobase (kb) interstitial loss of 12q21.31q21.32, classified as a variant of uncertain significance. A high-resolution maternal karyotype and SNP array performed on peripheral blood showed the same results confirming the CCR in our patient was maternally inherited.
Discussion: The high-resolution maternal karyotype clarified the cytobands and rearrangement orientation by identifying that insertion of 12q13.3q21.31 material at chromosome 5p14.3 is inverted. This report highlights the continued utility of high-resolution (greater than 550 band level) karyotypes in assessing CCRs. It is possible the small deletion on chromosome 12 occurred secondary to the CCR formation. As other small deletions or gene disruptions from the CCR cannot be excluded, emerging laboratory methods may assess this possibility. Here, we confirm prior literature reports that inherited CCRs are often maternal in origin. As carriers of CCR are at an increased risk of abnormal pregnancy outcome and infertility, genetic counseling is recommended to discuss recurrence risk. Other family relatives may be carriers of this CCR and should be evaluated using complementary cytogenomic methods.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session II

PB3220 HLA typing using capture-based next-generation sequencing

Authors:

S-K. Lai[1,2], M-J. Lin[3], C-S. Chen[4], P-L. Chen[1,5]; [1]Natl. Taiwan Univ. Hosp., Taipei, Taiwan, [2]Graduate Inst. of Genome and Systems Biology Degree Program, Taipei, Taiwan, [3]Dept. of Computer Sci., Johns Hopkins Univ., Baltimore, MD, [4]TBG Biotechnology Corp., Taipei, Taiwan, [5]Graduate Inst. of Med. Genomics and Proteomics, Taipei, Taiwan

Accurate and reliable HLA typing is crucial in various immunological applications, such as disease diagnosis and transplantation compatibility assessment. However, achieving precise and reliable HLA typing remains challenging due to the complexity and extensive polymorphism of HLA genes. In this study, we aimed to develop a set of capture probes and evaluate their performance using four distinct software programs: HISAT-genotype, HLA-VBseq, T1K, and QZtype. To accomplish this, we designed and utilized HLA capture probes specifically targeting our desired HLA loci and conducted NGS sequencing on a set of 38 samples from Human Pangenome Reference Consortium (HPRC). Subsequently, the generated NGS data were subjected to HLA typing analysis using the aforementioned software programs. In order to establish a benchmark, we aligned the HLA allele sequences obtained from the IPD-IMGT/HLA database to the HPRC publicly available whole-genome assemblies. The HLA alleles exhibiting the least sequence variations were selected as the HLA_annotate benchmark. Comparative analysis showed different performance characteristics among the software programs. T1K achieved a perfect concordance rate of 100% for HLA class I typing. For HLA class II typing, T1K demonstrated concordance rates ranging from 96% to 100% except for the challenging HLA-DRB1 region, where it achieved an 86% concordance rate. QZtype and HISAT-genotype exhibited comparable accuracy, with minor discrepancies compared to T1K. Inconsistencies with the HLA_annotate benchmark were attributed to intronic repeat regions and variations in exon weighting criteria used by different software programs. In conclusion, this study provides insights into the performance of HISAT-genotype, HLA-VBseq, T1K, and QZtype software programs for HLA typing. The designed HLA capture probes were effective, and no loss of heterozygosity was observed during typing. This approach has potential for precise and comprehensive HLA typing in research and clinical applications, improving immunological analyses and transplantation procedures.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session III

PB3221 Host and microbial determinants of cutaneous lesions detected simultaneously via RNA-Sequencing: Biallelic pathogenic CORO1A sequence variant, Mycobacterium tuberculosis, and Staphylococcus aureus Identification.

Authors:

L. Youssefian[1], F. Palizban[2], A. Saeidian[3], H. Hakonarson[4], E. Jouanguy[5], J-L. Casanova[6], H. Vahidnezhad[7]; [1]UCLA, Los Angeles, CA, [2]IBB, Tehran, Iran, Islamic Republic of, [3]Children s Hosp. of Philadelphia, Philadelphia, PA, [4]Children's Hosp. of Philadelphia, Philadelphia, PA, [5]Lab. of Human Genetics of Infectious Diseases, Necker Branch, INSERM U1163, Necker Hosp. for Sick Children,, Paris, France, [6]The Rockefeller Univ., New York, NY, [7]Univ. of Pennsylvania, Philadelphia, PA

Biallelic mutations in the CORO1A gene result in severe combined immunodeficiency (SCID) and susceptibilities to multiple infections, including granulomatous tuberculoid leprosy, epidermodysplasia verruciformis, human herpesvirus 4 (HHV-4), molluscum contagiosum, mucosal HHV-1, HHV-3, HHV-8, and visceral leishmaniasis. We report a novel association between COROA1 homozygosity, late-onset cutaneous tuberculosis (CTB), and Staphylococcus aureus infection. The immunophenotype of SCID was supported by clinical and immunological evaluations, including fluorescence-activated cell sorting and lymphocyte transformation tests. With a novel bioinformatic pipeline, we utilized whole-transcriptome sequencing (RNA-Seq) to concomitantly detect pathogenic viral and bacterial agents, the causal human pathogenic variant, and the consequences of detected sequence variants in the host. The proband, a 17-year-old male, was found to be homozygous for *CORO1A*: NM_007074:exon3:c.248_249del: p.P83Rfs*10, this variant was located inside the 26 Mb of runs of homozygosity on 16p11. RNA-Seq confirmed the deleteriousness of the *CORO1A* variant in four skin biopsies, revealing significant downregulation of *CORO1A*. Reads unaligned to the human genome were applied to a catalog of 926 different viruses and about 1500 pathogenic bacteria. β-HPV5, Mycobacterium tuberculosis (TB) H37Rv, and Staphylococcus aureus were detected. Collectively, we describe a previously unrecognized inborn error of immunity to CTB, indicating that it is autosomal recessive. Whole-transcriptome sequencing allows for the simultaneous detection of host and microbial determinants of widespread cutaneous lesions. A CORO1A deficiency can underlie EV and CTB. In addition, we innovatively harnessed the RNASeq data to detect abnormal microbiota underlying recalcitrant skin lesions.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session I

PB3222 Identification of Contiguous Gene Deletions of *SLC6A8*, *BCAP31*, and *ABCD1* in a Korean Infant Presenting with Hypotonia, Failure to Thrive, Hearing Loss, and Hepatitis

Authors:

Y-M. Kim[1], E. Kim[1], Y. Kwon[1], H. So[1], B. Lee[1], M. Kim[1], J. Min[1], Y. Lee[1], B. Kim[1], G. Seo[2], G-H. Kim[3], B. Lee[3], H-W. Yoo[4,3]; [1]Chungnam Natl. Univ. Sejong Hosp., Sejong, Korea, Republic of, [2]3billion, Seoul, Korea, Republic of, [3]Asan Med. Ctr., Seoul, Korea, Republic of, [4]CHA Bundang Med. Ctr., Seongnam-si, Gyeonggi-do, Korea, Republic of

Contiguous gene deletions involving *SLC6A8*, *BCAP31*, and *ABCD1* are associated with X-linked cerebral creatine deficiency syndrome-1 (MIM#300352), Deafness, dystonia, and cerebral hypomyelination (DDCH, MIM#300475), and X-linked adrenoleukodystrophy (MIM#300100), respectively. These deletions can result in severe clinical phenotypes, including significant failure to thrive, hypotonia, developmental delay, hepatic dysfunction, and hearing loss. We report the case of an 11-month-old boy presenting with growth failure, congenital deafness, muscle hypotonia, and hepatic dysfunction caused by de novo contiguous gene deletions of *SLC6A8*, *BCAP31*, and *ABCD1*. The patient, born to non-consanguineous parents at 40 weeks of gestation with a birth weight of 3.0 kg, exhibited failure to thrive and hypotonia. Congenital hearing loss and persistent elevation of liver enzymes were observed. The patient's height, weight, and head circumference were below the 3rd percentile. The laboratory findings showed persistent elevation of liver enzymes (ALT 81-138 U/L). At 3 months of age, hepatitis workup revealed positive urine and blood cytomegalovirus (CMV) PCR. Congenital CMV infection was diagnosed, and a hearing aid was provided. Metabolic evaluation, except for mild elevation of very long-chain fatty acids (VLCFA), was normal, and *ABCD1* sequencing was unremarkable. Brain magnetic resonance imaging (MRI) revealed a 2.3 cm arachnoid cyst in the left side of the cisternal magna and delayed myelination. Chromosome microarray, methylation tests for Prader-Willi syndrome, mitochondrial sequencing, and whole exome sequencing yielded negative results. At 22 months of age, the patient's development had halted at 10 months. Reanalysis of whole exome sequencing identified an 82.6 kb deletion at the Xq28 region, encompassing *SLC6A8*, *BCAP31*, and *ABCD1*. Mild elevation of VLCFA (C24/C22 1.642, C26/C22 0.045) was observed. Multiplex ligation-dependent probe amplification confirmed the deletions of *ABCD1*, *BCAP31*, and exons 12-14 of the *SLC6A8* gene. The mother did not have the same deletion. MR spectroscopy was consistent with cerebral creatine deficiency syndrome. Treatment with creatine monohydrate (200 mg/kg/day), glycine (150 mg/kg/day), and arginine (400 mg/kg/day) was initiated, resulting in the patient achieving independent walking after 3 months of treatment. Over 2 years, brain MRI showed no interval changes. In conclusion, we present a case of infancy with global developmental delay, failure to thrive, hepatic dysfunction, and hearing loss caused by rare contiguous gene deletions involving *SLC6A8*, *BCAP31*, and *ABCD1*.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session II

PB3223 Implementation of clinical genome sequencing-based panel testing: Experience and results from over 400 patients with suspected inherited cardiovascular diseases

Authors:

C. Hale[1], J. Platt[2,3], K. Dumas[1], D. Fisk[1], M. E. Grove[4], H. Ison[2], M. Kawana[2,5], A. Klinder[2], J. W. Knowles[2,5], J. N. Kohler[2], T. Moscarello[2], R. Murtha[2], T. Nguyen Dolphyn[1], M. Pariani[2], V. N. Parikh[2,5], M. V. Perez[2,5], C. Reavey[1], C. M. Reuter[2], K. Sallam[2,5], E. E. Smith[1], K. S. Weymouth[1], M. T. Wheeler[2,5], S. White[6], E. Spiteri[1,7], W. Qiao[1,7], Y. Yang[1,7], E. A. Ashley[2,5], S. A. Scott[1,7]; [1]Clinical Genomics Lab., Stanford Med., Palo Alto, CA, [2]Stanford Ctr. for Inherited Cardiovascular Disease, Stanford, CA, [3]Dept. of Pediatrics, Div. of Med. Genetics, Stanford, CA, [4]Color Hlth., Burlingame, CA, [5]Dept. of Med., Div. of Cardiovascular Med., Stanford, CA, [6]Fabric Genomics, Oakland, CA, [7]Dept. of Pathology, Stanford, CA

Although high-throughput sequencing is widely implemented for enrichment-based multigene panel diagnostic testing, enrichment-based exome sequencing is increasingly employed as a single platform for clinical panel testing. However, given that genome sequencing has less bias and more consistent coverage across coding regions and improved capacity to detect copy number variants (CNVs) compared to enrichment-based sequencing, in 2021 we validated and launched six genome-based diagnostic panels for inherited cardiovascular diseases [combined cardiomyopathy and arrhythmia (CMA), comprehensive cardiomyopathy (CCM), hypertrophic cardiomyopathy (HCM), aortopathy/familial thoracic aortic aneurysm/dissection (FTAAD), familial hypercholesterolemia, and comprehensive arrhythmia (CA)]. Five genome-based panels (CMA, CCM, HCM, FTAAD, and CA) include optional limited evidence genes for inclusion, and in total, 194 genes and 102 non-coding variants are interrogated across all six panels. To date, 431 genome-based panels (≥40X coverage) have been resulted, with the majority being CMA (73%) and CCM (14%). Outpatients and inpatients represented 69% and 31%, respectively, and most patients (81%) had not previously had prior genetic testing. The majority of patients were adults (97%) and the overall diagnostic rate across all six panels was 17%, which likely was influenced by age distribution as 76% of patients were ≥40 years of age. Inpatients had a higher diagnostic rate than outpatients (20% vs 15%; $p<0.0001$), and among patients that previously had a negative panel result (n=74), 3% had a new positive result by genome-based CMA panel testing, including pathogenic variants in the autosomal dominant *DSP* and *MAP2K1* genes. Among all patients subjected to genome-based panel testing, two patients had positive findings in two independent disease genes, and the spectrum of all reported diagnostic variants included frameshift (30%), missense (23%; including one mosaic), nonsense (19%), in-frame deletions (7%), and CNVs (6%). Importantly, genome-based panel testing across all patients reported 13 CNVs (3%; eight uncertain significance, two likely pathogenic, and three pathogenic), which included the ability to precisely identify most breakpoints. These results underscore the feasibility and utility of implementing genome-based diagnostic panel testing for Mendelian diseases, including improved coverage and accuracy, flexibility in content and design, CNV breakpoint interrogation, mosaic variant detection, a single laboratory workflow, and the opportunity to subsequently leverage non-diagnostic genome content for translational research.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session III

PB3224 Improved technical noise correction in integrated functional genomics datasets with application to pathogenic variant discovery in rare diseases

Authors:

C. Smail[1], R. L. Biswell[1], B. Koseva[1], J. C. Means[1], E. Grundberg[2], S. Younger[1], T. Pastinen[3]; [1]Childrens Mercy Kansas City, Kansas City, MO, [2]Childrens Mercy Res. Inst., Kansas City, MO, [3]Childrens Mercy Hosp. and Clinics, Kansas City, MO

The use of functional genomics resources (such as RNA-sequencing) has emerged as a powerful tool in the diagnosis of rare diseases where DNA-based methods are unsuccessful. This approach is used to measure the downstream molecular consequences of candidate pathogenic variants, revealed using outlier detection methods. Identification of outlier signals from rare disease patient cohorts requires a background distribution of gene function contributed from large-scale healthy/control data. This reference data is often sourced from external sources and the detection and removal of technical/non-biological noise is first required. However, a key limitation of standard noise correction methods is in the handling of non-linear effects when combining samples from imbalanced datasets, where rare disease samples are vastly outnumbered by controls. This creates a paradox central to the research described here; namely, large-scale control data should increase power for uncovering pathogenic functional profiles among rare disease samples, but in fact can substantially increase noise as statistical modeling is biased by the larger cohort. To enable accurate detection and removal of technical noise in imbalanced datasets, we developed a statistical method called SplineSeq. SplineSeq robustly models both linear and non-linear covariates using regression splines, removing confounding technical noise but retaining important biological outlier signal. Applied to a large-scale RNA-sequencing rare disease cohort, we observed that SplineSeq significantly improved technical noise reduction across 20% of all genes (mean increase in model fit (R2) = 4.8%). We applied SplineSeq to a patient case of affected siblings enrolled in the NIH Undiagnosed Diseases Network (UDN) presenting with microcephaly and quadriplegia, observing a 66% improvement in outlier rank in candidate gene *RARS2*. Applied to an additional UDN patient case presenting with diverse clinical phenotypes (including global developmental delay, acute metabolic decompensation, hyperketotic hypoglycemia, rhabdomyolysis) and subsequent intractable candidate gene list, we identified an under expression outlier event in the gene *TANGO2*. Repeating both analyses using standard noise correction approaches failed to identify any clear outlier signal in both cases. SplineSeq is applicable to multiple functional genomics modalities, including short-read Illumina RNA-sequencing and long-read PacBio Iso-Seq.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session I

PB3225 Improving the genetic diagnosis of patients with Congenital Hypogonadotropic Hypogonadism through the integration of genomics and transcriptomics data.

Authors:

A. Boizot[1,2], J. ZHAI[2,1], Y. Zouaghi[1,2], F. De Azevedo Correa[1,2], J. Meylan[1,2], C. Xu[1,2], L. Bartoloni[1,2], M. Adamo[1], G. Papadakis[1], A. Jacobs[1], N. Vionnet[1,2], N. Pitteloud[1,2], F. Santoni[1,2]; [1]Univ. Hosp. of Lausanne CHUV, Lausanne, Switzerland, [2]Univ. of Lausanne, Lausanne, Switzerland

Congenital Hypogonadotropic Hypogonadism (CHH) is a complex genetic disorder causing impaired gonadal development and function and absent puberty. More than 50 genes are nowadays considered to be causative for this disease, but our current knowledge of its etiology is incomplete. In general, more than 50% of CHH cases remain without genetic diagnosis.
To improve the diagnostic outcomes, we investigated the blood transcriptome of 40 CHH patients as a complementary approach of the standard genetic testing based on WES/WGS.
We developed a pipeline based on STAR integrated with custom designed modules to detect introns retentions caused by specific splicing variants. First, we utilized our pipeline to analyze the known CHH genes expressed in the blood.We found one CHH patient with a heterozygous synonymous variant in *CHD7*:c.2442G>A (p.Gln814=). This variant is not present in population databases, it is benign for the splicing predictors and it is classified as VUS by ACMG criteria. Our pipeline discovered that the variant creates an unannotated exon junction consequent to the skipping of exon 6, as frequent in reads as the normal junction. Sanger sequencing confirmed the splicing alteration. Subsequently, we decided to extend this analysis to the entire transcriptome.
An intriguing finding from this analysis was the observation that, on top of unannotated exon-exon junctions and intron retentions, several intra-exonic deletions were not detected by the standard genomic analysis but captured through our transcriptomics based approach. These deletions are mostly, but not only, located in repeated exonic regions and the reads covering the junctions (split-reads) are either unmapped or partially mapped by state-of-the-art DNA aligners (soft clipping). This suggest that the current clinical WES/WGS pipelines have a limited ability to align split-reads and an RNA-based complementary analysis is necessary not only to discover clinically relevant variants in known genes but also for the identification of novel causative genes. A clear disadvantage is the tissue-specificity of gene expression combined with the limited availability of human tissues that practically restrict this complementary approach to genes expressed in blood.
In conclusion, we have developed a comprehensive strategy to detect potentially clinically relevant variants that are undetectable or overlooked by standard clinical or research pipelines. By combining standard WES/WGS processing with the identification of aberrant junctions, intron retentions and exonic deletions, our pipeline improves the diagnostics yield in our cohort of CHH patients.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session II

PB3226 In search of "hidden" variants in craniofacial microsomia: benchmarking ultra-low frequency mosaic variants in deep exome sequencing

Authors:

N. Parmalee[1], C. T. Myers[2], P. Cliften[3], C. Heike[4], D. M. Jensen[1], M-H. M. Pham[1], N. Au[5], K. A. Loy[6], J. Perkins[1], J. T. Bennett[1]; [1]Seattle Children's Res. Inst., Seattle, WA, [2]Seattle Children's Hosp., Seattle, WA, [3]McDonnell Genome Inst., Washington Univ. Sch. of Med., St. Louis, MO, [4]Univ. of Washington, Seattle, WA, [5]Seattle Children s Res. Inst., Seattle, WA, [6]Univ. of Washington Sch. of Med., Seattle, WA

Structural birth defects (SBD) affect 3-6% of live births in the United States, accounting for >20% of deaths in the first year of life. Although genetic factors have been identified, the etiology of most isolated SBD remains unknown. Explanations could include 1) low variant allele fraction (VAF) that is missed by standard sequencing, or 2) absence of the variant in blood/saliva. The goal of our HiddenSeq project is to identify mosaic variants missed by conventional sequencing by performing ultra-deep (>2500X) exome sequencing of affected tissue from a patient with craniofacial microsomia (CFM), a common SBD.

To establish precision and recall of our ultra-low (<1% VAF) variant detection approach, we made two dilutions of HapMap samples (NA18517 into NA12878) to create pseudo-mosaic variants with VAFs of 0.5%, 1%, 2.5%, and 5%. These two benchmark samples were sequenced to >2500X coverage, aligned to hg38 and variants called using Dragen and MosaicHunter. Error correction using dual UMIs was performed. Truthsets were established by standard 100X exome coverage and called on the same pipeline. We defined the truthset as the 16,093 SNP and indel variants present in NA18517 that were not present in NA12878. We also sequenced DNA from affected tissue (pre-auricular skin tag) and blood from an individual with CFM and both parents, to identify variants present only in CFM affected tissue.

We detected 93% (n=15,012) of the truth set variants at low VAFs (5 and 2.5%), and 73% (n=11,698) of the truth set variants at ultra-low VAFs (1 and 0.5%.). 2,208 false positive variants were called in the low VAF dilution and 2,096 false positive variants were called in the ultra-low dilution, representing a false discovery rate of 12.8% and 15.2%, respectively. Application of filtering based on summary statistics of true and false positive reduced the number of false positives by 46% while removing 2-8% of true positives. Application of this pipeline to the ~25K variants called in the CFM samples, identified 611 variants private to CFM tissue. Applying the same filters from our benchmarking samples resulted in 303 variants of interest private to preauricular tag tissue. Manual inspection of these variants revealed >90% to be artifacts. ~25 candidates did not appear artifactual and await orthogonal validation.

Calling mosaic variants exome-wide from bulk tissues at VAFs nearing 1% or lower remains challenging. Ongoing work includes examining other callers and comparing alignment to the T2T CHM13 reference. Results of this work will have direct relevance to the broader field of calling somatic mutations genome-wide, as a part of the NIH Somatic Mosaicism across Human Tissues (SMaHT) Consortium.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session III

PB3227 Increasing clinical throughput: Automated DNA extraction of cultured cells.

Authors:

A. C. W. Brand, MS, A. S. Berlyoung, PhD, D. A. Johnson, MS, S. C. Lynch, J. Noya, MS, J. Spangler, MS, S. Yusuff, PhD, J. M. Devaney, PhD; GeneDx, Gaithersburg, MD

In genetic testing, quickly producing accurate and reliable data is vital to helping clinical providers communicate test results with patients. This process begins with DNA extraction, which is used for primary downstream testing as well as secondary confirmations of potential molecular variants of interest. Depending on sample type, manual DNA extraction using salting-out pellet precipitation can take 24 to 72 hours due to the length of the protocol and pellet rehydration times. We describe the implementation of an automated KingFisher Flex DNA extraction method for direct and cultured cell types at GeneDx that (1) reduced average DNA extraction time to three hours, (2) eliminated viscosity issues that exist with manual processes, (3) stabilized DNA concentrations post-extraction with the utilization of inverse magnetic particle processing, and (4) maintained downstream success for a wide range of clinical assays. 476 cell pellets were cultured and extracted, 90% of which passed GeneDx extracted DNA quality standards (≥10ng/μL). Compared to original specimen data, average DNA yield and nucleic acid purity slightly decreased using the KingFisher method; however, the average extraction time decreased by 87.5% (from 24 hours to 3 hours). Samples were not viscous, and concentration changed an average of 0.1 ng over 14 days. For downstream analysis, a subset of 82 DNAs were selected with the following cell types: One direct amniotic fluid, 28 cultured skin punches, 48 amniocytes, and five cultured chorionic villi. Extracted DNAs were distributed among 24 tests most common to GeneDx direct and cultured cell samples. Testing included Next-Generation and Sanger sequencing, Multiplex Ligation-dependent Probe Amplification, analysis of repeat expansions, arrays for the detection of copy number variants/regions of homozygosity, and maternal cell contamination. Datasets were generated for each test method comparing original specimen data to validation results, and no statistical difference ($p<0.05$) in data quality was observed. Downstream quality of test results was consistent with the manual extraction method, supporting automated KingFisher Flex DNA extraction of direct and cultured cell sample types as an efficient alternative to manual salting-out precipitation in high throughput clinical settings.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session I

PB3228 Insights from Cohesin Complex and Beyond: A single-center experience for cohesinopathy-related/like disorders

Authors:

J. Alaimo[1], V. Paolillo[1], F. Del Viso[2], A. Cohen[1], E. Farrow[1], S. Starling[1], C. Schwager[1], L. Bartik[3], K. Engleman[1], L. Cross[1], H. Welsh[1], E. Rush[1], S. M. Amudhavalli[1], M. Baig,[1], B. Sullivan[1], D. Zhou[1], E. Shaffer[1], T. Zion[1], E. Boillat[1], T. Pastinen[1], C. J. Saunders[1], I. Thiffault[1]; [1]Children's Mercy, Kansas City, MO, [2]Children's Mercy, kansas City, MO, [3]Children's Mercy, Kansa City, MO

Introduction: Cornelia de Lange syndrome (CdLS) is a well-characterized developmental disorder that is associated with genes encoding the cohesin complex. While a significant proportion is attributed to *NIPBL*, other components of the cohesin complex, combined with broader sequencing testing have led to the identification of individuals who have milder or atypical CdLS features. Moreover, CdLS-like phenotypes have been associated with defects in chromatin-associated proteins, such as *KMT2A, AFF4, EP300, TAF6, SETD5, SMARCB1, MAU2, ZMYND11, MED13L, PHIP, ARID1B, NAA10, BRD4* or *ANKRD11*, most of which have limited or no association with CdLS. Our study sought to identify cases for CdLS genotype/phenotype expansion along with the identification of novel genes. **Methods:** We queried our clinical genetics laboratory and research program "Genomic Answers for Kids" at Children's Mercy, Kansas City, for potential CdLS patients. This retrospective study included >5300 families who received genomic testing. A diagnosis of CdLS was established in a proband with heterozygous pathogenic variants in *NIPBL, RAD21, SMC3, or BRD4*, or hemizygous pathogenic variants in *HDAC8 or SMC1A*. We further extended the analysis to other components and regulators of the cohesin complex, which included many genes of unknown significance (GUS). The medical records were retrospectively reviewed for clinical and phenotypic assessment. **Results:** A total of 51 patients harbored private variants in the selected genes (absent in public datasets): 26 (51%) were *de novo* in origin (when a trio was performed), 35 (68%) had a diagnostic genotype, of which 15 (43%) were in "classic" CdLS genes. The proportion of cohesinopathies attributed to pathogenic variants in established CdLS genes was similar to the reported incidence. Highly suspicious variants of uncertain significance (VUS) were detected in 4 patients with "classic" CdLS. Ten individuals (20%) had deleterious variants in other components/regulators of the cohesin complex, not previously associated with disease. Moreover, we identified two patients, including one with classic CdLS features, with *JADE2* deficiency. Genotype/phenotype assessment suggests that additional factors are contributing to the phenotypic variability and severity. **Conclusion:** In conclusion, our study has successfully broadened our understanding of the genetic and phenotypic diversity observed in CdLS-like disorders. Our findings emphasize the importance of conducting pathway analysis to identify potential causal variants beyond the established CdLS genes. Furthermore, our results support the emerging concept of CdLS-related/like disorders**.**

Session Title: Molecular and Cytogenetic Diagnostics Poster Session II

PB3229 In-silico genotyping for Gaucher Disease using Gauchian: The good, the bad, the complex.

Authors:

J. Lichtenberg, E. Hertz, E. Sidransky, N. Tayebi; NIH, Bethesda, MD

Biallelic mutations in *GBA1* result in Gaucher disease (GD), the inherited deficiency of the lysosomal enzyme glucocerebrosidase. Variants in *GBA1* are also the most common genetic risk factor for Parkinson disease (PD). While multiple centers screen for mutant *GBA1* alleles when genotyping patients with parkinsonism, detecting of pathogenic variants, including recombinant alleles resulting from crossover events between *GBA1* and its homologous nearby pseudogene *GBAP1*, remains a challenge.

Recently, a new software tool named Gauchian was introduced to identify *GBA1* variants in whole genome sequencing studies and to detect recombinant alleles. It has been integrated into automatic diagnostic pipelines such as Illumina's Dragen4 to facilitate high-throughput detection of *GBA1* mutations in large cohorts. Using sequencing read depth across a 10kb region between *GBA1* and *GBAP1*, Gauchian employs a Gaussian mixture model to call copy number variants, with losses and gains presumed to represent potentially pathogenic fusion genes consisting of *GBA1* and *GBAP1* fragments. A database of a third of the known *GBA1* mutations is used to classify the cases further.

While Gauchian correctly classified the majority of Gaucher patients in a cohort of 96 individuals previously characterized by Sanger sequencing, it often failed to identify rare mutations and was unable to characterize de novo mutations. By relying on the identification of copy number changes in the intergenic region, the software runs the risk of misreporting the correct number of gene and fusion gene fragments in recombinant alleles. While Gauchian provides a step towards confident in-silico genotyping (sensitivity: 0.9, specificity 1.0), it still reports false positives and negatives that preclude its use in high-throughput variant screening in diagnostic laboratories. Another identified concern stems from an analysis of data from the 1000 Genomes project. Gauchian identified 28 cases out of 1435 individuals with *GBA1* variants. Most of the identified genotypes predicted severe GD that would present in infancy or childhood. suggesting that these were false positive cases.

Because of the significant number of misannotations, future studies should specifically compare the results of Sanger sequencing and copy number analyses with Gauchian predictions in a larger cohort including individuals known to carry complex alleles. While Gauchian may have value in a research setting, we recommend that prior to clinical use, all results be validated with Sanger sequencing and that Gauchian should not be used as a blackbox in larger diagnostic genotyping pipelines.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session III

PB3230 Integrating machine learning tools into a cloud based NGS platform to facilitate molecular diagnosis of Mendelian disorders.

Authors:

X. Wang[1], H. Yu[2], T. Zhuang[1], W. Xu[2], Y. Yang[1]; [1]AiLife Diagnostics, Pearland, TX, [2]Shandong Normal Univ., Jinan, China

Genome-wide sequencing has been growing rapidly in recent years, driven by advancements in sequencing technology and increasing demand for personalized medicine. In the meantime, clinical laboratories are overwhelmed with huge amounts of raw data but lack comprehensive tools for accurate and efficient clinical interpretations of NGS data. We have developed a cloud based NGS platform (AilisNGS®) with machine learning (ML) features that incorporates all components needed for clinical NGS, including LIS, HPO auto generation by natural language processing (NLP), knowledgebase buildup, secondary data analysis, data annotation, ML supported data interpretation, and clinical reporting.

The ML model for variant prioritization in AilisNGS® was established using a cohort of 657 symptomatic individuals previously referred to our lab for clinical exome sequencing. The exome data was initially processed in the AilisNGS® pipeline for secondary analysis and variant annotation. Clinical interpretation and reporting had been performed using rule based and manual methods according to our SOPs. About 26% of the cases were found to be positive cases with definitively contributing variants identified. For ML, fully annotated variants based on multiple public and internal databases, together with NLP generated HPO terms were fed into a multivariate regression model for training on variant ranking. The trained ML model was then applied to sets of testing data for evaluations. Cross validation data from the 657 samples (2/3 training and 1/3 testing) showed that an average of 94% and 98% of the positive variants rank within the top 5 and top 10 lists of the ML generated candidate variants respectively. While manual variant review can be time consuming, it takes only 36 seconds for the ML algorithm to perform training and testing for the 657 samples. The ML feature is being incorporated into our AilisNGS® pipeline for continuous and automatic learning and optimization of the ML model.

Our study showed that ML based models trained from well-annotated genomic data can accurately prioritize variants and enable efficient and accurate genomic data interpretation and molecular diagnosis. It should be noted that although ML can be a useful tool in NGS data interpretation, final manual review by trained professionals is still needed for clinical NGS cases at the current stage of ML implementation.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session I

PB3231 Integration of genomic and transcriptomic data for rare disease diagnosis.

Authors:

J. Major-Mincer[1], B. Moore[1], L. Botto[1], R. Mao[2], P. Bayrak-toydemir[2], A. Quinlan[1]; [1]Univ. of Utah, Salt Lake City, UT, [2]ARUP Lab., Salt Lake City, UT

Millions of individuals worldwide are affected by a rare disease. Typically, individuals with a rare disease will undergo a variety of clinical and diagnostic tests to try to provide a diagnosis. Though an estimated 80% of rare diseases have a genetic etiology, genetic diagnoses are lacking due to a myriad of factors. Genetic variant and phenotype association studies are statistically underpowered due to difficulties in assembling sufficiently large cohorts for rare disease. Evaluating the biological significance of genetic variants is challenging and many variants are left classified as “variant of unknown significance” (VUS). Lastly, patients presenting with complex phenotypes may necessitate the investigation of many variants across many genes to determine a correct diagnosis. While whole genome sequencing (WGS) has been transformative in determining the genetic cause of rare disease cases, the diagnostic rate remains at approximately 50% due to the detailed challenges. Clinical laboratories have begun to utilize RNA sequencing (RNA-seq) to identify candidate genes and functionally validate the effect of such VUSs. Diagnostic laboratories typically rely on manual investigation of VUSs with paired RNA-seq expression events, including aberrant total expression, alternative splicing, and allelic imbalance, for confirmation. Even with the rich information RNA-seq can provide, rare disease diagnostics presents unique interpretation challenges. The relevance of proximal or distal variation to combinations of such expression events may entail the ascertainment of tens to hundreds of potentially relevant VUS, leading to scalability challenges when there is no clear prior genetic candidate. We will present our progress in developing a computational framework to automate the integration of paired WGS and RNA-seq data to identify and prioritize gene/variant candidates that are statistically supported by a multi-omics perspective. An emphasis will be placed on concise representation that reduces the interpretation burden on the reviewing clinical team. We will validate our approach on diagnosed patients in collaboration with the Undiagnosed Diseases Network and the Utah NeoSeq Project, which develop and apply state-of-the-art methodologies to diagnose individuals with rare diseases. The tools will then be applied to undiagnosed individuals from the same cohorts to evaluate cases that remained undiagnosed after WGS analysis.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session II

PB3232 Intentionally degenerate probe design for genotyping multiallelic variants

Authors:

B. Main, B. Wong, J. Gollub; Thermo Fisher Scientific, Santa Clara, CA

The Applied Biosystems™ Axiom™ genotyping microarray platform uses a two-color, hybridization- and ligation-based assay. Biallelic and multiallelic SNPs and indels can be genotyped. Multiallelic genotyping is less robust compared to biallelic because it relies on more probe sequences and a complex calling algorithm with more degrees of freedom. In many applications, such as research in inherited diseases, only pathogenic alleles need to be detected and distinguishing non-pathogenic alleles is not necessary. Here, we present the intentionally degenerate probe design (IDPD) method that simplifies multiallelic genotyping by reducing the number of alleles to distinguish. To accomplish this, IDPD probes ambiguously detect two or more non-pathogenic alleles in the same channel. To estimate the scale of pathogenic markers that are amenable to IDPD, we applied IDPD to a set of 20,232 multiallelic variants in ClinVar. Probe designs were successful for 12,452 (62%) of the variants. The next step would be to compare concordance of IDPD versus standard multiallelic probe design (MA) on an array. However, DNA samples with pathogenic alleles are not readily available. We therefore applied IDPD to a set of 48,497 complex multiallelic variants that have alleles observed in 1000 Genomes Project samples. After standard QC filtering, IDPD outperformed MA with 60% more markers achieving high call rate and concordance (>0.95) to 1000 Genomes Project reference calls. Thus, IDPD is an important new tool that can be used to substantially increase the number of complex pathogenic variants that can be interrogated by the Axiom platform.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session III

PB3233 *Investigating the molecular mechanism of a complex genomic rearrangement causing 6X amplification at 13q33.3*

Authors:

P. Kaur[1], C. M. Grochowski[1], H. Du[1], P. J. Hastings[1], K. Machol[2,3], J. R. Lupski[1,3,4,5,6]; [1]Dept. of Molecular and Human Genetics, Baylor Coll. of Med., Houston, TX, [2]Dept. of Molecular and Human Genetics, Baylor Coll. Med., Houston, TX, [3]Texas Children's Hosp., Houston, TX, [4]Human Genome Sequencing Ctr., Baylor Coll. of Med., Houston, TX, [5]Dept. of Pediatrics, Baylor Coll. of Med., Houston, TX, [6]Baylor Genetics, Houston, TX

Copy number variation (CNV) leading to gene dosage changes are associated with several genomic disorders. When found within a complex genomic rearrangement (CGR), these deviations from diploid state pose a challenge to clinical interpretation partly due to current sequencing technology limitations such as short read lengths and computational challenges of mapping to repetitive sequences. Herein, we present a CGR at 13q33.2 in a 2-years-old male with developmental delay, recurrent infections, hypergammaglobulinemia, laryngeal cleft, thrombocytopenia, fatty liver and splenomegaly. Whole exome sequencing and chromosomal microarray revealed an apparent 6x amplification of LIG4, inherited from phenotypically normal father, however its boundaries could not be ascertained. No other clinically significant variants were identified. In order to better delineate this higher order amplification, we enrolled the family in our IRB approved research protocol and employed short read whole genome sequencing (WGS) followed by read depth analysis using VizCNV plotting tool (https://github.com/BCM-Lupskilab/VizCNV). This revealed a more complex genomic structure than previously detailed on clinical testing. Five copy number changes were seen in an apparent duplication-triplication-6xAmplication-triplciation-duplication event at 13q33.2. All five segments were in positive orientation with length ranging from 5,805 to 240,926 bp. Within these copy number changes, three junctions (blunt fusion, 4 bp insertion and 2 bp deletion) were mapped to nucleotide level resolution. The largest segment had 6X copy number and spanned LIG4, ABHD13 and TNFSF13B. MYO16, was located in region of copy number change truncating this amplified copy of the gene. These findings were confirmed by long read WGS with reads spanning more than one junction. Cohort analysis of over 60,000 individuals who underwent clinical CMA testing revealed four additional patients with CNVs in same relative region affecting the gene LIG4 who had developmental delay, autism, sensory dysfunction and seizures among other findings. The clinical significance and molecular mechanism(s) of gains within 13q33.3 remain to be established. To further resolve this CGR and its clinical significance, we will employ WGS in the four additional patients with gains of LIG4, optical genome mapping to visualize the structure in a linear fashion and expression studies of involved genes. This study emphasizes the complexity of structural mutagenesis and interpretive challenges for clinical genomicists. Understanding CGRs in human disease paves path for evidence based clinical practice.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session I

PB3234 † Lessons learned from the nationwide 10K rare disease genome project.

Authors:

G. Seo[1], H. Han[1], D-W. Kim[1], R. Khang[1], E. Lee[1], J. Han[1], J. lee[1], H. Lee[1], M. Kim[2], B. Lee[3]; [1]3billion, Seoul, Korea, Republic of, [2]Seoul St. Mary s Hosp./The Catholic Univ Korea, Seoul, Korea, Republic of, [3]Asan Med Ctr, Seoul, Korea, Republic of

As part of the pilot study for the South Korea National Project of Bio-big Data supported by the Korea National Institute of Health (KNIH), ~10,000 suspected rare genetic disease patients (4,657) and family members (5,306) were recruited through 17 hospitals between 2021 and 2022. Whole genome sequencing (WGS) was performed on all participants using TruSeq PCR-free library preparation kit and NovaSeq6000 (Illumina). SNV/INDELs and structural variants (SV) including CNVs, inversions, translocations, and repeat expansions were called and interpreted using a custom developed AI-based system, EVIDENCE, that adopts the ACMG variant classification guideline. Once the variants were annotated, filtered, and prioritized, each patient data was manually curated by medical geneticists for reporting of the diagnostic variants and the results were reviewed by the clinical committee to confirm the genotype-phenotype correlation before the return of results to the patient. Overall, 31.6% of the patients received a conclusive molecular diagnosis. Patients with at least one family member sequenced yielded higher diagnostic rate (DR) of 40.8% compared to patients without family members (25% DR). The DR varied by disease categories, and the two groups with the highest DR were ‘neurology/neurodevelopmental disorders (47.3%)’ and ‘dysmorphic/congenital abnormality syndromes (42.9%)’. The lowest DR of 9.2% was observed for ‘cancer’ and the overall DR without ‘cancer’ increased to 37.8%. Of all diagnosed patients, 28.3% were diagnosed with at least one SVs. Most of them had CNVs (97.2%), 1.5% had repeat expansions, and the rest had other SVs including LINE-1 insertions and UPD. Importantly, 3.9% of the diagnosed patients could not have been diagnosed by WES or CMA as the variants were deep intronic or copy-neutral/complex SVs. Additional 5.2% of the patients received inconclusive results and 16 of them had deep intronic variants that were predicted to alter splicing, requiring further study. An essential component of the successful completion of this project within just a year was cost-effective distribution of the tasks across multiple sites and a highly streamlined process including the automated and accurate analytical system handling large genomic dataset. Although further workup such as re-evaluating the patients, reanalyzing the WGS data, integrating RNAseq data and/or performing long-range sequencing is warranted for the undiagnosed patients, this study demonstrates that WGS is a preferred first-line genetic diagnostic tool for a large portion of the patients with suspected rare genetic diseases yielding high diagnostic rate while being fast and accessible.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session II

PB3235 Long-Chain Fatty Acid Oxidation Disorder Gene Variants and Newborn Screening Results from a Sponsored Gene Panel Program

Authors:

V. Rangel Miller[1], H. Gutierrez[1], O. Japalaghi[1], L. Konczal[2], D. Marsden[1], D. McKnight[3], M. Mcnutt[4], M. Vela-Amieva[5], N. Miller[1]; [1]Ultragenyx Pharmaceutical Inc., Novato, CA, [2]Univ. Hosp. Cleveland Med. Ctr., Cleveland, OH, [3]Invitae Corp., San Francisco, CA, [4]UT Southwestern Med. Ctr., Dallas, TX, [5]Natl. Inst. of Pediatrics, Mexico City, Mexico

Introduction: Long-chain fatty acid oxidation disorders (LC-FAOD) are rare, life-threatening but treatable autosomal recessive conditions that result from impaired utilization of long-chain fatty acids for energy production. Clinical signs may include hypoglycemia, cardiomyopathy, cardiac arrhythmias, retinopathy, and rhabdomyolysis. LC-FAOD may be detected through newborn screening (NBS) or plasma/dried blood spot acylcarnitine (AC) analysis. Genetic testing can aid diagnosis, but access to follow-up genetic testing has historically been limited.
Methods: Clinicians suspecting LC-FAOD accessed a no-charge gene panel consisting of the LC-FAOD genes (*ACADVL*, *CPT1A*, *CPT2*, *HADHA*, *HADHB*, *SLC25A20*) plus 19 genes associated with disorders having a similar presentation or overlapping AC profiles. Samples were submitted to Invitae® for clinical testing sponsored by Ultragenyx. De-identified results and optional clinical data were used for this analysis.
Results: As of APR2023, 84 of 1060 patients tested received a presumed LC-FAOD molecular diagnosis (a combination of 2 or more pathogenic variant(s), likely pathogenic variant(s) or VUS): 62% *ACADVL*, 19% *CPT2*, 11% *HADHA*, 4% *HADHB*, 2% *CPT1A*, 2% *SLC25A20*. NBS results were provided for 526 patients as positive (suggestive of increased risk of LC-FAOD) or negative (suggestive of no increased risk of LC-FAOD). A positive NBS and a presumed molecular diagnosis were identified in 58/526 patients. Among 468/526 patients without a molecular diagnosis, NBS results were 95% positive, 5% negative. AC analysis results were provided for 503 patients as positive (suggestive of LC-FAOD), inconclusive (neither normal nor suggestive of LC-FAOD), or negative (normal). A presumed molecular diagnosis was found in 51/503 patients, and AC results were 65% (33/51) positive, 35% (18/51) inconclusive. Among 452/503 patients without a molecular diagnosis, AC results were 73% (331/452) inconclusive, 21% (93/452) positive, 6% (28/452) negative. Additionally, 103 patients had only one LC-FAOD variant identified (83% *ACADVL*). 84/103 patients had NBS results provided: 99% (83/84) positive. 54/103 patients had AC results provided: 78% (42/54) inconclusive, 15% (8/54) positive, 7% (4/54) negative.
Conclusion: Diverse genetic findings were observed among patients tested through this gene panel. VUS and single heterozygous variants were frequent, highlighting the diagnostic challenges for patients with abnormal NBS for LC-FAOD and/or abnormal AC profiles. These data demonstrate the value of gene panel testing to aid LC-FAOD diagnosis and highlight the need for clinical and biochemical evidence to resolve VUS.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session III

PB3236 † Long-read sequencing addressed clinically unsolved problems remained by expanded carrier screening: thalassemia as an example.

Authors:

Y. Zheng[1], M. Chau[2], Y. Cao[1], E. Dong[1], C. CHAN[3], A. LIAUW[4], S. TONG[5], S. LAU[1], A. KWAN[1], M. SHI[1], S. JIANG[1], Y. Kwok[1], T. Leung[1], K. Choy[6]; [1]The Chinese Univ. of Hong Kong, Hong Kong, China, [2]Baylor Coll. of Med., Houston, TX, [3]Hong Kong Maternal & Fetal Med. Clinic, Hong Kong, China, [4]Union Hosp., Hong Kong, China, [5]United Christian Hosp., Hong Kong, China, [6]Chinese Univ. of Hong Kong, Hong Kong, China

Background: Widespread next-generation sequencing- (NGS) based expanded carrier screening (ECS) could support an informative reproductive decision. However, it is limited by a short-read sequencing strategy, and thus could not adequately evaluate several prevalent conditions related to low sequence complexity or repeat regions. This study aimed to assess the utility of long-read sequencing (LRS) in solving these challenges by taking homologous α- and β-globin gene clusters as examples. **Methods:** We retrospectively reviewed NGS-based ECS results of 782 individuals from 2019-Mar to 2021-Sep, among which 14.1% (110/782) cases were screened positive as thalassemia carriers. Inconclusive results were obtained for 0.3% (2/782) of cases. There were 4/476 (0.8%) hemoglobin (Hb) Bart's hydrops at-risk couples (ARCs), and 1/476 (0.2%) Hemoglobin H-constant spring (Hb H-CS) ARC identified. However, alpha-thalassemia deletion subtypes cannot be delineated for future genetic workup. Targeted LRS was applied to 104/774 (13.4%) of the selected cases with sufficient samples, including 1) Ten cases in the unsolved group: The two inconclusive cases, the four Hb Bart's ARCs, and one Hb H-CS ARC and 2) Ninety-two other thalassemia carrier cases in the validation group. Variant analysis was blinded from previous testing results. **Results:** LRS resolved all (12/12) cases in the unsolved group: phased one case with both *HBA1* deletion and *HBA2* triplication as HKαα (Hong Kong αα) allele, thus clarifying the non-carrier status; revealed ECS missing variant Hb Quong Sze for another case with abnormal hemoglobin parameters; delineated deletion subtype as SEA for all four Hb Bart's ARCs; and THAI for Hb H-CS ARC. For the validation group, all 97 events detected by the NGS-based ECS panel were validated by LRS. Overall, 2.8% (3/108) additional events missed by the NGS-based ECS panel were revealed by LRS, leading to a change of thalassemia carrier status for 1.9% (2/104) cases after internal curation. **Conclusions:** LRS is more comprehensive for thalassemia screening. The potential of LRS acting as a supplement of NGS-based ECS panel for challenging genes for reducing residual risk deserves further investigation.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session I

PB3237 Making the call: Trends among successful strategies to conclude the diagnostic odyssey for participants at the Pacific Northwest Undiagnosed Diseases Network clinical site.

Authors:

E. Blue[1,2], S. Sheppeard[1], A. Allworth[1], E. Kaitryn[1], J. Chen[1], D. Miller[1,2,3], E. Rosenthal[1], A. Stergachis[1,2], S. Strohbehn[1], N. Dargie[1], D. Doherty[1,3,2], A. Folta[1], M. Gillentine[3], M. Horike-Pyne[1], G. Mirzaa[4], J. Ranchalis[1], U. Schwarze[1], M. Bamshad[1,3,2], P. Byers[1], S. Chanprasert[1], I. Glass[1,2], K. Leppig[5], W. Raskind[1], V. Sybert[1], M. Wener[1], F. Hisama[1,2], K. Dipple[3,1,2], G. Jarvik[1,2], University of Washington Center for Mendelian Genomics (UW-CMG), Undiagnosed Diseases Network (UDN); [1]Univ. of Washington, Seattle, WA, [2]Brotman Baty Inst., Seattle, WA, [3]Seattle Children's Hosp., Seattle, WA, [4]Seattle Children's Res. Inst., Seattle, WA, [5]Group Hlth.Cooperative, Kaiser Permanente Washington, Seattle, WA

The Undiagnosed Diseases Network (UDN) offers comprehensive clinical and multiomics profiling to individuals with rare disorders who have exhausted local resources during their long diagnostic odyssey. The Pacific Northwest UDN clinical site has provided diagnoses to dozens of individuals since 2018 through our interdisciplinary approach and deep, diverse range of expertise. Here, we offer metrics and trends among our successes and obstacles to diagnosis, each supported by an example case.

We performed a comprehensive review of the 258 applications reviewed by our center over the last five years, up to May 5, 2023. Through our careful review of existing medical records, we diagnosed 25 (10%) applicants without admitting them to our program. The overwhelming majority of these were adults (88%), and only 16% involved genetic diagnoses.

Among the 134 participants enrolled in our study (52% of applicants), we diagnosed 26 (27%) of the 95 participants evaluated to date. We are pursuing strong candidates for another 18 (19%), while evaluations are pending or ongoing for the remaining participants. Most evaluated participants with new diagnoses are pediatric (58%) and most have a genetic diagnosis (88%). A "genome-first" approach would have provided many (35%) of these genetic diagnoses while avoiding costly chart review, as pathogenic or likely pathogenic variants were identified in the original clinical exome or genome analysis.

Our program emphasizes that the value of reanalysis and fresh clinical and molecular perspectives cannot be overstated. Our depth of clinical expertise applied to medical records across time and providers benefited all applicants and directly led to 24 diagnoses (47%), while reanalysis of clinical sequence data led to 10 molecular diagnoses (38%). Atypical phenotypic presentations could be explained by multiple rare disease diagnoses in a single participant (n >= 2) or novel disease mechanisms (n = 2) or genes (n = 3). Two cases of mosaicism required sequencing of multiple tissue sources, while research technologies including long-read DNA sequencing and transcriptomics resolved the boundaries and phenotypic consequences of pathogenic structural variants (n = 3).

Despite our best efforts, diagnostic rates were slowed by common obstacles in rare disease research, including the "n of 1" problem requiring patient-specific experiments in model systems and the patience to wait for matches within gene- and phenotype-based patient networks. These results emphasize the need for precise data sharing among networks of experts who are then provided time and targeted experimental resources tailored to individual patient needs.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session II

PB3238 Measuring genome wide sensitivity for detecting copy number variations (CNVs) using a sampling methodology.

Authors:

A. Roter[1], B. Bolstad[1], J. Schmidt[2], L. Wu[2]; [1]Thermo Fisher Scientific, Healdsburg, CA, [2]Thermo Fisher Scientific, Santa Clara, CA

Chromosomal Microarrays are the first tier research tool for chromosomal abnormalities. But measuring the sensitivity to detect CNVs for all 3 billion base pairs of the human genome using real samples is an impossible task. As a reasonable proxy, we use a sampling method on microarray data generated on chromosome aberration cell lines from the National Institute of General Medical Science Chromosome Abnormality Human Genetic Cell Repository. These cell lines have large chromosomal aberrations due to aneuploidy, triploidy and unbalanced translocations that create large regions of non-neutral copy number across a large portion of the genome. Genomic DNA from 41 chromosome aberration samples was processed and run 3 to 7 replicates on the CytoScan HD Accel microarray, using the accelerated 2-day protocol using 100µg of input DNA. This generated a data set of 195 microarrays with large aberrations covering ~ 50% of the genome. Stretches of log ratio data with lengths of 25 to 150 markers were sampled from the aberration regions and inserted into 100 marker stretches of log ratios randomly sampled from copy number neutral regions. These stretches of marker data were run through a Hidden Markov Model copy number segmentation and calling algorithm and scored as detected if 90% of the markers were called as the correct copy number by the algorithm. This was repeated for every 25 bp sliding window in the aberration regions, resulting in a test of over 130,000 sampled regions. For regions of copy number loss, the average sensitivity for detecting 25 marker segments of at least 25kbp was 97.4%. For regions of copy number gain, the average sensitivity for detecting 50 marker segments of at least 50kbp was 96.5%. Analysis of the large chromosome aberrations showed that the mean reproducibility for detecting losses or gains of ≥ 100kbp was 100% (95% confidence interval of 98.8 to 100%). In addition, the false discovery rate for 400kbp segment with ≥50 markers was 2.3% (2.7% for gains and 1.8% for losses), corresponding to a PPV of 97.7%.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session III

PB3239 Minimizing uncertainty and increasing equity in genetic testing: the role of machine learning tools in the classification of genetic variants

Authors:

F. Facio[1], A. Colavin[1], B. Johnson[1], S. Aradhya[1], J. Reuter[1], S. Padigepati[1], T. Manders[1], J. Nicoludis[2], D. Stafford[1], A. Wahl[1], Y. Kobayashi[1], R. Nussbaum[1], K. Nykamp[1], L. Fresard[1]; [1]Invitae, San Francisco, CA, [2]Genentech, San Francisco, CA

INTRODUCTION: Variants of uncertain significance (VUS) remain a challenge in the translation of genetic information to clinical care. This study describes an Evidence Modeling Platform (EMP) designed to validate and integrate novel variant evidence sources. This EMP combines clinical genomics expertise with machine learning (ML) models built from various types of input data (e.g. evolutionary conservation, population frequency, protein structure) to provide evidence for a variant's pathogenicity. Here we present the **validation** and the **impact** of the EMP models on variant classification (VC) for individuals undergoing testing for diverse hereditary conditions.
METHODS: EMP was validated by calculating the concordance in VC with and without the input of its models for novel variants observed at Invitae® from May 2022 to May 2023. Impact was measured by assessing the proportion of individuals with ≥1 variants classified as Benign (B), Likely Benign (LB), Likely Pathogenic (LP), Pathogenic (P) after the application of EMP evidence. For each clinical area and clinician-reported ancestry group analyzed, 10 subsets of 5,000 individuals were sampled.
RESULTS: **Validation**: 27,477/496,209 novel variants were eligible for validation, as they received evidence from at least one EMP model and had reached a non-VUS classification without that evidence. Across 8,517 pathogenic variants and 14,960 benign variants, the EMP models achieved 98% Negative Predictive Value and 90% Positive Predictive Value. **Impact**: Overall, 22.3% of individuals had at least one variant classified based on the input from EMP models (21.6% from VUS to B/LB, 1.4% to P/LP). Impact varied by clinical area, being the greatest among individuals tested for immunology (60.5%) and neurology (48.5%). Impact also varied by ancestry, with the greatest impact seen among individuals of Asian (33.7%) and Black (28.7%) ancestries compared to White individuals (18.5%, two-sided t-test p<1e-10).
CONCLUSION: A systematic approach to integrating ML models into VC yields highly accurate interpretations, allowing a significant proportion of individuals with a VUS to receive a more definitive result. EMP models are largely informed by biologically-grounded hypotheses, not biased by the ancestral origin of the evidence, leading to more equity in VC.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session I

PB3240 Mitochondrial dysfunction detected by peripheral blood mononuclear cells analysis in patients with Williams-Beuren syndrome

Authors:

I-C. Ling[1], T-L. Liao[2], D-M. Niu[1,3,4], Y-J. Chen[1,3,2]; [1]Dept. of Pediatrics, Taipei Veterans Gen. Hosp., Taipei, Taiwan, [2]Dept. of Life Sci. and Inst. of Genomic Sci., Natl. Yang-Ming Chiao Tung Univ., Taipei, Taiwan, [3]Inst. of Clinical Med., Sch. of Med., Natl. Yang Ming Chiao Tung Univ., Taipei, Taiwan, [4]Genetic Counseling Ctr., Taipei Veterans Gen. Hosp., Taipei, Taiwan

Williams-Beuren syndrome (WBS) is a rare genetic disorder caused by the deletion of about 28 genes on chromosome 7q11.23. Secondary mitochondrial dysfunction frequently accompanies genetic disorders and plays a significant role in their pathogenesis even without major mitochondrial gene defects. Previous studies have implied that individuals with WBS might have decreased mitochondrial function. However, practically, there are no feasible methods for screening mitochondrial function in WBS patients.

Peripheral blood mononuclear cells (PBMCs) analysis through Seahorse XFe analyzer and flow cytometry were used to evaluate mitochondrial function in this study. For XFe analyzer, cellular respiration, and bioenergetics are measured, where mitochondrial mass, reactive oxygen species (ROS), and membrane potential are estimated via flow cytometry.

PBMCs from 30 healthy individuals (HI) and 15 WBS patients were collected. In the bioenergetic analysis by Seahorse XFe analyzer, a significant reduction of mitochondrial respiration was found in WBS patients, including decreased maximal respiration, decreased spare respiratory capacity, and increased proton leakage. The overall bioenergetics health index (BHI) was lower in WBS than in normal healthy individuals. BHI were 1.20±0.44 and 1.93±0.29 for patients with WBS and HI, respectively. In flow cytometry, ROS and mitochondrial mass were higher in WBS than in HI; while the mitochondrial membrane potential was lower in WBS than in HI. All of these could imply that WBS patients have decreased mitochondrial function even though they have more mitochondria.

In conclusion, bioenergetic analysis and flow cytometry were performed on blood samples of patients with WBS and HI. Our results revealed patients with WBS had lower BHI, lower mitochondrial membrane potential, and higher ROS. All of these indicate mitochondrial dysfunction in WBS patients and enhancement of mitochondrial function may improve WBS clinical manifestation.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session II

PB3241 Molecular and phenotypic characterization of a group of Colombian patients with clinical suspicion of a RASopathy.

Authors:

M. Suarez Rubiano[1], S. Bello Uyaban[1], J. Lores Espinosa[2], H. Pachajoa[3], M. Galvez[4]; [1]Gencell – Genuino Res. Group, Bogotá, Colombia, [2]Pontifical Javeriana Univ., Cali, Colombia, [3]Icesi Univ. - Valle de Lili Fndn.., Cali, Colombia, [4]Gencell – Genuino Res. Group, Bogota, Colombia

Purpose: RASopathies are rare diseases with a worldwide incidence of 1 in 1.000 live births. This group of diseases present common clinical characteristics, secondary to germline variants in the RAS-MAPK pathway. In Colombia, there is only one study characterizing Noonan Syndrome (a RASopathy) in a specific geographic region of the country. A wider knowledge of these diseases' spectrum in our country is needed.
Objective: To perform a genotypic and phenotypic characterization in a group of Colombian patients with a clinical suspicion of a RASopathy.
Methodology: Observational, retrospective, cross-sectional study including subjects who underwent next-generation sequencing (NGS) analysis of genes related to RASopathies. Two groups were defined based on clinical suspicion: 1) Neurofibromatosis and 2) Other RASopathies. The analysis includes a description of pathogenic (P) and likely pathogenic (LP) variants, as well as a general phenotypic description of the two groups.
Results: A total of 127 subjects, distributed among the five Colombian geographic regions (Andean, Pacific, Orinoquia, Caribbean and Amazon) were included. Fifty-four percent were male. The most frequent diagnosis suspicion was Neurofibromatosis in 77 (60.6%) subjects, who had a median age of 10 years (IQR: 3-16) at the time of molecular testing. P or LP variants were identified in 29 (37.6%) of these patients, with NF1 being the most affected gene (96.5%) and null variants (frameshift, nonsense and deletions) being the most frequent (75%). We found 50 (39.3%) cases with clinical suspicion of other RASopathies, with a median age of 7.5 years (IQR: 3-15) at the time of genetic testing. P or LP variants were identified in 21 (42%) of these cases. PTPN11 mutations were predominant (52.3%) and missense variants were the most frequent in this group (85.71%). Overall, 50 subjects had a confirmed molecular diagnosis of a RASopathy, 56% with Neurofibromatosis type 1 and 30% with Noonan syndrome. Other less frequent syndromes included Leopard, cardiofaciocutaneous, Schwannomatosis and Noonan-like.
Conclusions: We describe the largest known genotypic-phenotypic characterization in a Colombian group of patients with a suspected RASopathy. NGS is an useful tool for molecular confirmation of RASopathies, given the genetic heterogeneity of this spectrum. We observed a delayed median age for molecular diagnosis, which in the era of precision medicine could impact access to personalized therapies.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session III

PB3242 Molecular characterization of a colombian group of patients with clinical suspicion of dystrophynopathies.

Authors:

M. Galvez[1], S. Bello Uyaban[2], A. Eslava[1], S. Castellar[3]; [1]Gencell-Genuino Res. Group, Bogota, Colombia, [2]Gencell-Genuino Res. Group, Bogotá, Colombia, [3]Univ. Natl. de Colombia-Inst. Roosevelt, Bogota, Colombia

Purpose:
Dystrophynopathies are the most common inherited muscular dystrophy (incidence 19.5/100.000 live male births). These are caused by mutations in the *DMD* gene. Some of these lead to milder manifestations, known as Becker, and the severest form known as Duchenne. Initial clinical features should be suspected by primary physicians, but in low-income countries as Colombia they are commonly suspected by clinical specialists. There are a limited number of studies concerning dystrophynopathies in our country. Therefore, understanding the underlying genetic causes and the health professionals requesting molecular testing in Colombia is relevant.
Objectives:
To describe the molecular findings in a group of Colombian male patients who underwent molecular testing of the *DMD* gene within the context of a personal and/or family history of a suspected distrophinopathy.
Methodology:
Retrospective, descriptive, cross-sectional study analyzing the results of *DMD* molecular testing and the medical specialty requests, between 2019 and 2022. Initially, a multiplex ligation-dependent probe amplification (MLPA) was carried out, and when negative, followed by next-generation sequencing (NGS). The molecular results were analyzed considering only pathogenic and likely pathogenic variants.
Results:
We included 346 male patients, of which seven were familial cases, belonging to three families. Median age at the moment of molecular testing was 9 years (IQR 6-14) and the demographic analysis showed a distribution of subjects in four out of the five Colombian geographic regions (Andean, Pacific, Orinoquia and Caribbean). Health professionals requesting the genetic test included pediatric neurologists (31.2%), clinical geneticists (26.6%), physiatrists (16.7%), pediatricians (12.2%) and others (13.3%). A pathogenic or likely pathogenic variant was identified in 41% of cases, confirming a dystrophinopathy in 142 patients. Copy number variants (CNV) were identified in 91 (64%) individuals, 82.4% of them corresponded to deletions. Single-nucleotide variants were found in 51 cases (36%), of which nonsense were the most frequent (54.9%), followed by frameshift (25.5%) and splicing site (17.6%).
Conclusions:
CNV were the most frequently identified variants, accordingly to worldwide genotypic description associated with *DMD* gene, which highlights the importance that health professionals should consider first a MLPA-based diagnosis strategy, and only in negative cases, requesting a sequencing technique during primary attention. We found a low proportion of molecular requests by primary care programs, which could impact early diagnosis and treatment.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session I

PB3243 Molecular diagnosis for patients with autism spectrum disorder

Authors:

N. Miyake[1], T. Yoshinori[2], R. Fukai[3], E. Koshimizu[3], S. Miyatake[3], A. Takata[4], N. Matsumoto[3]; [1]Natl. Ctr. for Global Hlth.and Med., Tokyo, Japan, [2]Sagami Women's Univ., Sagamihara, Japan, [3]Yokohama City Univ. Graduate Sch. of Med., Yokohama, Japan, [4]RIKEN Ctr. for Brain Sci., Wako, Saitama, Japan

Autism spectrum disorder (ASD) is a heterogenous neurodevelopmental disorder characterized by persistent deficits in social communication and interaction across multiple contexts, and restricted, repetitive pattern of behavior, interests, or activities, with onset during the first year of life. Monogenic and polygenic causes have been indicated. We investigated monogenic factors in 405 individuals (351 trios, 24 quads, and 2 quintet families) affected with ASD using trio-based whole exome sequencing (WES). We reached the molecular diagnosis in 66 affected individuals (16.3%). Total 55 causative variants including 45 single nucleotide variants (SNVs) and 10 small insertion deletions (indels) were classified “Pathogenic” (28 variants) and “Likely pathogenic” (27 variants) based on the American Collage of Medical Genetics/Association for Molecular Pathology guidelines. They were confirmed by Sanger sequencing. We also analyzed copy number variants (CNVs) using WES data by eXome Hidden Marcov Model, and reached molecular diagnosis in 13 affected individuals (3.2%). 13 CNVs were confirmed by quantitative PCR. The molecular diagnostic rate in our cohort was higher in females than in males, which is consistent with “female protective effect” as previously reported. We analyzed familial cases (24 quads and 2 quintets), and contrary to expectations, we found only one family shared the pathological variant. Interestingly, the molecular diagnostic rate in familial cases (4/54, 7.4%) was lower than simplex cases (62/351, 17.7%), which might indicate a difference in genetic architecture between simplex and multiplex families. As one attempt, we simulated the diagnostic yield of WES data using known disease-related aberrant genes for each year since 1992, and calculated that the average percentage of newly resolved cased was 0.63% (0%~2.5%) per year in our ASD cohort. We re-recognized the importance of re-analysis using the latest annotation data.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session II

PB3244 Molecular Diagnostic Yield of Exome Sequencing and Chromosomal Microarray in Short Stature: A Systematic Review and Meta-analysis

Authors:

Q. Li[1], Z. Chen[1], J. Wang[1], K. Xu[1], X. Fan[2], C. Gong[3], Z. Zhao[1], G. Li[1], J. Cai[1], Z. Wu[4], T. Zhang[1], N. Wu[1]; [1]Peking Union Med. Coll. Hosp., Beijing, China, [2]Dept. of Pediatric, The second affiliated Hosp. of Guangxi Med. Univ., Guangxi, China, [3]Dept. of Endocrinology, Genetics and Metabolism, Beijing Children's Hosp., Capital Med. Univ., Natl. Ctr. for Children's Hlth., Beijing, China, [4]Peking Union Med. Coll. Hosp., Chinese Academy of Med. Sci., Beijing, China

Background: The diagnostic yield of exome sequencing (ES) and chromosomal microarray analysis (CMA) for short stature cohorts is currently uncertain, despite their widespread use. A definitive diagnostic yield has not been established.
Methods: A systematic literature search was conducted in three databases (PubMed, Embase, and Web of Science) in February 2023 to investigate the diagnostic yield of ES and CMA in short stature. Eligible studies for meta-analysis were those that had at least 10 participants with short stature who were diagnosed using either ES or CMA, and the number of diagnosed patients was reported. Out of 5,222 identified studies, 20 were eventually included in the study. Two independent investigators extracted relevant information from each study, which was then synthesized using proportional meta-analysis to obtain the overall diagnostic yield of ES and CMA.
Results: The overall diagnostic yield of ES among the cohorts (10 study cohorts comprising 1,350 individuals) and CMA among the cohorts (14 study cohorts comprising 1,070 individuals) was found to be 27.1% (95% CI, 18.1%-37.2%) and 13.6% (95% CI, 9.2%-18.7%), respectively. A subgroup meta-analysis was also performed to assess if the diagnostic yield varied depending on whether ES was used as a first-tier or last resort test. Although no statistically significant difference was observed between the first-tier and last resort groups, a slightly higher diagnostic yield was observed in the former. Furthermore, a meta-regression was carried out to investigate how the diagnostic yield varied over time. There was an increase in the percentage of positively diagnosed patients over time, although this was not statistically significant.
Conclusions: The results of this systematic review and meta-analysis suggest that ES and CMA may be warranted to identify molecular disorders of children with short stature. The diagnostic yield of ES was higher than that of CMA, and there was a slightly higher diagnostic yield when ES was used as a first-tier test. Although there was no statistically significant increase in diagnostic yield over time, the trend suggests that these tests may become even more effective in the future. These findings have important implications for clinical practice, as they suggest that ES and CMA should be considered as diagnostic tools for children with short stature.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session III

PB3245 Molecular genetic analysis of patients with skeletal disorders: Identification of ultra-rare monogenic diseases.

Authors:

L. Dvorakova[1], M. Reboun[1], R. Bakalar[1], P. Slavikova[1], R. Svacinova[1], V. Todorovova[1], K. Peskova[1], A. Baxova[2], T. Honzik[1]; [1]Dept. of Pediatrics and Inherited Metabolic Disorders, Gen. Univ. Hosp. and 1st Faculty of Med., Charles Univ., Prague, Czech Republic, [2]Inst. of Biology and Med. Genetics, Gen. Univ. Hosp. and 1st Faculty of Med., Charles Univ., Prague, Czech Republic

Background: Monogenic skeletal disorders are a heterogeneous group of disorders affecting the development, growth and homeostasis of bone and/or cartilage. They remain a diagnostic challenge due to their phenotypic diversity.**Methods:** Sequencing of 334 genes was performed using a custom enrichment system (Roche) and the MiSeq platform (Illumina). Variant calling was performed using an in-house pipeline including the ANNOVAR annotation tool and the CNV kit. **Results:** Molecular genetic analysis was performed in 109 patients, the causative mutation was found in 36 of them. Low diagnostic yield is observed mainly in children with isolated short stature (height below -2.5 SD) and isolated osteoporosis (Z-score below -3). Some of the diagnosed patients had specific signs leading to a high suspicion of the disease and molecular genetic analysis confirmed the diagnosis. For example, Aarskog-Scott syndrome (*FGD1*, two probands), fibrodysplasia ossificans progressiva (*ACVR1*, one proband), hypophosphatemic rickets (*PHEX*, three probands), multiple epiphyseal dysplasia (*SLC26A2*, four probands). The other patients had less specific signs, in three of them an extremely rare diagnosis was confirmed. In one patient, two undescribed variants (c.82+2T>G and c.258G>C, which may lead to missense and/or splicing mutation) were identified in *B3GAT3*. The clinical symptoms were fully compatible with this finding and led to the diagnosis of glycosaminoglycan linkeropathy (osteoporosis, repeated fractures, radioulnar synostosis, valve involvement, joint dislocation, hypermobility). In the second patient with IUGR/SGA, growth failure (-2.9 SD), microcephaly and delayed speech development a deletion of the distal part of the long arm of chromosome 15 including *IGF1R* was identified. We conclude that haploinsufficiency of *IGF1R* is a highly probable cause of the patient's phenotypic symptoms. The third patient was a newborn with craniosynostosis, prominent forehead, midface hypoplasia, proptosis, limb contractures, clubfoot, long toes, genital anomalies (originally thought to be female, normal male profile was found by array CGH). Two previously described mutations in *POR* confirmed the diagnosis of Antley-Bixler syndrome.**Conclusions:** Skeletal disorders show a wide variety of phenotypic manifestations and a high degree of heterogeneity in the genetic basis. Next-generation sequencing is the method of choice to confirm the diagnosis. One of the reasons why some patients remain undiagnosed is a mutation in a gene not included in the panel. We are now testing the yield of whole exome sequencing in undiagnosed patients.Support: MH CR-RVO-VFN00064165

Session Title: Molecular and Cytogenetic Diagnostics Poster Session I

PB3246 Molecular genetic testing of congenital adrenal hyperplasia due to 21-hydroxylase deficiency should include comprehensive chimera analysis.

Authors:

Q. Lao[1], S. Kollender[1], D. Burkardt[1], A. Schulman[1], A. Moon[1], D. Merke[1,2]; [1]NIH Clinical Ctr., Bethesda, MD, [2]The Eunice Kennedy Shriver Natl. Inst. of Child Hlth.and Human Dev., Bethesda, MD

Congenital adrenal hyperplasia (CAH) is a group of autosomal recessive diseases of steroid hormone biosynthesis with 21-hydroxylase deficiency CAH due to defects in *CYP21A2* accounting for 95% of the cases. Both neonatal screening and diagnosis are based on biochemical testing. Molecular genetic testing is a vital second tier means of CAH diagnosis and is essential for genetic counseling, but is not suitable for massive parallel sequencing methodologies due to the pseudogene interferences across the genome locus where *CYP21A2* is mapped, and chimeric genes due to large gene conversions or unequal crossovers (also called 30-kb deletions) during meiosis commonly exist. Most commercial molecular tests detect pathogenic CAH alleles of minor pseudogene conversions and chimeric genes but rarely analyze the latter in detail. In our large CAH cohort at the U.S. National Institutes of Health Clinical Center (n=420), chimeric genes account for 28% of the CAH alleles (242 of 850), and a comprehensive analysis of the chimeric genes is clinically meaningful in 20% (39/197) of patients carrying a chimeric allele. *CYP21A1P-TNXA/TNXB*, or CAH-X chimeras, are associated with hypermobility type Ehlers-Danlos syndrome manifestations. The vast majority of known *CYP21A1P/CYP21A2* chimeras are null 38/197mutations, but attenuated chimeras corresponding to a milder phenotype also exist and have been historically under-detected. Identification of attenuated chimeras may resolve some common genotype-phenotype discrepancies because they are misdiagnosed as either a P30L variant associated with the nonclassic mild phenotype or a null chimera associated with the most severe salt-wasting phenotype by current methodologies, but in fact attenuated chimeras are often associated with a moderate simple virilizing phenotype. Although, in general, we agree with the European Molecular Genetics Quality Network practice guidelines for molecular genetic testing of CAH due to *CYP21A2* defects, comprehensive chimera analysis should be included. Early diagnosis of *CYP21A1P-TNXA/TNXB*, or CAH-X chimeras, is especially valuable because monoallelic presence is associated with the Ehlers-Danlos phenotype including possible cardiac anomalies regardless of CAH status. A genetic diagnosis thus offers valuable early awareness.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session II

PB3247 Monosomy 7 in Philadelphia-PositiveCML

Authors:

N. Abdelmoula, S. Aloulou, B. Abdelmoula, S. Kammoun; Genomics of Signalopathies at the service of medicine UR17ES36, Med. Univ. of Sfax, Sfax, Tunisia

Background: Chronic myeloid leukemia (CML) is a clonal hematopoietic stem cell disorder, characterized by reciprocal translocation t(9,22)(q34;q11). Occasionally, CML patients harbor additional chromosomal anomalies during the chronic phase. While some additional chromosomal aberrations are common such as trisomy 8, others such as monosomy 7 (5 percent) are rare. The additional chromosomal aberrations may be factors of prognostic significance. Our aim was to report a rare case of association of an additional monosomy 7 in a Ph+ CML patient. In this cytogenetic CML presentation, we described related hematologic characteristics as well as the clinical course. Methods: A 43-year-old Tunisian female patient was diagnosed with CML at the chronic phase based on peripheral blood and bone marrow examination. She was treated by Busulfan therapy for 13 months. At the clinical level, she was in remission. Hematological assessment was carried out using blood cell count, peripheral blood smears, bone marrow aspirate and biopsy, as well as bone marrow cytogenetic analysis. Results: Blood cell count revealed aregenerative normocytic normochromic anemia, impaired production of red blood cells with punctate basophilia (the presence of numerous basophilic granules that are dispersed through the cytoplasm of erythrocytes) in peripheral blood smear, as well as reticulin myelofibrosis on bone marrow biopsy. There were also abnormalities of the megakaryocytic cell lineage at sternal puncture. Megakaryocytic dysplasia consisted of numerous micro-megakaryocytes and hypo-lobed nuclei in megakaryocytes. Cytogenetic analysis revealed the Philadelphia translocation associated with monosomy 7 in all 30 analyzed metaphases: 45,XX, t(9,22)(q34;q11), -7. The clinical course was fatal and the patient died in the context of severe infection one month after the evaluation and 14 months after the CML diagnosis. Conclusion: Abnormalities involving chromosome 7 are one of the most frequent chromosomal abnormalities in myeloid disorders including myelodysplastic syndromes and acute myeloid leukemia, and they are associated with an adverse prognosis. In the Ph+ CML patient reported here, we identified additional monosomy 7 at the chronic phase. This rare additional abnormality suggested an acceleration of the CML or a myelodysplastic syndrome secondary to alkylating agent treatment. The aggressive clinical course in the reported patient confirms that the monosomy 7, as an additional chromosomal abnormality, indicates CML transformation/progression.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session III

PB3248 Mosaic: A multiplexed in-vivo drug screening platform at the intersection of single-cell RNA sequencing and small molecule oncology.

Authors:

K. Jin; Vevo Therapeutics, San Francisco, CA

Herein we present MOSAIC, a multiplexed in vivo drug discovery platform. MOSAIC combines advances in single-cell RNA sequencing with in-vivo xenograft model systems. The core principle behind the platform is combining cells from tens to hundreds of cell lines or patient-derived xenograft models into a single MOSAIC tumor and then implantation of that multiplexed tumor into 3D organoid or in-vivo mouse models. Tumors are then treated appropriately in the model system over an extended time course with small molecule oncology drugs and then post-treatment, dissociated and prepared for single-cell RNA sequencing. Cells are deconvolved according to patient of origin, mouse identity, and drug treatment without the need for barcoding of models. This allows for simultaneous interrogation of multiple model systems in a single mouse, dramatically reducing the scale of experimentation needed while simultaneously increasing the scale and data output by way of high-throughput transcriptome based readouts. We demonstrate utility and scalability of the MOSAIC method in this abstract by presenting data from many models across many compounds.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session I

PB3249 Multimodal sequencing approaches to end the diagnostic odyssey of patients with suspected rare monogenetic diseases.

Authors:

J. A. Madden[1], W. Shao[2], C. A. Genetti[1], H. Li[3], D. A. Johnson[4], K. McWalter[4], J. Noya[4], K. Arvai[4], V. Gainullin[4], K. Schmitz-Abe[1], V. D. Ustach[4], S. Rockowitz[2,1], P. B. Agrawal[1], C. A. Brownstein[1], S. Newman[4], J. Chin[5], J. M. Devaney[4], P. Kruszka[4], A. H. Beggs[1]; [1]The Manton Ctr. for Orphan Disease Res., Div. of Genetics and Genomics, Boston Children's Hosp., Harvard Med. Sch., Boston, MA, [2]Res. Computing, Boston Children's Hosp., Boston, MA, [3]Dept. of Data Sci., Dana-Farber Cancer Inst., Boston, MA, [4]GeneDx, Gaithersburg, MD, [5]GeneDx, Gaithersburg, MA

Despite multi-year diagnostic odysseys, many individuals with neuromuscular and developmental disorders never receive a genetic diagnosis. To develop best practices for comprehensive, multimodal, genomic diagnostic evaluations, we combined Illumina-based short-read genome sequencing (GS), with PacBio HiFi long-read GS, and RNA sequencing to investigate six trio/quad families enrolled by The Manton Center for Orphan Disease Research at Boston Children's Hospital. Each case had a strong suspicion of monogenic disease and had extensive phenotyping and genetic testing but remained undiagnosed. DNAs were extracted from either fresh peripheral blood or previously frozen and banked whole blood specimens, both of which yielded high-quality data by short and long-read GS technologies. Short-read GS data were processed through a DRAGEN Bio-IT Platform and VExP and long-read data were analyzed via the PacBio Human Whole Genome bioinformatics workflow, with results of both integrated and delivered through Genuity Sequence Miner via custom scripts. De novo long-read genome assemblies with Hifiasm were compared with the recent draft human pangenome with PGR-TK and structural variants (SVs) and CNVs of interest were identified with a novel kmer counting approach. Candidate genes for focused variant evaluation were prioritized based on a novel machine-learning phenomics algorithm, "Multiscore", as well as clinical domain expert input based on comprehensive clinical phenotyping. SV analysis of the long-read data identified variants of interest in three of the six families. Notable findings include a de novo 119 kb intragenic deletion of *TTN* predicting in-frame loss of exons 34-152 (NM_001267550) that was confirmed on RNAseq of skeletal muscle mRNA in a proband with neuromuscular disease, and copy number gain of the triplicated region of *NEB* that may be responsible for nemaline myopathy in another proband. A novel duplication within the 15q13 region in a 26-year-old male with a complex neurodevelopmental disorder is also under further study. Short-read analysis also identified candidate single nucleotide variants in this and other cases. RNAseq is pending to evaluate the potential impact of a de novo splice region variant in *UBA6* in this proband as well as the impact of a *ZEB2* deletion in two brothers with intellectual disability, psychiatric disorders, and connective tissue abnormalities. These results demonstrate the potential of multimodal and 'next generation' genomic approaches to generate new candidate variants for patients previously unable to receive a diagnosis through traditional short-read sequencing methods alone.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session II

PB3250 Nanopore long-read sequencing analysis in patients with congenital malformations and balanced chromosomal abnormalities

Authors:

H. Murakami[1,2], Y. Enomoto[2], T. Sakashita[1], T. Kumaki[2], N. Aida[2], A. Imamura[1], K. Kurosawa[2], H. Kaneko[1]; [1]Gifu Prefectural Gen. Med. Ctr., Gifu, Japan, [2]Kanagawa Children's Med. Ctr., Yokohama, Japan

Accurate evaluation of breakpoints is crucial for analyzing balanced chromosomal abnormalities in patients with congenital anomalies. However, conventional chromosome analysis methods face difficulties in precisely pinpointing breakpoints of structural abnormalities that do not involve copy number variations. In this study, we employed Nanopore long-read sequencing to perform breakpoint analysis in two patients presenting with multiple malformations and developmental delay. These patients exhibited a *de novo* inversion of chromosome 3q and a complex balanced translocation t(9;17;20), respectively. The first patient presented with bicoronal craniosynostosis, characteristic facial features, and cerebellar malformation. Breakpoint analysis revealed a breakpoint in the distal centromere of 3q24, located 7 kb downstream of the 3′ untranslated region of *ZIC1*. *ZIC1*, a member of the Zic family, encodes a transcription factor with zinc finger domains that is essential for the normal development of the cerebellum. Loss-of-function of *ZIC1* causes Dandy-Walker malformation, while gain-of-function leads to a multiple congenital anomaly syndrome. On the basis of the clinical similarities, we concluded that the abnormalities in this patient were caused by the transcriptional dysregulation of *ZIC1*. We hypothesize the underlying molecular mechanisms of transcriptional dysregulation of *ZIC1* such as the abnormalities in topologically associated domains encompassing *ZIC1* or abnormal transcriptional extension of 3'-terminus of *ZIC1*. The second patient presented hydrocephalus, esophageal atresia, and developmental delay. Through the analysis, we identified 12 breakpoints and 4 microdeletions. Notably, a breakpoint at 20q12q13.11 cleaved *PTPRT* gene, which has been associate with developmental delay resulting from haploinsufficiency. These findings demonstrate the clinical utility of long-read sequencers in analyzing balanced chromosomal abnormalities.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session III

PB3251 National Validation of Long-read Whole Genome Sequencing for Characterizing Chromosomal Genomic Rearrangements in Sweden

Authors:

A. Lindstrand[1], F. Lenner[2], J. Eisfeldt[1], A. Ameur[2], M. Ek[1], H. Ehrencrona[3], L. Lovmar[4], M. Melin[2], L. Feuk[2], Genomic Medicine Sweden - Rare Diseases; [1]Karolinska Inst, Stockholm, Sweden, [2]Uppsala Univ., Uppsala, Sweden, [3]Lund Univ Hosp, Lund, Sweden, [4]Sahlgrenska Univ. Hosp., Gothenburg, Sweden

Clinical cytogenetic laboratories often require comprehensive analysis of chromosomal genomic rearrangements (CGRs) which can range from gross chromosomal events such as translocations and inversions to supernumerary ring/marker chromosomes, and small deletions or duplications. In order to fully understand the complexity of a specific event and its associated clinical consequences, it is imperative to locate the breakpoint junctions and to resolve the derivative chromosome structure. This task, however, often surpasses the capabilities of conventional short-read sequencing technologies. In contrast, emerging long-read sequencing techniques present a compelling alternative for clinical diagnostics. In Sweden, the Genomic Medicine Sweden (GMS) rare diseases consortium has validated PacBio Revio long read whole genome sequencing (lrWGS) for clinical digital karyotyping of unresolved CGRs nationwide. The first 20 samples included in this validation study were collected from six of the seven Swedish university regions. For most cases new blood samples were not available, and DNA extracted by standard clinical protocols stored for years in the clinical biobanks was used. The sequencing yield ranged from 65-111Gb of HiFi data per sample, correlating with extraction protocols and storage time for samples. We have established a national analysis pipeline and a shared database for variant calling and filtering, which enables both reference-based and *de novo* assembly analysis. Additionally, tools for analysis and filtering of differential methylation will be incorporated. The findings include a balanced translocation between 17p and the 19 centromere, previously undetected by both short read sequencing and optical mapping. Both derivatives were fully assembled after *de novo* analysis. The included validation samples also encompass inversions and complex CGRs. In conclusion, the GMS rare diseases consortium has successfully validated PacBio Revio lrWGS as a clinical analysis for chromosomal rearrangements.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session I

PB3252 † Neuron-specific chromatin disruption at CpG islands and aging-related regions in Kabuki syndrome

Authors:

L. Boukas[1], T. Luperchio[2], A. Razi[3], K. Hansen[4], H. Bjornsson[5]; [1]Children's Natl. Hosp./Johns Hopkins Univ., Washington, DC, [2]Johns Hopkins Sch. of Med., Ellicott City, MD, [3]Johns Hopkins Sch. of Med., Baltimore, MD, [4]Johns Hopkins Univ, Baltimore, MD, [5]Univ. of Iceland/Johns Hopkins Sch. of Med., Reykjavik, Iceland

Many Mendelian developmental disorders caused by coding variants in epigenetic regulators have now been discovered. Epigenetic regulators are broadly expressed, and each of these disorders typically exhibits phenotypic manifestations from many organ systems. An open question is whether the chromatin disruption - the root of the pathogenesis - is similar in the different disease-relevant cell types. This is possible in principle, since in all these cell-types the causative gene has the same kind of function (e.g. methylates histones) and is disrupted by the same germline variant. In particular, because intellectual disability is the most prevalent phenotype of these disorders and blood is easily accessible, the blood/brain comparison is of major interest. The blood/brain comparison can also shed light upon another question of much recent interest: whether the same locations that comprise blood-derived diagnostic episignatures can inform the chromatin basis of abnormal neurodevelopment, as has been hypothesized. Here, we focus on mouse models for Kabuki syndrome types 1 and 2, and characterize genome-wide chromatin accessibility changes using ATAC-seq in a well-powered experiment (12 mutant vs 11 wild-type mice). We find that the chromatin accessibility changes in neurons are mostly distinct from those in B or T cells from blood. Zooming into the KS1 blood episignature locations, we see that they are no more likely to be disrupted in neurons than random locations. By contrast, they are significantly disrupted in B/T cells, suggesting that this result is not due to human-mouse differences.

We find that the discrepancy between neurons and B/T cells is not because the neuronal abnormalities occur at regulatory elements that are only active in neurons. Neurons, but not B or T cells, show preferential chromatin disruption at CpG islands and at regulatory elements linked to aging, such as promoters of genes regulating the IGF1-receptor and mTOR pathways. This is true in both Kabuki syndrome types, pointing towards a unified pathogenesis at the chromatin level downstream of the distinct causative genes. Moreover, the accessibility changes in the mutant neurons suggest they are epigenetically “older” than their wild-type counterparts.

Our findings reveal that the same variant, in the same epigenetic regulator, has cellular-context-specific effects. They also provide novel insights into the molecular pathogenesis of Kabuki syndrome. Finally, they suggest that blood-derived episignatures, while accurate as a diagnostic tool, may not be well-suited for understanding the mechanistic basis of neurodevelopment in Mendelian disorders of the epigenetic machinery.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session II

PB3253 Next generation sequencing in the clinical setting: four years' experience from a private laboratory in Brazil.

Authors:

N. Sobreira[1,2], A. Planello[2], D. Alencar[2], D. Bermeo[2], J. Sobreira[2], M. Gomes[2], T. Cavalcanti[2], L. de Castro[2], E. Palmero[2], H. Galvão[2], R. Martin[3], C. Scapulatempo-Neto[2]; [1]Johns Hopkins Univ, Crownsville, MD, [2]Diagnósticos da América S.A., DASA, São Paulo, Brazil, [3]Johns Hopkins Univ, Crownsville, MA

Introduction: Next generation sequencing (NGS) technology has been progressively incorporated to clinical laboratory testing in Brazil. Current applications include detection of either germline or somatic variants, and laboratories may focus their analysis on single gene, multi-gene panels, exome or whole-genome sequencing. Objectives: In this study, we aimed to report the implementation of NGS tests in the DASA laboratory in Brazil, and discuss the types of NGS tests ordered per year and the diagnostic yield by disease group and by test type.Methods: We retrospectively analyzed the NGS results from 3025 patients investigated in our laboratory between 2018 and 2021. NGS methods included single gene, multi-gene panel and WES. A descriptive statistic was used to evaluate the diagnostic yield of NGS tests results, which were categorized as positive, inconclusive, or negative. The clinical indications were: neurodevelopmental disorders, late-onset neurological disorders, neuromuscular disorders, syndromic disorders, skin/hair disorders, hematological/immunologic disorders, endocrine disorders, ophthalmologic disorders, inborn errors of metabolism (IEM), skeletal disorders, cardiovascular disorders, reproductive planning, cancer, genitourinary disorders, gastrointestinal disorders, and hearing disorders. Results: Out of the total 3025 NGS tests, 420 were single gene, 1158 were multi-gene panel, and 1447 were WES. From 2018 to 2020, the most frequently ordered test was muti-gene panel, followed by WES. However, in 2021, WES became the most frequently ordered test in our laboratory routine. The primary clinical indication for WES were syndromic features (76.2%) and neurodevelopmental disorders (74%). In terms of diagnostic yield, WES had the highest detection rate (32.7%), the highest inconclusive rate and the lowest negative rate; nonetheless, the diagnostic yield remained relatively constant, and the negative results decreased with time. Among the clinical feature groups, hearing disorders (50%) and skeletal disorders (43.15%) had the highest diagnostic yield. When stratified by WES, skeletal and hearing disorders had 55% and 50% detection rate, respectively. The rate of positive secondary findings in WES was 3%, and regarding the detected variants, predicted protein-truncating accounted for 25% of the P/LP variants and 7% of VUS.Conclusion: Our data show that in a short period of time WES already became the most frequently order test clinical routine of DASA, a private laboratory in Brazil, for being more cost-effective due to its ultra-high throughput, scalability, speed, and higher diagnostic yield.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session III

PB3254 Non mosaic trisomy 22. A case report.

Authors:

C. Mutinda[1,2], J. Wachira[3], N. Gachara[4], T. Mutinda[3], W. Sikolia[5]; [1]Univ. of Nairobi, Nairobi, Kenya, [2]Pumwani Maternity Hosp., Nairobi, Kenya, [3]Paediatrician, Mp Shah Hosp., Nairobi, Kenya, [4]Paediatrician and Paediatric Cardiologist, Univ. of Nairobi, Nairobi, Kenya, [5]Obstetric Gynecologists and fetal maternal medicine, Aga Khan Univ. Hosp., Nairobi, Kenya

Background: Majority of trisomy 22 zygotes end up as spontaneous abortions. Postpartum survival ranges from 3-4 days to a maximum of 3 years. Most trisomy 22 errors occur in oogenesis, predominantly in the first meiotic division. We present a case of a complete trisomy 22 that lived for 52 days.Methods: A case report Results: Live female neonate born via caesarean section at 37 weeks gestation of nonconsanguineous parents who had 2 sons. Pregnancy was flagged by an ultra sound at 30 weeks that showed severe intrauterine growth restriction and unspecified cardiac anomalies. Amniotic Fluorescence In Situ Hybridization (FISH) ruled out T13, T18, T21 and sex chromosomal aneuploidy. Apgar score at birth was 6 1, 7 5, 910, birth weight was 1710g (<2.5th centile), length 40 cm (<3rd centile) and head circumference of 31cm (10th centile). Physical exam, the neonate had dolichocephaly, sparse hair, flat square face, full cheeks, short columella, depressed wide spaced nasal bridge, long and smooth philtrum, macrotia, low set ears posteriorly rotated with preauricular pits, short neck, narrow shoulders and widely spaced nipples. A cardiac echo revealed Double Outlet Right Ventricle (DORV) with normally related great arteries and an interrupted aortic arch type B. A chromosomal microarray reported Arr [GRCh37] 22q11.1q13.33(17153988_51178264) x3. A follow up karyotype confirmed Karyotype 47, XX, +22 Complete, not mosaic. Parental karyotype results Father 46, XY Mother 46, XX, 21ps+. The child had poor weight gain and respiratory distress. The parents gave a Do Not Resuscitate order on day 48 of life, the child died at the age of 52 days. Conclusion: Trisomy 22 though rare can be diagnosed in utero. Access to comprehensive prenatal genetic services are paramount to enable parents make informed reproductive decisions.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session I

PB3255 † Not just one: the utility of whole genome sequencing for making a dual molecular diagnosis

Authors:

M. Chau[1,2], C. Settler[2], W. Bi[1,2], H. Dai[1,2], C. Eng[1,2], P. Liu[1,2], L. Meng[1,2], N. Owen[1,2], R. Rigobello[1,2], L. Vossaert[1,2], C. Wu[1,2], F. Xia[1,2], X. Zhao[1,2]; [1]Baylor Coll. of Med., Houston, TX, [2]Baylor Genetics Lab., Houston, TX

Introduction: Whole genome sequencing (WGS) offers a comprehensive solution for genetic investigation by capturing a wide range of genomic variations in a single test. Conventional phenotype-driven and stepwise genetic testing is time-consuming and may yield inconclusive results. Moreover, the genetic investigation paradigm often stops at the point of identifying a single diagnosis, potentially overlooking additional underlying genetic etiologies. WGS is now often utilized as a first- or second-tier test for patients with multiple congenital anomalies (MCA), developmental delay and intellectual disability (DD/ID). Limited data exist on the frequency of patients receiving multiple genetic diagnoses pertaining to their phenotype.
Methods: This is a retrospective evaluation of WGS results at Baylor Genetics. Demographic data, clinical history, diagnostic findings and the frequency of patients who had multiple reportable molecular findings were investigated.
Results: Reportable multilocus findings were detected in 48 cases; including 13 cases with dual molecular diagnoses (pathogenic or likely pathogenic variants); 24 cases with one diagnosis in conjunction with reportable finding(s) in another locus [2 variants in an autosomal recessive (AR) locus or 1 variant in an autosomal dominant (AD) locus]; 7 cases with one molecular diagnosis and one reportable variant in an AR locus; and 4 cases with one molecular diagnosis and an actionable secondary finding. Among the 13 cases with dual diagnoses, the most common clinical indications included MCA (n=10), DD/ID (n=4), seizures (n=3), and failure to thrive (n=1). All 13 cases involved at least one AD gene(s) (AD+AD=3, AD+AR gene with two variants=7, AD+X-linked=2, AD+mitochondrial DNA=1) among the dual diagnoses. Copy number variants contributed to a high impact finding in 8 of the 13 cases. Parental samples were available in 8 cases which revealed that *de novo* findings accounted for at least one of the diagnostic variants in 7 trios.
Conclusion: This study demonstrates the utility of WGS to provide reportable multilocus findings. This finding is significant not only in ending diagnostic odysseys but also in demonstrating the need to account for heterogenous variant types and multiple molecular etiologies associated with complex patient phenotypes. In addition, WGS can also aid in optimizing management and expediting diagnoses of critically ill patients with complex phenotypes. Multiple molecular diagnoses will also improve the understanding about the clinical effects of multiple variants at more than one locus.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session II

PB3256 Novel deletion 19(q13.31q13.32) with syndromic presentation.

Authors:

J. Pappas[1], E. Moran[2]; [1]NYU Clinical Genetic Services, New York, NY, [2]NYU-Hosp Joint Diseases, New York, NY

19q13.32 microdeletion syndrome has been reported in four cases in the medical literature (Castillo A et al, 2014). Common features of these cases were facial asymmetry, ptosis, oculomotor paralysis, orofacial clefts, micrognathia, kyphoscoliosis, aortic defects and colonic atony. There is a ClinVar entry [VCV000057097.1] of a pathogenic deletion in 19q13.32 presenting clinically with genu valgum, impaired language development and ostium primum atrial septal defect (Kaminsky EB et al, 2011). We describe a 19 year old female with a del19(q13.31q13.32). The deletion in our case overlaps a portion of the deletion in the reported cases and includes the deletion in ClinVar. Our case was referred to us at 32 months because of macrocephaly, hypotonia, developmental delay, dysmorphic features, surgically repaired cleft soft palate, multiple resolving VSDs, G-tube, history of aspiration pneumonia and bilateral hip dysplasia. She was born full term via NSVD to a 15.5 year old primigravida after uncomplicated pregnancy. Birth weight was 3.295 kg and length 48 cm, appropriate for gestational age. Delayed developmental milestones prompted early intervention therapies and included: sitting at 9 months, cruising 31 months and no words at 32 months. We periodically followed the patient and at age 19 our case presented with minimal expressive language, obesity (BMI 32.39 kg/m^2), short stature (0.03 centile, Z = -3.42), obstructive sleep apnea, hyperphagia and aggressive oppositional behavior. She had menarche at 12 but amenorrhea since. She had right hip dislocation and left epiphysiodesis for leg length discrepancy. Her features were distinctive: short neck, small and low set ears with right preauricular pit, flat face, bilateral microphthalmia, exotropia with history of nystagmus, lateral eyebrow flaring, sparse scalp hair, tented upper lip, inverted nipples, left accessory nipple, small hands and feet, sacral dimple and thick nails with rapid growth. None of these features were familial. Whole genome CN/SNP microarray revealed arr19q13.31q13.32(49,662,092-52,484,949)x1 (NCBI36/hg18). Phenotyping similarities between our case and the cases in Castillo A et al, 2014 included developmental delays, behavioral disorders, congenital heart defects, orofacial clefts and skeletal abnormalities. These abnormalities are likely associated with deletion of the overlapping segment and further support their association with the 19q13.32 microdeletion syndrome. The facial features and minor malformations appear to be different in our case and the cases with del19q32.2 and suggest a unique syndromic association with the deletion 19(q13.31q13.32).

Session Title: Molecular and Cytogenetic Diagnostics Poster Session III

PB3257 Optical genome mapping as a potential routine clinical diagnostic method.

Authors:

H. Barseghyan[1,2,3], D. Eisenreich[1], H. Erdmann[1], A. Hallermayr[1], E. Lindt[1], K. Sendelbach[1], U. Koehler[1], T. Neuhann[1], A. Abicht[1,4], E. Holinski-Feder[5,6]; [1]MGZ - Medizinisch Genetische Zentrum, Munich, Germany, [2]Childrens Natl. Res. Inst., Washington, DC, [3]Dept. of Genomics and Precision Med., The George Washington Univ., Washington, DC, [4]Dept. of Neurology, Ludwig-Maximilians-Univ., Munich, Germany, [5]MGZ - Medizinisch Genetisches Zentrum, Munich, Germany, [6]Dept. of Med. IV, Ludwig-Maximilians-Univ., Munich, Germany

Both cytogenetics and molecular genetics methods have helped to identify genetic diagnoses for many disorders. However, despite this success, many remain undiagnosed due to our inability to sensitively identify structural variants (SVs) with currently available technologies, that are limited in resolution (karyotype), detection of only unbalanced events (microarray), targeted nature (FISH) or utilization of short reads (genome sequencing). Optical Genome Mapping (OGM) promises to address these limitations by capturing patterns of fluorescent labels within long DNA molecules (>150kbp) that allow for accurate identification of SVs starting from 500bp in size even in highly repetitive regions of the human genome.

We have performed validation of OGM using 72 samples for which diagnoses had been achieved with orthogonal methods. In total, OGM was 98% concordant with previous diagnoses only missing one translocation case due to breakpoints being in centromere and one mosaic variant in an FSHD case. We also tested 25 undiagnosed cases for which microarray and exome sequencing were uninformative to investigate the increased diagnostic yield provided by OGM.

To select for potential clinically significant variants, we filtered SVs based on quality, size, frequency, and gene overlap. We identified two likely pathogenic deletions involving two clinical genes, *CLTC* and *DHX30*. We also identified several potentially interesting insertions in disease causing genes needing further investigation. Lastly, we identified a translocation involving *SON* gene and mosaic deletion involving *TSC2* gene. Both SVs were diagnostic and were validated via targeted long-read sequencing. In summary, we show that OGM detects large SVs, has high concordance with other cytogenetic methods and can provide diagnosis in negative cases.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session I

PB3258 Optical genome mapping identified a *KMT2C* exonic deletion in a Kleefstra syndrome subject resulting from maternal mosaicism.

Authors:

B. Brassell, A. Ortega, C. Bilancia; Bionano Lab., Salt Lake City, UT

Recent advances in genomic technology have significantly increased diagnostic yield in patients with genetic disorders. Here, optical genome mapping (OGM) was used to further investigate an apparently de novo deletion in *KMT2C*. The proband is an 8-year-old male presenting with clumsiness, drooling, sensory disorder, speech articulation and developmental coordination disorder, fine motor delay, hyperactivity, facial weakness with asymmetry, and hypotonia. Family history indicates that both parents are apparently healthy, a maternal uncle presents with attention deficit hyperactivity disorder (ADD) and learning disabilities, and the proband's younger twin siblings were born prematurely at 30 weeks.

In the proband, OGM revealed a heterozygous 338 kbp deletion (within hg38 chr7:152,245,774-152,590,586 window) affecting exons 1-14 of *KMT2C* (NM_170606.3), which was previously reported as likely pathogenic by chromosomal microarray (CMA) in 2018. Disease-causing variants in *KMT2C* are associated with autosomal dominant Kleefstra syndrome (MIM: 617768). Further examination by a geneticist determined that the deletion is explanatory for his symptoms and is consistent with a diagnosis of Kleefstra syndrome. Parental testing by qPCR indicated this deletion to be de novo. However, OGM testing in the maternal sample revealed the 338 kbp *KMT2C* deletion with a lower variant allele frequency (VAF) of 0.28, suggesting mosaicism. A confirmatory CMA was performed on both the proband and maternal sample. This testing both confirmed the deletion in the proband and the deletion present in the mother that was not called by the CMA interpretation software and was identified by manual review by expert geneticists. Further follow-up studies are ongoing in this family and the younger twin siblings will be tested to determine the effect of the *KMT2C* deletion in this family.

In conclusion, this case demonstrated that OGM, as an advanced genomic technology, can be used to reveal the genetic etiology of diseases caused by structural variations (SVs) even with lower-level mosaicism. Mosaicism in unaffected parents of children with Kleefstra syndrome has been reported for the *EHMT1* gene (PMID: 29416845, 34357686), but not for *KMT2C*. Thus, the variant reported here is the first published for this gene.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session II

PB3259 Optimization of ORACollect buccal sample extraction for clinical molecular analysis.

Authors:

A. Berlyoung, A. C. Brand, S. C. Lynch, D. A. Johnson, J. Noya, J. Spangler, S. Yusuff, J. Devaney; GeneDx, Gaithersburg, MD

As a patient-centered health intelligence company, GeneDx continually focuses on improvements to laboratory processes. Efforts to streamline methods, reduce costs, and improve turnaround times provide critically important answers to patients. This will facilitate rapid diagnosis, leading to swift treatment and significantly improved prognosis. We recently launched a novel extraction method for our highest quantity sample, ORAcollect buccal (OCB), based on a commercially available kit (ThermoFisher Scientific).

The extraction utilizes automation for plate setup with the OCB samples and reagents. The entire method is performed in under two hours, requires no by-hand steps, and is of minimal cost ($3.70/sample). To measure quality and quantity of the DNA extraction, we utilized a UV method that determines the amount of DNA and any contaminants. To further scrutinize the quality of the extracted DNA, we analyzed the samples on massively parallel sequencing (gene panels, exomes, and genomes), microarrays, multiplex-ligation dependent probe amplification reactions, repeat expansion assays, and sanger sequencing. For all lab techniques, we used current measures of quality for each method to determine if the extracted DNA would pass in our clinical lab. The pass rate for the extracted samples was >95%.

With our modifications to the extraction method, we were able to consolidate resources and limit reagent waste. This modified laboratory procedure created a novel instrument protocol for extraction of OCBs that produces high quality DNA that is usable for all of our current molecular laboratory techniques. This protocol serves as a model archetype for future works to improve laboratory efficiency, simplify workflows, and provide rapid answers for patients.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session III

PB3260 Pathogenic tandem splice acceptor variants escape screening by*in silico*tools

Authors:

F. Frost[1], S. Gu[2,3], A. Elbert[4,5], M. Malicdan[1,6], C. Boerkoel[4,5], P. Cherukuri[2,3]; [1]NIH, Rockville, MD, [2]Imagenetics, Sanford Hlth., Sioux Falls, SD, [3]Sanford Sch. of Med., Sioux Falls, SD, [4]Dept. of Med. Genetics, Univ. of British Columbia, Vancouver, BC, Canada, [5]Provincial Med. Genetics Program, B.C. Women's Hosp. and Hlth.Ctr., Vancouver, BC, Canada, [6]Human Biochemical Genetics Section, Med. Genetics Branch, Natl. Human Genome Res. Inst., NIH, Bethesda, MD

Introduction. Tandem splice acceptor sites are a known mechanism of alternative splicing. Such tandem sites serve as loci of transcriptional regulation by modulation of either the transcript abundance or the function of the resulting protein. Previous literature has demonstrated that variants modifying existing tandem acceptor sites or creating novel sites can cause mendelian disease. Despite recent advances in *in silico* splicing prediction tools and understanding of tandem splice acceptors, predicting the pathogenicity or deleteriousness of variants remains difficult and, as a result, are rarely contextualized in clinical reporting. **Methods.** In this study we characterize the prevalence of variants creating novel tandem acceptor sites in healthy (gnomAD) and disease (ClinVar) reference populations. To characterize tandem splice acceptor variants further in the context of Mendelian disease, we screened for such variants among 1236 individuals from the NIH Undiagnosed Diseases Program (UDP). To evaluate the pathogenicity of those variants among130 individuals who had available RNA sequencing data, each novel tandem splice acceptor variant was assessed for evidence of function and pathogenicity. **Results.** Assessment of nucleotide trimer frequency at the exon-intron boundary showed a depletion of AG-containing trimers from position -3 to -14, a region that often overlaps the polypyrimidine tract, an acceptor site enhancer motif. SpliceAI scores of AG-gain variants in both ClinVar and gnomAD showed that acceptor gain scores are higher in the AG-depleted region compared to the rest of the intron and exon window. AG-gain variants within that same window showed similarly high SpliceAI acceptor gain scores in the UDP cohort but occur infrequently (mean = 10 per case). Evaluation of empirical RNASeq data for novel acceptor site usage in UDP cases showed that variants with verified usage were more abundant in the -3 to -14 window compared to the rest of the intron. Furthermore, although variants verified to create utilized novel acceptor sites tended to have high SpliceAI acceptor gain scores, they had a PPV of 0.344, demonstrating the incomplete ability of *in silico* tools to assess AG-gain variants.**Conclusions.** The pathogenicity of variants affecting splicing are difficult to assess. We show that a subset of those variants, those which create tandem splice acceptor sites, frequently lead to the usage of a novel acceptor site. Predicting the function of those variants using *in silico* tools is inconsistent. This supports the use of empirical evaluation of AG-gain variants by RNA sequencing.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session I

PB3261 PDIVAS: Pathogenicity predictor for Deep-Intronic Variants causing Aberrant Splicing

Authors:

R. Kurosawa[1], K. Iida[2], M. Ajiro[3], T. Awaya[1], M. Yamada[4], K. Kosaki[4], M. Hagiwara[1]; [1]Kyoto Univ., Graduate Sch. of Med., Kyoto, Japan, [2]Kindai Univ., Faculty of Sci. and Engineering, Osaka, Japan, [3]Natl. Cancer Ctr. Res. Inst., Tokyo, Japan, [4]Keio Univ. Sch. of Med., Tokyo, Japan

Deep-intronic variants that alter RNA splicing were ineffectively evaluated in the search for the cause of genetic diseases. Determination of such pathogenic variants from a vast number of deep-intronic variants (approximately 1,500,000 variants per individual) represents a technical challenge to researchers. Thus, we developed a Pathogenicity predictor for Deep-Intronic Variants causing Aberrant Splicing (PDIVAS) to easily detect pathogenic deep-intronic variants. PDIVAS was trained on an ensemble machine-learning algorithm to classify pathogenic and benign variants in a curated dataset. The dataset consists of manually curated pathogenic splice-altering variants (SAVs) and commonly observed benign variants within deep introns. Splicing features and a splicing constraint metric were used to maximize the predictive sensitivity and sensitivity, respectively. PDIVAS showed an average precision of 0.92 and a maximum MCC of 0.88 in classifying these variants, which were the best of the previous predictors. Genome sequencing analysis using PDIVAS with 95% sensitivity identified an average of 27 pathogenic candidates per individual. Furthermore, the causative variants were more efficiently prioritized than the previous predictors in simulated patient genomes. Incorporating PDIVAS into variant interpretation pipelines will enable efficient detection of disease-causing deep-intronic SAVs and contribute to improving the diagnostic yield. PDIVAS is publicly available at https://github.com/shiro-kur/PDIVAS.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session II

PB3262 Performance of Copy Number Variant Detection Tools in Clinical Applications: A Comparative Benchmark for Short-Read Whole Genome Sequencing

Authors:

F. De La Vega[1], S. Irvine[2], P. Anur[1], K. Potts[1], L. Kraft[1], R. Torres[1], S. Truong[3], Y. Lee[3], S. Han[3], V. Onuchic[3], J. Han[3], P. Kang[1]; [1]Tempus Labs, Inc., Chicago, IL, [2]Real Time Genomics, Ltd., Hamilton, New Zealand, [3]Illumina, Inc., San Diego, CA

As whole genome sequencing (WGS) becomes cheaper, it is set to become a standard approach in clinical settings due to its superior copy number variant (CNV) detection. Current tools for short-read WGS CNV calling need to be evaluated in clinical settings where orthogonal confirmation of CNVs may be required and there is a higher priority placed on sensitivity over specificity compared to research uses. We evaluated several CNV detection tools designed for short-read WGS data, including Delly, DRAGEN™ 4.0, CNVkit, CNVpytor, Lumpy, Manta, and Parliament2, as well as two newer tools: Cue, a machine learning-based method, and DRAGEN 4.2's integrated CNV caller that combines breakpoint and depth-based calls. We used data from independent PCR-free libraries of the HG002 reference cell line, sequenced to a mean depth of 50X using paired-end 2x150bp reads on Illumina NovaSeq™ 6000 and X-Plus instruments. Benchmarking CNVs is often aimed to evaluate event-level similarities, but in clinical contexts, the primary concern is whether a variant disrupts protein structure. Thus, to calculate accuracy we evaluated CNV overlaps with coding exons, defining a match as an event intersecting an exon with the same dosage direction as the truth set. The event's contribution is adjusted by the number of exons spanned to account for events overlapping multiple exons. Using GRCh37-defined exon boundaries, we confined our analysis to exons intersecting the HG002 GIAB v0.6 SV truth set for events of 1-100kb, including 13 deletions and 4 duplications that overlap 45 and 8 exons, respectively. Given the limited examples in the truth set, we placed simulated gene models across the truth set to increase deletion and duplication exon overlaps to 125 and 20, respectively. We show that all callers struggle with detecting single-exon events (typically <5kb) and duplications. However, DRAGEN 4.2 and Parliament2 demonstrated the highest combined sensitivity (86% and 94%, respectively), with the former showing superior specificity (66% vs 22%, respectively). Remarkably, despite a reduction in specificity (down to 42%), DRAGEN's v4.2 "high sensitivity mode" achieved 96% and 100% combined sensitivity for events >1kb and >5kb, respectively. This could be advantageous in clinical settings where sensitivity is paramount and confirmatory tests are performed.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session III

PB3263 Phenotypic Variations of a Novel TRMT10A Gene Mutation in a Consanguineous Malian Family

Authors:

M. Barros Perez, J. Lee, R. Vela Sosa, N. Khardoush, M. Banikazemi; NYC H+H/Metropolitan Hosp. Ctr., New York, NY

Microcephaly, short stature, and impaired glucose metabolism 1 (MSSGM1) (OMIM #616003) is a rare autosomal recessive condition caused by TRMT10A gene variations. MSSGM1's clinical manifestations include impaired glucose homeostasis or non-autoimmune diabetes, microcephaly, epilepsy, cognitive impairment, and failure to thrive. Less frequently, delayed puberty and joint laxity have been reported. To the best of our knowledge, a total of 20 cases within 13 families have been documented. In this report, we describe the variable phenotypic manifestations of the MSSGM1 syndrome in three siblings from one family, with a variation in the TRMT10A gene. Our case involves three of the six children from a family of African descent from Mali with third-degree consanguinity. The eldest sibling was referred for genetic evaluation at 4 years of age due to developmental delay, failure to thrive, and microcephaly. Initial genetic work-up including karyotype, SNP microarray, FISH, and metabolic testing resulted normal. In 2016, at 13 years of age, a repeat SNP microarray showed ROH of 233 MB. The middle sibling, also with microcephaly and developmental delay, had an SNP microarray with multiple ROH, and between these, the mutual ROH was located in Chromosome 4q21.23q24. After reviewing genes in this locus, only the gene TMRT10A was a candidate for sequencing, which revealed an autosomal recessive homozygous nonsense mutation c.483G>A (p.Trp161*), which had not been reported previously. Soon after, the middle sibling was found to have the variation, and eight years later, the youngest sibling, who had intrauterine growth restriction and microcephaly, also tested positive for the homozygous mutation. The affected siblings, a currently 20-year-old female, a 17-year-old male, and a 10-month-old female, demonstrated intrauterine growth restriction, microcephaly and progressed to have failure to thrive. Microcephaly has been seen in around 85% of reported individuals. Short stature was observed in the elder two siblings, reported in approximately 94% of the cases. All of the siblings exhibited developmental delay. Glucose homeostasis disturbances observed were neonatal hypoglycemia and type 2 diabetes mellitus in the eldest and neonatal hypoglycemia in the youngest sibling. The middle one was not affected by diabetes. Delayed puberty was observed in the older two siblings. Unique features of the eldest sibling were seizure, central hypothyroidism, and joint laxity.
We discuss the variability of the MSSGM1 phenotypes within the members of a family. The scarcity of reported cases underlines the need for further exploration of this rare genetic condition.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session I

PB3264 Physical linkage to resolve complex variants in the X-chromosome opsin locus

Authors:

A. Naik[1], B. Guan[1], D. Elul[2], E. B. Haggerty[3], W. Zein[1], C. Lwin[1], Z. Li[1], D. Blain[1], B. P. Brooks[1], S. J. Tumminia[1], G. Kerry[1], D. H. Brainard[4], G. K. Aguirre[5], R. B. Hufnagel[1]; [1]Natl. Eye Inst., NIH, Bethesda, MD, [2]2Dept. of Psychology, Univ. of Pennsylvania, Philadelphia, PA, [3]3Dept. of Neurology, Univ. of Pennsylvania, Philadelphia, PA, [4]Dept. of Psychology, Univ. of Pennsylvania, Philadelphia, PA, [5]Dept. of Neurology, Univ. of Pennsylvania, Philadelphia, PA

The X-opsin array encodes for red and green opsin and is associated with X-linked color vision deficiencies, high grade myopia, Bornholm eye disease (BED), X-linked cone dystrophy (XLCD), and blue cone monochromacy (BCM). The X-opsin array consists of highly homologous *OPN1LW* (L) and *OPN1MW* (M) genes arranged in multiple tandem copies, increasing the probability a gene conversion event or that unequal crossover occurs. Current gold standard sequencing technologies are unable to resolve the variation between gene copies that may contribute to disease. Here we present a novel method to resolve complex genotypes of the X-opsin array. We designed a tiered approach to assess copy number, structural, and sequence variants to comprehensively genotype the X-opsin array. The test involves digital droplet PCR (ddPCR) copy number analysis for the locus control region (LCR), L and M genes, the recurrent inactivating variant c.607T>C p.(Cys203Arg) (C203R), and exon-3 haplotypes. Proximal and distal gene(s) are genotyped separately by long-range PCR followed by nested PCR and Sanger sequencing or next-generation sequencing. The ddPCR milepost assays were employed to determine gene order. Nanopore long-read sequencing was used for validation. Male patients (n=31) with X-opsin array disorders from 29 families were used to demonstrate the utility of the test. Pathogenic genotypes of 100% (29/29) probands with X-opsin array disorders were resolved. Of these, seven required milepost assays to determine gene arrangement. Of 17 males with color discrimination disorder, 14 deutans and three protans were molecularly confirmed by haplotype and spectrally sensitive amino acid residues. Three probands with BED, XLCD, or cone-rod dystrophy had splice-altering exon-3 haplotypes. Of nine BCM probands, six had LCR deletions, and three had p.C203R in expressed gene copy(s), including two brothers harboring p.C203R in three of four gene copies in which the C203= allele could not be resolved by long-read sequencing alone. The ddPCR-centered diagnostic test provides an economical and efficient tool for genotyping the X-opsin array, elucidating genotype-phenotype correlations, and informing future gene-directed therapies. The milepost assay is critical for the determination of genotypes of expressed genes on the array in a subset of patients. The applicability of the assay, however, is dependent on DNA quality and integrity. The ddPCR milepost approach could be applied to other genomic regions that are difficult to genotype with current technologies.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session II

PB3265 Precise and reliable detection of regions of homozygosity and copy-neutral loss of heterozygosity in the human genome

Authors:

L. Wu, B. Bolstad, J. McGirr, A. Roter, D. Black, J. Wallace; Thermo Fisher Scientific, SANTA CLARA, CA

Human genome often harbors regions that are in allelic imbalance or homozygosity without an associated DNA copy-number change. This type of genomic events can occur in somatic cells or in the germline. In the absence of a copy number variation, this is referred to as copy-neutral loss of heterozygosity (CN-LOH) or as a region of homozygosity (ROH). In this study, for simplicity, we use the term of loss of heterozygosity (LOH) to stand for both ROH and CN-LOH. Genome-wide detection of LOH with high accuracy and reproducibility plays a vital role in understanding tumorigenesis in neoplastic disorders and unravelling genetical causes of constitutional disorders. We employed a newly developed chromosomal microarray analysis (CMA) platform to ascertain LOH at a genome-wide scale. This new CMA has a shorter laboratory turnaround time of two days and expanded genomic content. It provides a whole-genome copy-number profile of a sample that includes neutral copy numbers and genome-wide SNP genotyping, which allows an effective LOH detection. The sensitivity of the new array platform was measured on 49 HapMap cell line samples. These samples have 200 known LOH regions with reportable length greater than 3 megabases (MB). We conducted LOH calling and visual confirmation using B-allele frequencies (BAF) in version 4.4.1 of the Chromosome Analysis Suite (ChAS) software. The LOH-calling algorithm in ChAS is based on a binomial statistical test that adapts to SNP call rate in the sample. The new array platform reaches 98.5% sensitivity in detecting LOH greater than 3MB in length and demonstrates very high reproducibility. DNA from sample NA18550 with twelve reportable LOH regions was run by two different operators on different instruments and dates. ChAS analysis shows that the results of both runs target the same set 12 LOH regions with 100% concordance. In addition to these results on cell line DNA, the accuracy of LOH calling of the new array platform using prenatal samples will be reported.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session III

PB3266 † Precision Genomics in 9p Minus Syndrome

Authors:

T. Turner, E. Sams, K. Valle, S. Couteranis, R. Slaugh, J. Neidich, X. Cui, F. Cole, P. Dickson, R. Mitra, J. Milbrandt; Washington Univ. Sch. of Med., St. Louis, MO

Complex structural variation underlies several rare syndromes. One example is 9p minus syndrome, which is caused by deletions on the p-arm of chromosome 9 with co-occurring translocations in 50% of affected individuals. In collaboration with the Chromosome 9 p minus network, we have used several strategies to assess the genomic features of 9p minus syndrome including our published results on karyotypes in >700 individuals with 9p minus syndrome and our unpublished results on short read whole genome sequencing (WGS) in 42 individuals, PacBio HiFi long-read WGS in 10 individuals, and Illumina Complete Long Read WGS in 14 individuals. It is known that the deletions in 9p minus syndrome do not share one specific set of breakpoints. In our large-scale karyotype analysis, we found that the most common deletions include the 9p24 locus and the second most common include 9p24 to 9p22. Furthermore, by assessing translocation information from the karyotypes, we found that there is no preference for a specific chromosome arm; rather, there are examples of individuals with 9p minus syndrome with structural variants involving several different chromosome arms, respectively. By WGS and overlapping deletion assessment, we have identified one of at least two hypothesized critical region(s) for 9p minus syndrome. This region encompasses the gene RFX3, deleted in 80% of individuals with 9p minus syndrome. We also utilized a machine learning based approach to evaluate gene dosage-phenotype correlations among individuals with 9p minus syndrome, compared with control individuals from the 1000 genomes project. This analysis also revealed *RFX3* as the top gene implicated in 9p minus syndrome. Previously, *RFX3* has been implicated in autism, and we hypothesize that it may be involved in the developmental delay phenotype in 9p minus syndrome. Beyond the identification of genes shared across all individuals with 9p minus syndrome, we have also utilized long read WGS to precisely catalogue all genomic variation within individuals with 9p minus syndrome. This includes complex variation involving translocations, deletions, and mosaic duplications within single individuals. Combining several modern genomic technologies may yield insights into 9p minus syndrome and may be broadly important for the genomic study of syndromes with complex structural variation.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session I

PB3267 Proteome-scale probabilistic modeling of human genetic variation discovers novel developmental disorder genes

Authors:

R. Orenbuch[1], J. Frazer[2], M. Dias[2], H. Spinner[1], D. Franceschi[1], D. Marks[1]; [1]Harvard Med. Sch., Boston, MA, [2]Ctr. for Genomic Regulation, Barcelona, Spain

Our inability to identify causal variants across patients' genomes is a major limitation in genome sequencing for diagnostics. This is especially problematic for variants in genes not previously associated with disease. To overcome this issue, we introduce the first unified model of genetic variation across the full human proteome. Our model, popEVE, enables us to rank all variants according to their pathogenicity and to obtain a measure of missense constraint at variant-level resolution. We find this model to significantly outperform all other models, including meta-predictors, at distinguishing patients with developmental disorders from unaffected individuals, based on their de novo mutations, while also achieving state of the art performance at predicting readouts from deep mutational scanning experiments. Finally, we argue that our model can play a valuable role in identifying many rare genetic disorders which may never be discoverable with burden tests. We provide evidence that there are likely thousands of such disorders, some of which are present in the 31k trio meta-cohort, partially explaining the current low yield of patient sequencing in the context of rare disease.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session II

PB3268 Rapid DNA extraction from buccal swab samples: enhancing efficiency and performance.

Authors:

L. Willing[1], N. Seip[2], M. Salazar-Rondon[2], L. Torres-Benito[2]; [1]BioEcho Life Sci., Inc, Wouburn, MA, [2]BioEcho Life Sci. GmbH, Cologne, Germany

The extraction of high-quality DNA from buccal swab samples is of utmost importance in various fields of genetic research, forensic science, and medical diagnostics. However, conventional DNA extraction methods often involve time-consuming protocols and provide suboptimal results in terms of DNA concentration, yield, and integrity. This can significantly hinder downstream applications such as PCR testing, short tandem repeat (STR) analysis, and genetic sequencing. In this study, we present a novel and efficient method for rapid DNA extraction from buccal swab samples: a tailored lysis, followed by a single purification step. The primary objective of this research was to streamline the DNA extraction process while maintaining high-quality DNA suitable for subsequent approaches. Dry buccal swab samples were collected from a diverse set of individuals, and DNA extraction was performed with our novel method and with a silica-based technique. Our results demonstrated significant improvements in multiple aspects of DNA extraction. Firstly, the total processing time was reduced by 85 % compared to the conventional method, allowing for faster sample throughput. Secondly, the DNA extraction using our method resulted in comparable DNA concentration and superior yield. Notably, our technology ensured better preservation of DNA integrity, as evidenced by higher DIN values obtained. To validate the extracted DNA for downstream applications, we performed STR analyses. The results obtained from the DNA extracted with our method exhibited concordance with a standard DNA extraction method, demonstrating the suitability of the extracted DNA for molecular genetic analysis. In conclusion, the presented EchoLUTION technology presents a significant advancement in the extraction of DNA from dry buccal swab samples. The reduction in processing time, coupled with the enhanced efficiency of our technology, ensures increased sample throughput and enables more accurate and reliable downstream applications, such as STR analysis, genetic investigations, and further microarray analysis, and sequencing.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session III

PB3269 Relative contribution of variant prioritization and phenotype similarity to performance of automatic genome analysis algorithms

Authors:

S. Khanam[1], A. J. Marcello[1], Z. Soens[2], V. Ustach[2], J. X. Chong[1,3]; [1]Univ. of Washington, Seattle, WA, [2]GeneDx, Gaithersburg, MD, [3]Brotman-Baty Inst., Seattle, WA

Genome analysis algorithms (GAAs) attempt to automate analysis of patient phenotype and sequence data to identify pathogenic variants in individuals with suspected Mendelian conditions (MCs). Most successful GAAs such as XRare, LIRICAL, and Exomiser combine variant-driven and phenotype-driven prioritization, in which patient phenotypic similarity is compared to a reference set of known MCs. Performance is assessed in small cohorts, typically obtained from gene discovery or clinical diagnostic research programs or in sets of published cases. For example, the performance of LIRICAL was assessed in a 100,000 Genomes Project cohort (N=116) and XRare was assessed in a Melbourne Genomics Health Alliance Cohort (N=59). These cohorts may not be representative of the heterogeneous set of individuals receiving genetic testing. Most affected individuals worldwide lack access to similar clinical expertise, even in high resource settings, especially as genetic testing makes inroads into diverse areas of medicine in which clinicians do not necessarily have explicit genetics training (e.g., neurology, cardiology). Moreover, phenotype data available to diagnostic testing laboratories may be extracted computationally, with variable accuracy, from a diverse set of sources such as medical records and notes, test requisition forms, etc. We assessed GAA performance in >20,000 individuals with a wide variety of categorical phenotypes (e.g., autism, skin disorders, seizures, deafness) in whom diagnostic testing identified a Likely Pathogenic or Pathogenic (LP/P) variant(s) that was likely explanatory. We used phenotype data submitted during testing and considered the gene with the reported LP/P variant to underlie the correct diagnosis. Because variant-driven prioritization in all algorithms relies on pathogenicity predictors, a fast-evolving area of research, GAAs may differ in performance purely due to external factors (e.g., date of last update). To understand the impact of phenotype detail and the efficacy of GAAs operating on phenotype alone, we analyzed our cohort without including sequence data. GAAs ranked the correct gene highest in only ~0.3-3% of cases and among the top 10 genes in ~3-13% of cases. We are in the process of assessing performance when GAAs use sequence data simultaneously with phenotype. Our analyses highlight a gap in performance by GAAs that seemingly perform well in small cohorts with phenotype data collected in ideal conditions but are compromised when heterogeneously collected phenotype data are used, as well as their dependency on variant prioritization to achieve good performance.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session I

PB3270 Reporting the common African PMS2 variant c.2182_2184delACTinsG results with 100% confidence.

Authors:

T. Wijesuriya[1], A. Smith[1,2], L. Huang[1,2], W. Qin[1]; [1]Children's Hosp. of Eastern Ontario, Ottawa, ON, Canada, [2]Univ. of Ottawa, Ottawa, ON, Canada

The *PMS2* gene has 15 exons and encompasses ~37 kb of genomic sequence on chromosome 7. A partial duplication of the *PMS2* gene, as a part of a larger repeat element, results in a transcribed pseudo copy of *PMS2*, termed *PMS2CL*. *PMS2CL* consists *PMS2* exons 9 and 11-15, whereas exon 10 is missing due to an Alu-mediated deletion. Sequencing analysis of *PMS2* is extremely difficult due to these highly homologous regions between *PMS2* and *PMS2CL*. Currently in our lab, we use a custom designed Next Generation Sequencing (NGS) panel (Roche, KAPA HyperPlus) as the initial test for all patients referred for hereditary cancer testing. Read alignment, single nucleotide variants and small insertion/deletions calling are performed using NextGENe software. This can be technically challenging for genes like PMS2 that are prone to misalignment Due to the presence of high sequence homology, any variants detected by NGS within exons 11-15 of *PMS2* are confirmed by long range PCR (LR-PCR) amplifying *PMS2* exons 11-15 followed by Sanger sequencing. Only variants confirmed by *PMS2* LR-PCR and Sanger sequencing are reported. Variants that are not confirmed are assumed to be in *PMS2CL*. However, we haven't confirmed that these variants are in fact in *PMS2CL* using a LR-PCR assay targeting the pseudogene. We have detected c.2182_2184delACTinsG variant by NGS, which is in exon 13 of either *PMS2* or *PMS2CL*. This is a truncating variant with an allele frequency of ~2.5% in African population (gnomAD). Distinguishing the origin of this variant is very important as if it is detected in the *PMS2* gene, it is considered to be pathogenic; however, if it is in the *PMS2CL* pseudogene it is considered as having no clinical impact. This variant was not confirmed by our *PMS2* LR-PCR and Sanger sequencing analysis, suggesting it is a variant in *PMS2CL*. In order to further verify this variant is in *PMS2CL*, we designed highly specific LR-PCR primers that only amplify the *PMS2CL* pseudogene. We were able to confirm this c.2182_2184delACTinsG variant is indeed in the *PMS2CL* pseudogene using the *PMS2CL* LR-PCR and Sanger sequencing analysis in 8 patients in which the variant was detected by NGS but was not confirmed by the *PMS2* LR-PCR and Sanger sequencing. Being able to perform LR-PCR for both *PMS2* and *PMS2CL* simultaneously for variants detected by NGS allows us the added confidence in distinguishing the origin of any variants discovered in these highly homologous exons and to report them appropriately.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session II

PB3271 Resolution of non-coding variants by RNA sequencing in rare disease.

Authors:

A. Marshall[1], G. Del Gobbo[1], Y. Liang[2], M. Couse[2], K. M. Boycott[1], K. D. Kernohan[3]; [1]Children's Hosp. of Eastern Ontario Res. Inst., Univ. of Ottawa, Ottawa, ON, Canada, [2]Ctr. for Computational Med., The Hosp. for Sick Children, Toronto, ON, Canada, [3]Children's Hosp. of Eastern Ontario Res. Inst., Univ. of Ottawa; Newborn Screening Ontario, Ottawa, ON, Canada

The clinical implementation of next-generation sequencing has revolutionized genetic diagnostics of rare disease, yet many exome and genome analyses remain inconclusive. This is due in part to challenges with variant prioritization and interpretation, particularly for intronic variants. The accuracy of *in silico* tools to predict the effect of a variant on RNA expression and/or splicing outside the canonical splice site remains low and therefore these variants are almost always variants of uncertain significance (VUSs).

Following inconclusive genetic testing, families affected with undiagnosed rare disease consented to enrollment in the Care4Rare Canada research program and RNA sequencing (RNA-Seq) analyses were performed on affected individuals and has been useful for resolution of VUSs in a number of projects we have analyzed. For example, an affected individual with suspected short-rib thoracic dysplasia 3 (SRTD3) had clinical genetic testing which identified a likely pathogenic frameshift variant and an intronic VUS in *DYNC2H1* in trans. Research RNA-Seq investigations revealed significantly decreased *DYNC2H1* gene expression in the proband as well as a novel splice junction as a result of the intronic variant, which again allowed reclassification of the variant to LP. In a second study, an individual with medium-chain acyl-coenzyme A dehydrogenase (MCAD) deficiency had only one likely pathogenic variant detected in *ACADM* by clinical sequencing. RNA-Seq detected a novel intron inclusion in this gene; an SNV was detected within this intron inclusion, was confirmed to be heterozygous in the individual's genomic DNA and was classified as likely pathogenic. Finally, in two siblings with suspected Joubert syndrome, RNA-Seq was also able to identify their second causative variant. This variant was intractable to detection by routine exome and genome sequencing analyses as it is a 57 bp deletion in a repetitive intronic region, but was detectable by RNA-Seq as it leads to novel intron inclusion in the resulting transcript.

These studies highlight the benefits of RNA-Seq in obtaining a diagnosis for patients with rare disease where additional data is required after inconclusive clinical testing, especially for variants with a potential mRNA splice impact. While we recognize that non-coding variants play a role in disease, we currently lack robust tools to accurately interpret and classify them. RNA-Seq can assess the impact of these non-coding variants via their effect on gene expression, RNA stability and RNA processing and therefore could be considered as a follow up clinical test in the future.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session III

PB3272 RNA Sequencing as a Routine Diagnostic Tool in Medical Genetic Practice, Solving A Case with Gaucher Disease

Authors:

H. Onay[1,2], E. Goksoy[3]; [1]Multigen Genetic Diseases Diagnosis Ctr., Izmir, Turkey, [2]Gene2Info Sag.Hiz., Istanbul, Turkey, [3]Dept. of Pediatry, Pediatric Metabolic Diseases Subdivision, Adnan Menderes Univ., Faculty of Med., Aydın, Turkey

In routine medical genetic practice sequencing is the main molecular genetic technique for diagnosis of the genetic disorders and sequencing of the coding exons and exon-intron boundaries is the main sequencing approach. DNA is the main genomic material used in sequencing in routine practice. In recent years RNA sequencing was began to use as a diagnostic tool in limited number of diagnostic genetic laboratories for limited number of genetic disorders. The main barriers for RNA sequencing in routine practice are rapid degradation of RNA and expression of the RNA in certain tissues. In metabolic genetic disorders such as Gaucher, Pompe and Fabry RNA sequencing can be performed from the blood because there is limited expression of the genes responsible for metabolic genetic disorders in the blood. A 6-year-old girl was referred to our laboratory due to bicytopenia. Her motor and mental milestones were normal. She had suffered from hepatosplenomegaly since the age of 2 years. Physical examination revealed a pale appearance, distended abdomen, palpable liver and spleen up to the groin, and prominent veins on the anterior abdominal surface. She had anaemia, thrombocytopenia, moderate elevation in liver function tests and high ferritin. In metabolic tests; chitotriosidase level (809.8 µmol/L.hr; NR 200.0) and lyso-GB1 level (832.1 ng/mL; NR <14.0) were quite high. Leukocyte beta-glucosidase levels were found to be low (0.40 nmol/mg/hr; NR 1.00). Ultrasound revealed massive hepatosplenomegaly. With all these findings, the patient was clinically diagnosed with Gaucher disease type 1. *GBA* sequencing from DNA was performed. Heterozygous c.1126A-G [p.N409S-N370S)] pathogenic variant was detected. GBA Whole Gene Sequencing (including all introns) and MLPA tests were performed in order to find the second causal variant in the patient but turned to be negative. Then the RNA sequencing was performed from the blood. Interestingly second causal pathogenic variant, c.475C-T (p.R159W), was detected in the RNA in heterozygous state. Heterozygous c.1126A-G [p.N409S-N370S)] variant was also detected in the RNA. Segregation analysis was performed. Father is carrier for the c.1126A-G [p.N409S-N370S)] variant but mother is carrier for the c.475C-T (p.R159W) variant in RNA. After re-analyses of the sequencing data from DNA obtained from blood, c.475C-T (p.R159W) variant was detected but allele frequency is 5%. Eventually, the patient was genetically diagnosed and received treatment. This patient shows the importance of RNA sequencing in routine practice. Single step RNA sequencing from the blood might be an optimal solution for metabolic genetic disorders.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session I

PB3273 RNA-based diagnosis of a child with SMA disease

Authors:

S. Durmus[1], U. Guven Gulhan[1], C. Kolakoglu[1], O. Akdas[1], I. Ovali[1], E. Evin[2], F. E. Sevilgen[1], E. Nikerel[1], P. Pir[1], M. Adli[3], E. Unsal[2], L. Ozer[2], S. Aktuna[2], A. Ulgenalp[4], U. Yis[4], V. Baltaci[2], A. O. Caglayan[4], T. Cakir[1]; [1]PhiTech AS, Kocaeli, Turkey, [2]Mikrogen Ltd, Ankara, Turkey, [3]PhiTech Inc, Chicago, IL, [4]Dokuz Eylul Univ., Izmir, Turkey

Monogenic rare diseases challenge the patients, families, and healthcare systems with approximately 5 years spent on average to achieve a successful diagnosis. The conventional methods are based on the detection of disease-causing mutations using gene panels or exome sequencing, and still, they fail to provide a successful diagnosis for more than 50% of the patients.

Here, we report a case for which diagnosis was not possible using the conventional methods for years after the emergence of the abnormality in muscle functions. Whole Exome Sequencing was previously applied to the blood samples from the family, consisting of healthy parents and the diseased child. A homozygous variant with uncertain significance (VUS) was detected in the *SMN1* gene of the child. The same variant was detected in the mother as heterozygous, while the father had a deletion in the related variant region. However, evidence of loss of function was needed to demonstrate the pathogenicity of the detected VUS. RNA sequencing (RNA-Seq) provides information on the genes with aberrant expression, which is not possible to obtain via genome sequencing. Therefore, RNA-Seq was applied to the blood samples from the family, and expression levels were identified at the gene and transcript levels. Transcript level expression analysis showed that, in the mother and father, no significant loss of expression was observed in the transcripts of the *SMN1* gene. However, there was a 100% loss of expression in the canonical and biologically relevant transcript of the gene in the child. Functional analysis via RNA-Seq demonstrated the pathogenic effect of the identified VUS in the function of SMN1 protein, leading to SMA disease. A muscle biopsy from the child was further subjected to RNA-Seq and transcript-level expression analysis, confirming the complete loss of expression of the *SMN1* gene in the muscle tissue. Consequently, the RNA-based diagnosis of the patient was approved by the clinical committee, and drugs for treatment were prescribed. While DNA holds genetic code, the functional effects of genetic variants can only be detected through the functional genome, and we here demonstrated that RNA has the power to capture the functionality of the protein coded by the candidate gene, *SMN1*.

We propose the use of our approach in rare disease diagnosis, particularly for patients who could not be diagnosed with genome profiling. In the near future, we expect to demonstrate the success of our approach with many other cases.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session II

PB3274 RNA-seq based rare diseases diagnostics at the Pan-European consortium Solve-RD

Authors:

V. Yepez[1], C. Mertes[1], K. Ellwanger[2], R. Luknarova[1], C. Doornbos[3], A-S. Denomme-Pichon[4], A. Vitobello[4], D. Beijer[5], M. Synofzik[5], D. Mei[6], R. Guerrini[6], B. Estévez-Arias[7], D. Natera-de Benito[7], A. Segarra[8], J. Müller[9], A. Topf[8], L. Matalonga[10], G. Demidov[2], S. Beltran[10], A. Hoischen[3], P-B. t'Hoen[3], A. Esteve-Codina[10], J. Gagneur[1]; [1]Technical Univ. of Munich, Garching bei München, Germany, [2]Univ. of Tübingen, Tübingen, Germany, [3]Radboud Univ. Med. Ctr., Nijmegen, Netherlands, [4]CHU Dijon Bourgogne, Dijon, France, [5]Hertie Inst. for Clinical Brain Res., Tübingen, Germany, [6]Meyer Children's Hosp. IRCCS, Florence, Italy, [7]Hosp. Sant Joan de Déu, Barcelona, Spain, [8]Newcastle Univ. and Newcastle Hosp. NHS Fndn. Trust, Newcastle, United Kingdom, [9]Univ. Coll. London, London, United Kingdom, [10]CNAG-CRG, Ctr. for Genomic Regulation (CRG), Barcelona, Spain

RNA sequencing has emerged as a powerful complementary approach to DNA sequencing to discover disease-causing gene regulatory defects in individuals affected by genetically undiagnosed rare disorders. However, as large rare disease research consortia are implementing RNA-seq-based diagnostics, new data analysis challenges arise to cope with the heterogeneity of diseases, various tissues and sample sizes, and the multiplicity of interpreters. Here we present new algorithms and methodologies that have been developed by the RNA-seq working group of the Pan-European consortium Solve-RD to address these issues. The Solve-RD RNA-seq dataset consists of >450 RNA-seq samples collected from 4 different tissues, from individuals tested at more than 20 clinics belonging to 6 European Reference Networks. The vast majority of the samples also have matching whole exome or whole genome sequencing. The samples were analyzed using the workflow DROP v1.3.3, which includes the recent release of the aberrant splicing caller FRASER 2.0 to call aberrant events in RNA-seq data. Two new RNA-seq quality control modules were added to DROP to confirm sex and tissue annotations. Both helped to detect a handful of mislabelled samples. To identify candidate genes and variants, we considered i) a DNA-first approach, in which candidate variants from DNA sequencing are functionally validated by RNA-seq and ii) an RNA-first approach in which candidate genes are first identified by aberrant expression or splicing and DNA data is then re-examined. The two approaches proved to be complementary and reported on average, a handful of candidates per sample. The DNA-first approach benefits from a much reduced multiple testing burden. It falls under the ACMG guidelines, whereby the RNA-seq is considered as a functional assay. We devised annotation guidelines for the RNA-first approach that integrates scoring of an aberrant expression or splicing event at four levels: i) raw data support of the outlier calls, ii) gene function, iii) identification of a variant with a plausible regulatory mechanism, and iv) variant segregation. To further increase the consistency of our annotations, we have conducted an in-presence workshop named Solvathon, combining training and collective data inspection sessions, and offered weekly consultancy hours. As of June 2023, the RNA-seq analysis has led to 5 new solved cases (4 harboring a deletion and 1 a deep intronic variant missed by exome sequencing), and a handful of strong candidates. Extensive case-by-case analysis is ongoing. Altogether, our new algorithms and processes contribute to improved methodologies for RNA-seq-based rare disease diagnostics.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session III

PB3275 Selected aspects of the process of implementation of tandem repeats characterisation into routine laboratory testing processes based on massively parallel sequencing

Authors:

I. Lojová[1,2,3], M. Kucharik[2,4], Z. Pös[1,2,4], A. Zatkova[1,4], E. Tothova Tarova[1], J. Budis[4,2,5], T. Szemes[2,3,4], J. Radvanszky[2,3,1,4]; [1]Inst. of Clinical and Translational Res., BioMed. Res. Ctr. of the Slovak Academy of Sci., Bratislava, Slovakia, [2]Comenius Univ. Sci. Park, Bratislava, Slovakia, [3]Dept. of Molecular Biology, Faculty of Natural Sci., Comenius Univ., Bratislava, Slovakia, [4]Geneton Ltd., Bratislava, Slovakia, [5]Slovak Ctr. of Scientific and Technical Information, Bratislava, Slovakia

Introduction: The introduction of massively parallel sequencing (MPS) methods initiated the rapid development of methods for genotyping sequence variants at the genomic level. The evaluation of tandem repeat (TR) motifs has become possible with some delay, mainly due to their high variability and complexity. We are reporting on specific qualitative and quantitative measures of reliability in the context of characterizing TR motifs in genomic diagnostics. **Material and Methods:** Whole genome sequencing data, using short-read MPS, were generated for 62 individuals. For characterisation of appr. 50 clinically relevant TR loci we used the modified version of a TR dedicated tool Dante. **Results:** According to us, for efficient genotyping of TR loci there are certain aspects of quality management, which may yet differ from the actual recommendations for SNVs. The most important parameters include: 1) uniformity of locus definition and nomenclature (may be more complex than for SNVs); 2) sufficient depth of locus coverage (higher than for SNVs); 3) the possibility of visual verification (dedicated tools with more complex functions); and 4) laboratory validation of relevant findings (may be still necessary, using alternative methods, in certain specific cases). **Conclusion:** The MPS is currently an advanced methodology, which in the case of SNVs has been refined and translated into clinical practice. The characterization and assessment of TRs presents multiple challenges in both secondary and tertiary data analysis, which means that the official guidelines currently applicable to SNVs will need to be extended to include interpretation and reporting of TRs. **Grants:** This work has been supported by the PANGAIA project H2020-MSCA-RISE-2019 (Grant agreement ID: 872539) funded under H2020-EU.1.3.3. Programme; by the ALPACA project H2020-MSCA-ITN-2020 (Grant agreement ID: 956229) funded under H2020-EU.1.3.1. Programme; by the Operational Program Integrated Infrastructure within projects ITMS: 313011F988 and ITMS:313021BUZ3, co-financed by the ERDF; and by the Scientific Grant Agency (VEGA_2/0146/23).

Session Title: Molecular and Cytogenetic Diagnostics Poster Session I

PB3276 † Semiautomated Interpretation of Genomic Disease Contribution to Unknown Causes of Infant Mortality.

Authors:

L. Protopsaltis[1,2], R. Baer[3,4], M. Bainbridge[1,2], B. Cao[1,2], Y. Ding[1], K. Ellsworth[1,2], L. Forero[1,2,3], E. Frise[5], L. Guidugli[1,2], Y. H. Kwon[1,2], J. Le[1,2], E. Ontiveros[1,2], M. J. Owen[1,2], E. Sanford[1,2], L. Van Der Kraan[1,2], M. Wright[1,2], M. Yandell[1,5], L. Jelliffe-Pawlowski[4], G. Bandoli[3], C. Chambers[3], S. F. Kingsmore[1,2]; [1]Rady Children's Inst. for Genomic Med., San Diego, CA, [2]Rady Children's Hosp., San Diego, CA, [3]Dept. of Pediatrics, Univ. of California San Diego, San Diego, CA, [4]California Preterm Birth Initiative, Univ. of California San Francisco, San Francisco, CA, [5]Fabric Genomics, Oakland, CA

Quantitative understanding of the causes of infant mortality shapes public health, surveillance, and research investments. However, the contribution of single-locus genetic diseases to infant mortality is poorly understood. We recently reported that 46 (41%) of 112 infant deaths within a large pediatric hospital system in San Diego County (SD), CA, were associated with genetic diseases using whole genome sequencing (WGS) and semiautomated interpretation. Subsequently, we performed post-mortem WGS (>30-fold, Illumina) of de-identified, archived, newborn dried blood spots collected from 2005-2018 in an additional 914 SD infant deaths. Here we report initial results of diagnostic analysis of the first 211 of these infant deaths. Death certificates listed ICD-10 code R95 (sudden infant death syndrome) as the cause of death in 175 (83%) of 211 infants, and R99 (ill-defined and unknown cause of mortality) or W75 (other accidental threats to breathing) in the remaining 36 (17%). Variants identified with DRAGEN v3.9 underwent semiautomated interpretation with Fabric Genomics' Artificial Intelligence Classification Engine (ACE) and GEM with the HPO phenotype "death in infancy". Variant diplotypes in disease-associated genes with positive Bayesian GEM scores were reviewed manually according to ACMG classification guidelines. 71 (34%) of 211 infant deaths were associated with at least one pathogenic (P) or likely pathogenic (LP) variant, of which <5 were assessed to have had autosomal dominant (AD) disorders associated with infant mortality, such as Left Ventricular Noncompaction (OMIM IDs 615396, 615092). The remaining 69 infant deaths either had P/LP variants in genes associated with later-onset AD disorders or represented carrier status for autosomal recessive (AR) disorders. Analysis of variants associated with AR, mitochondrial, and X-linked disorders in addition to copy number and structural variants is ongoing and will be reported together with results of manual curation of variants where ACE or GEM did not report. We will also evaluate this cohort for likely deleterious variants that occur recurrently and that have biologic plausibility as novel causes of infant mortality. It is much more difficult to identify genetic disorders that potentially contributed to infant mortality in WGS of de-identified samples than in the previous cohort with deep phenotypes derived from electronic health records. However, ACE and GEM are effective for automated prioritization of P and LP variants in disease-associated genes. Together these studies should clarify the contribution of single-locus genetic diseases to infant mortality in Southern California.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session II

PB3277 Single exonic CNV analysis uncovers masked complex genome rearrangements in recessive disorder

Authors:

T. Naruto, Y. Kuroda, Y. Saito, Y. Enomoto, K. Kurosawa; Kanagawa Children's Med. Ctr., Yokohama, Japan

Advances in next-generation sequencing and genome capture technologies have made exome sequencing a cost-effective approach for mutation detection in inherited diseases. However, computationally predicting copy number variants (CNVs) from exome sequence data is a challenging task. Numerous programs are available, but with varying sensitivities and less sensitivity to detect smaller single-exon CNVs. Bardet-Biedl syndrome (BBS) is a genetically heterogenous disorder characterized by retinopathy, obesity, polydactyly, urogenital abnormalities, and intellectual disability. BBS is a multisystemic disorder requiring lifelong medical care and support. Because of the variable expressivity of BBS, a molecular genetic approach combined with clinical diagnosis is crucial. The proband was an 11-year-old boy. He was born at 39 weeks of gestation to nonconsanguineous parents. At birth, he was noted to have bilateral postaxial polydactyly. From the age of three years, he developed progressive vision loss and night blindness and was diagnosed clinically with Bardet-Biedl syndrome. Whole exome sequencing revealed a pathogenic SNV of ClinVar in BBS9. Since it was consistent with the patient's clinical symptoms and recessive inheritance, we searched for a causative variant in the other allele. three CNV detection programs (XHMM, CoNIFER, and CNVnator) failed to find CNV in BBS9. A detailed investigation of the CNVs of each exon converted from read depth log 2 ratio, we identified a deletion of exon 17 as a candidate variant. To elucidate the deletion interval, we employed Nextera XT (Illumina) sequencing of a PCR product of approximately 22 kb containing exons 16 to 18. We identified a deletion of 3.6 kb and insertions of uncertain origin in the adjacent 239 bp and 22 bp. From this result, it was inferred that exon 17 was skipped, resulting in a codon stop and causing the onset of BBS. These results indicated that the deletions of exons may have potential complex rearrangements around them.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session III

PB3278 Somatic genetic rescue in a patient with RUSAT2 syndrome carrying a *MECOM* LoF variant

Authors:

A. Buijs[1], S. Holwerda[1], R. Bladergroen[2], L. Suurenbroek[2], L. van Vulpen[1], M. Bartels[1], R. Kuiper[2], M. Jongmans[1]; [1]Univ. Med. Ctr. Utrecht, Utrecht, Netherlands, [2]Princess Máxima Ctr. for Pediatric Oncology, Utrecht, Netherlands

We present a unique case of clonal hematopoiesis of indeterminate potential (CHIP) by means of copy neutral loss-of-heterozygosity (CN-LOH) of the long arm of chromosome 3, potentially resulting in a rescue mechanism for MECOM loss-of-function (LoF) in radioulnar synostosis with amegakaryocytic thrombocytopenia-2 (RUSAT2) syndrome (OMIM 165215). A 20 years-old woman was diagnosed with RUSAT2 syndrome, caused by a *de novo* heterozygous missense variant in the *MECOM* gene. Since diagnosis, mild cytopenias remained stable (hemoglobin 7.0-8.4 mmol/L; leukocytes 3.2-5.1x10^9/L; platelets 127-161x10^9/L). Routine bone marrow (BM) analyses showed decreased cellularity. Single nucleotide polymorphism (SNP) array analyses on BM DNA samples were performed for surveillance of genomic alterations associated with progression and/or malignant transformation, requiring stem cell transplantation. We detected acquired CN-LOH of 66 Mb of the long arm of chromosome 3 (band 3q22.1->3qter) including *MECOM*, with an increasing percentage of mosaicism in time up to ~10%. Locus-specific amplicon-based deep sequencing (4,500X) demonstrated an identical quantitative shift in time in favor of the wildtype *MECOM* allele. Our data indicate that the expanding clone in the BM cells has undergone mitotic recombination resulting in wildtype *MECOM* on both alleles. Somatic genetic rescue has been reported in other bone marrow failure (BMF) syndromes, conferring selective advantage of cells that lost the germline pathogenic variant(s) over cells carrying the pathogenic allele(s). We are, for the first time, witnessing expanding CHIP in RUSAT2 syndrome, suggesting somatic genetic rescue for MECOM LoF. We will continue to follow the course of the CN-LOH 3q and its consequences in relation to the mild cytopenias in this case. Our study underscores the importance of periodic genomic screening of BM in BMF syndromes in clinical decision-making.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session I

PB3279 Spontaneous induction of Chromosome 1 Trisomy in reprogrammed iPSC line is associated with accelerated growth and Proliferation.

Authors:

C. Pearce[1], S. Borrowman[1], S. Ternes[2], A. Kamath[2], J. Nagy[1], M. Schultz[1], M. Tung[1], B. Chandra[1]; [1]Univ. of Iowa, Iowa City, IA, [2]Cellular Engineering Technologies, Coralville, IA

Induced pluripotent stem cell (iPSCs)-based model systems allow for a wide array of studies, and especially the direct reprogramming of patient-derived cells into iPSCs further enables the implementation of personalized medicine through their broad applications in drug discovery, regenerative medicine, and disease modeling. Although iPSC reprogramming methods have been well established and highly reproducible over the past decade, reprogrammed iPSC cell lines must be grown and maintained in artificial culture media supplemented with multiple growth factors for an extended time period, which may result in cellular or genetic aberrations, such as spontaneous differentiation of the cells or a spontaneous gain or loss of chromosomes. Here we report a rare occurrence of spontaneous gain of an extra copy of chromosome 1 in a patient derived Gaucher disease iPSC line. The parent fibroblasts were successfully reprogrammed using Episomal reprogramming factors, and pluripotency was confirmed by both immunohistochemistry and Flow Cytometry. This gain of chromosome 1 was discovered after 15 passages when the cell line was sent for karyotyping. A karyotype performed on the parent of origin fibroblasts was normal. The mechanisms of these cellular and genetic aberrations are not well studied but are typically attributed to suboptimal culture techniques and replication stress. Replication stress can lead to replication fork stalling, which results in double-strand DNA breaks. iPSCs also undergo rapid replication by avoiding certain cell cycle checkpoints, which is also a likely cause of the genetic aberrations the cells could acquire, particularly when compared to somatic cell lines. In humans, trisomies and other aneuploidies are known to cause multiple congenital anomalies and developmental delays and, depending on the chromosome, can often lead to miscarriages. Pathogenic mechanisms of certain trisomies are known or are being studied, such as trisomy 21 (down syndrome), 18 (Edwards syndrome), and trisomy 13 (Patau syndrome), and are believed to be due to maternal meiotic non-disjunction. However, there is limited information available on Trisomy 1, due to its constitutional non-viability. Our reprogrammed iPSC cell lines grew rapidly with a significantly faster doubling time in comparison to the other lines that were being grown concurrently, leading us to believe that trisomy 1 has some effect on growth. This may allow for further investigation into the pathogenic mechanisms and consequences of trisomy 1 in somatic disorders such as cancers.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session II

PB3280 Subtle Cellular Phenotypes Inform Pathological and Benign Genetic Mutants in the Iduronidase-2 Sulfatase Gene

Authors:

W. Buchser[1], S. Elia[1], S. Hurt[1], A. Viswanathan[1], S. Le[1], K. Kelley[1], B. Doray[1], P. Dickson[2]; [1]Washington Univ. in St Louis, Saint Louis, MO, [2]Washington Univ. Sch. of Med., Saint Louis, MO

BackgroundInterpreting the pathogenicity of genetic variants in molecular DNA testing for inherited conditions presents significant challenges, leading to many "variants of uncertain significance" (VUS). Mutations in the IDS gene, implicated in mucopolysaccharidosis type II (Hunter syndrome), where deficiency of the lysosomal enzyme iduronate-2-sulfatase (I2S) impairs the degradation of specific glycosaminoglycans, requiring enzyme replacement therapy. Despite available biochemical assays, numerous IDS VUS remain unclassified, highlighting the need for innovative classification methods. MethodsWe devised an evaluative platform, Deep Cellular Phenotyping, leveraging high-content imaging and machine learning to generate composite phenotypes based on cellular morphology. This platform enables individual cell sequencing after imaging to identify cellular variants. The approach was applied to the IDS gene for functional variant classification.ResultsAfter introducing a dozen IDS mutations into the A549 cell line, imaging revealed distinct differences in lysosomal structures, dye accumulation in intracellular vesicles, and the distribution of a LAMP1-GFP reporter. These image-based features facilitated differentiation between known mutant and wild-type cells. Notably, cells with pathogenic IDS mutations showed a reversal of these anomalies upon exposure to recombinant I2S enzyme. Only a fraction of the cells was rescued after 72-96 hours, implying some long-term damage to the organelles in these particular cells.ConclusionOur findings suggest image-based screening as a promising tool for detecting subtle distinctions, potentially paving the way for new methods of variant classification and elucidating a broader spectrum of variant effects.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session III

PB3281 Targeted sequencing of 50+ pathogenic repeat expansions using Twist target enrichment and PacBio HiFi sequencing.

Authors:

T. Han[1], H. Corbitt[1], E. M. Lampraki[2], F. Tornos[1], S. Holt[2], D. Stucki[2], E. Toro[1], S. Kingan[2], C. Locklear[1]; [1]Twist BioSci., South San Francisco, CA, [2]Pacific BioSci.s, Menlo Park, CA

Background/objectives: The expansion of unstable genomic short tandem repeats (STRs) has been identified as the causal DNA mutation in more than 30 Mendelian diseases. PCR-based DNA fragment analysis, short-read sequencing, or legacy molecular genotyping methods like Southern blot are frequently used to analyze STR expansions, but are not capable of determining the exact length and sequence composition. Targeted sequencing allows for high-resolution characterization of dozens to thousands of gene regions at a scale and cost that is more accessible than whole genome sequencing. **Methods:** Here we describe the performance of a Twist target enrichment panel sequenced with PacBio long HiFi reads to measure the size of STR expansions with base-pair resolution. Using a proprietary algorithm, we designed a gene panel targeting 50+ regions with known pathogenic repeat expansion alleles in human samples. We started with 200 nanograms of fragmented gDNA. After end-repair, A-tailing, and adapter ligation, unique dual indices for sample barcoding were added during PCR. Every 8 samples were pooled for an overnight hybridization. After capture and wash, the post-capture libraries were then amplified and converted into SMRTbell library and sequenced on PacBio Sequel IIe system (also compatible with the Revio system) with resulting HiFi read lengths of ～5 kilobases. We genotyped repeat copy numbers using Tandem Repeat Genotyping Tool (TRGT) and visualized reads spanning repeats using TRVZ. **Results:** We benchmarked performance of the long-read capture workflow in reference samples, of which 7 have characterized pathogenic repeat expansions alleles with diverse repeat base composition. Preliminary analysis showed two CAG repeats in the *HTT* and *AR* genes, a CTG repeat expansion in *DMPK*, a GAA repeat expansion in *FXN*, and GC-rich expansions in several other genes could be captured and sequenced. We also tested four different commercially available PCR polymerases and compared sequence coverage across the samples and targets for each condition. We found concordance between observed and expected repeat number for the genes *HTT*, *DMPK*, *AR*, and *PABPN1*. Optimizing the coverage of a few genes and testing on clinical samples with expanded alleles are ongoing. **Conclusion:** From this pilot study, we provide design and workflow guidance to researchers interested in targeted long-read sequencing for scalable and cost-efficient hybrid capture of 50+ different repeat expansions with long read lengths, minimizing coverage bias, and maximizing accuracy to fully capture all variant types. *[For Research Use Only. Not for use in diagnostic procedures.]*

Session Title: Molecular and Cytogenetic Diagnostics Poster Session I

PB3282 The diagnostic yield of plasma-based cell free DNA in individuals with matched tissue-based venous and arteriovenous malformations

Authors:

D. Jensen[1], N. Au[1], M-H. M. Pham[1], N. Parmalee[1], V. Dmyterko[1], T. T. Cook[1], K. Zenner[2,1], S. Wilcox[1,3], C. T. Myers[4], M. A. Gillentine[4], G. Strub[5,6], R. A. Bly[2,7], G. M. Shivaram[8], G. Richter[9,6], J. A. Perkins[2,7], J. T. Bennett[10,7,11]; [1]Seattle Children's Res. Inst., Seattle, WA, [2]Dept. of Head & Neck Surgery, Univ. of Washington, Seattle, WA, [3]Univ. of Michigan, Ann Arbor, MI, [4]Seattle Children's Hosp., Seattle, WA, [5]Arkansas Children's Res. Inst., Seattle, AR, [6]Univ. of Arkansas, Fayetteville, AR, [7]Vascular Anomalies Program, Seattle Children's Hosp., Seattle, WA, [8]Dept. of Radiology, Univ. of Washington, Seattle, WA, [9]Arkansas Children's Res. Inst., Little Rock, AR, [10]Seattle Childrens Res. Inst., Seattle, WA, [11]Dept. of Pediatrics, Div. Genetic Med., Univ. of Washington, Seattle, WA

Purpose: Most venous and arteriovenous malformations (VeM and AVM) are caused by somatic activating mutations in oncogenes. Molecular diagnosis currently requires surgically excised tissues, but there is increasing demand for non-invasive molecular diagnostics due to the development of targeted medical therapies. Here we report the diagnostic yield of cfDNA in a cohort of 61 individuals with matched tissue and plasma samples. We also compared diagnostic yield of peripheral blood and "lesional" blood collected directly from the vascular malformation. **Methods:** We included only individuals with known tissue-based pathogenic mutations in *PIK3CA*, *TEK*, *MAP2K1*, *BRAF*, or *KRAS*. cfDNA was extracted from the plasma and mutations were assayed using digital droplet PCR or high depth custom NGS panel. **Results:** Mutations were identified in plasma based cfDNA in a third of the cohort (21/61 individuals). The diagnostic yield was higher in VeM (16/36 individuals) than in AVM (5/25 individuals). Surprisingly, the diagnostic yield of intralesional plasma samples was higher (10/17 VeM and 1/3 AVM) than peripherally collected plasma (8/35 VeM and 4/25 AVM). All plasma samples were also compared to matched cellular "pellets," where mutations were detected in 5/36 individuals with VeM and in 0/25 individuals with AVM. In a single individual with an AVM, the mutation was not detected in plasma collected prior to glue embolization, but was detected in a peripheral sample collected immediately post-embolization. **Conclusion:**In a cohort of 61 individuals with known tissue-based somatic mutations as gold-standard comparators, we demonstrate the utility of plasma based cfDNA as a non-invasive diagnostic analyte. Overall, mutations were detected in cfDNA in 44% of our VeM cohort and 20% of our AVM cohort. The detection rate in plasma cfDNA is superior to that in the cellular ("blood pellet") fraction (13% VeM and 0% AVM), likely reflecting an enrichment of endothelial cell derived DNA in the plasma. Direct sampling of intralesional plasma increased the diagnostic yield (58% VeM and 1/3 AVM) over peripheral plasma, which is also consistent with an endothelial cell origin of these mutations. Although the diagnostic rates in cfDNA are inferior to that in tissues, our results suggest a route to molecular diagnosis, and companion targeted medical therapy, in individuals with VeM or AVM for whom surgical tissues are unavailable. Our results also support the collection of plasma for diagnostic testing during embolization or sclerotherapy procedures.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session II

PB3283 The first case of a Korean patient with the *WAC* germline mutation c.451C>T(p.Arg151*)

Authors:

C. Cheon, I. Kim; Pusan Natl. Univ. Children's Hosp., Yangsan-si, Korea, Republic of

Backgrounds: DeSanto-Shinawi syndrome is an autosomal dominant genetic condition caused by loss-of-function mutations in *WAC* and is characterized by a multitude of dysmorphic features along with behavioral abnormalities, and intellectual disability. Congenital heart defects associated with DeSanto-Shinawi syndrome were rarely reported in patients with pathogenic variants in *WAC*. **Methods:** We performed WES on DNA obtained from the peripheral blood of the patient and his parents, and variants found were compared and filtered. The coding and flanking intronic regions were enriched using Agilent solution (Agilent Technologies) and sequenced using the HiSeq 2500/4000 system (Illumina). Conventional Sanger sequencing was used to sequence at least 1 causative or rare variant for a second, independent confirmation. **Results:** Here, we report a case of a 13-year-old Korean girl patient with dysmorphic facial features such as a prominent forehead, high arched palate, anteverted nares, low nasal bridge, long philtrum, micrognathia, and intellectual disability. She was born at 40+3 weeks, 3.14kg via cesarian delivery. She underwent coartoplasty and mitral valve valvotomy with papillary muscle splitting. She had medullary nephrocalcinosis, VUR (left grade III), and epilepsy. Brain MR revealed several small microbleeds in the cerebrum on the SWI image and left mild periventricular leukomalacia. Whole exome sequencing revealed a novel heterozygous variant, c.451C>T (p.Arg151*), in the *WAC* gene. **Conclusions:** To the best of our knowledge, this is the first case of DeSanto-Shinawi syndrome in Korea, which expands the phenotypic and genotypic spectrum of this rare syndrome and provides deeper insights by delineating the clinical features of our patient.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session III

PB3284 The NIH INCLUDE Project: Transforming research for people with Down syndrome across the lifespan.

Authors:

M. Parisi[1], S. Bardhan[1], M. Brown[2], L. Garcia[1], **H. Li**[2], A. Mazzucco[3], G. Pearson[2], L. Ryan[4], C. Schramm[2], B. Schwartz[2], E. Tarver[4]; [1]NICHD/Natl. Inst.S OF HEALTH, Bethesda, MD, [2]NHLBI/Natl. Inst.S OF HEALTH, Bethesda, MD, [3]Office of Director/Natl. Inst.S OF HEALTH, Bethesda, MD, [4]NIA/Natl. Inst.S OF HEALTH, Bethesda, MD

Introduction: In response to a Congressional directive in fiscal year (FY) 2018, the NIH launched the INCLUDE (INvestigation of Co-occurring conditions across the Lifespan to Understand Down syndromE; https://www.nih.gov/include-project) Project. INCLUDE is studying conditions that affect those with Down syndrome (DS) and the general population, such as Alzheimer's disease, autism, celiac disease, congenital heart disease, and diabetes, and expanding the number of investigators engaged in DS research. Methods: The INCLUDE Project supports DS research through three major components: (1) Conduct targeted, high-risk, high-reward basic science studies on chromosome 21; (2) Assemble a large study population of individuals with DS; and (3) Include individuals with Down syndrome in existing and future clinical trials. There are currently 19 active Funding Opportunities, including five focusing on career development and fellowship awards, and several focusing on increasing diversity. The INCLUDE Data Coordinating Center (DCC; https://includedcc.org/) is promoting data sharing through its cohort-building efforts and by creating tools for data visualization and analysis through the INCLUDE Data Hub (https://portal.includedcc.org/login). The DCC has launched a training initiative for diverse data scholars through a summer internship program. In addition, the DS-Connect® registry (https://dsconnect.nih.gov/) is enhancing research participation by connecting families with research opportunities of interest to them. Results: NIH has invested $258 million in the past 5 years on 269 new INCLUDE projects. In FY2022, the INCLUDE investment was $75 million for total NIH DS funding of $113 million. Over three dozen trainees have been supported during the past 5 years. DS-Connect® has over 5800 registered participants and has promoted research enrollment for 100 projects, 18 INCLUDE-funded, and 5 clinical trials. The INCLUDE DCC launched the INCLUDE Data Hub with clinical and 'omics data from multiple studies, and currently includes over 2600 whole genomes. A DS Research Plan was recently published outlining goals and objectives for the next 5-10 years of DS Research. Conclusions: The INCLUDE Project has stimulated DS research funding since its launch, while the DS-Connect® registry and INCLUDE DCC are facilitating subject recruitment. A major cohort and biobanking initiative will be launched in FY24. Efforts to increase the diversity of research participants and investigators through community outreach are essential to create a more representative cohort that will lead to improved treatments and quality-of-life for those with DS.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session I

PB3285 The Role of Large Copy Number Variants in The Etiology for Chronic Kidney Disease Revealed through A Commercial Next-Generation Renal Gene Panel (Renasight™)

Authors:

F. Fang[1], J. Fierro[1], B. Sapra[1], J. Choi[1], S. Punj[2], **Y. Xue**[1]; [1]Fulgent Genetics, Temple City, CA, [2]Natera, Inc, San Carlos, CA

Introduction Disease-causing genetic variants are present in ~10% of adults and ~20% of children with chronic kidney disease (CKD). Large copy number variants (LCNVs) are associated with Mendelian disease, common traits, and sporadic diseases, including CKD. However, the role of LCNVs in the etiology for CKD in clinical practice is not yet well characterized. **Method:** A commercially available next-generation sequencing (NGS)-based 385 gene kidney disease panel test was performed for clinical purposes on patients seeing general and transplant nephrologists in the US. We used an in-house developed algorithm for CNV analysis. 1103 consecutive unique patient samples with deletion/duplication of >100kb and chromosomal aneuploidies were analyzed. CNVs were classified and Pathogenic (P) and Likely Pathogenic (LP) variants were reported. Positive findings included a monoallelic P/LP variant in an autosomal dominant or X-linked gene and biallelic P/LP variants in autosomal recessive genes and well-documented large chromosomal structural alterations. **Result:** Positive LCNV findings were identified in 26% (287/1103) of cases. A total of 42 positive results were identified across 18 chromosomes. Most frequently occurred positive results were: 17q12 microdeletion (12.7%, 140/1103, including *HNF1B*), 47, XXY (2.63%, 29/1103), 4q22.1 microdeletion (1.81%, 20/1103 including *ABCG2, DMP1, PKD2*), 45, X (1.09%) and 47,XXX (1.09%). Multiple diagnosis due to large CNV were identified in 3.35% (37/1103) of cases. Trisomy 12p, trisomy 13, trisomy 15 and trisomy 18 have been identified in 1, 3, 2, and 1 case respectively. Some of them were detected as mosaic trisomy and were reported as incidental findings. **Conclusion:** LCNV plays an important part in attributing to inherited kidney diseases. Detecting LCNVs in clinical exome-based broad kidney disease panel benefits the diagnosis, prognosis and management of CKD patients.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session II

PB3286 Transcriptional landscape of *MLIP* in pediatric and adult myopathy patients.

Authors:

A. Gagné[1,2], S. Audet[1,2], J. Mezreani[1,2], L. Hamilton[1], V. Triassi[1], A. Rihoux[1,2], E. O'Ferrall[3], J. Karamchandani[4], S. Donkervoort[5], C. Bonnemann[5], M. Tetreault[1,2]; [1]CHUM Res. Ctr., Montreal, QC, Canada, [2]Dept. of NeuroSci.s, Univ. of Montreal, Montreal, QC, Canada, [3]Dept. of Neurology and Neurosurgery, Montreal Neurological Inst., Montreal, QC, Canada, [4]Dept. of Pathology, Montreal Neurological Inst., Montreal, QC, Canada, [5]Neuromuscular and Neurogenetic Disorders of Childhood Section, Neurogenetics Branch, Natl. Inst. of Neurological Disorders and Stroke, Natl. Inst.s of Heallth, Bethesda, MD

Progress in genomic research has been driven by technological innovation, such as high-throughput sequencing. Despite great progress, there is a significant number of patients without a molecular diagnosis. One reason for this diagnostic shortfall is the challenge associated with interpretation of clinically relevant variants. To address this, *in vivo* or *in vitro* validation is required. Recently, bi-allelic variants in the gene *MLIP* (Muscular LMNA-interacting protein) were identified. Little is known about the biological role of MLIP, except the fact it interacts with *LMNA* (Lamin AC) a known muscle disease gene. *MLIP* has a complex transcriptional architecture with at least 7 isoforms with a tissue-specific expression pattern. Thus, the interpretation of MLIP variant is even more difficult and requires functional validation. Using long-read sequencing we uncovered unannotated transcripts of MLIP as well as a compensatory alternative event in an adult patient affected with a distal myopathy. Since biallelic variants have been identified in both pediatric and adult patients, we believe that difference in transcriptional landscape could explain, at least partially, the variability in age-of-onset and disease severity. To answer this question, we performed long-read sequencing on pediatric patients and compared the different splicing events. A better characterization of MLIP isoforms in muscle will contribute to our understanding of disease variability and ultimately will increase our understanding of MLIP's function in muscle.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session III

PB3287 Trio analysis reveals *de novo* rare variants in Congenital Hypogonadotropic Hypogonadism

Authors:

J. ZHAI[1,2], Y. Zouaghi[1,2], F. De Azevedo Correa[1,2], A. Boizot[1], R. Quinton[3], M. Bonomi[4,5], G. Szinnai[6,7], V. Lozovanu[8], F. Phan Hug[1], M. Adamo[1,2], A. Jacobs[1,2], G. Papadakis[1], N. Niederlander[1], A. Messina[1,2], F. Santoni[1,2], J. Acierno[1], I. Habibi[1], N. Pitteloud[1,2]; [1]Service of Endocrinology, Diabetology and Metabolism, Lausanne Univ. Hosp., Lausanne, Switzerland, [2]Faculty of Biology and Med., Univ. of Lausanne, Lausanne, Switzerland, [3]Dept. Translational & Clinical Res. Inst., Univ. of Newcastle-upon-Tyne, Newcastle-upon-Tyne, United Kingdom, [4]Dept. of Endocrine and Metabolic Diseases and Lab of Endocrine and Metabolic Res., IRCCS Istituto Auxologico Italiano, Milan, Italy, [5]Dept. of Med. Biotechnology and Translational Med., Univ. of Milan, Milan, Switzerland, [6]Dept. of Pediatric Endocrinology and Diabetology, Univ. Children's Hosp. Basel, Univ. of Basel, Basel, Switzerland, [7]Dept. of Clinical Res., Univ. Hosp. Basel, Univ. of Basel, Basel, Switzerland, [8]"Iuliu Hatieganu" Univ. of Med. and Pharmacy Cluj-Napoca, Cluj-Napoca, Romania

Background: Congenital Hypogonadotropic Hypogonadism (CHH) and Kallmann Syndrome (KS), an anosmic subtype of CHH (MIM147950), constitute rare genetic disorders defined by delayed puberty, infertility and GnRH deficiency. Despite the identification of numerous disease-associated genes, the genetic etiology remains elusive in approximately 25% of cases. **Objectives**: The aim of this study was to investigate the genetics of CHH in a cohort of well-characterized trios (n=21) using whole-genome sequencing (WGS). **Methods**: The WGS data from all trios were processed through a Sentieon-empowered pipeline, which included sequence alignment, variant calling, and subsequent annotation by ANNOVAR. Prioritization of candidate variants was conducted using VariantMaster. **Results**: So far, we have identified 4 heterozygous *de novo* variants in known CHH genes: one in *FGF17* and three in *FGFR1*. In addition, we have identified two putative *de novo* mutations in *BICRA* and *BARHL1*, genes not previously know to be associated with CHH.In a patient with KS, we found a *de novo* variant in *BICRA*, a gene involved in chromatin remodeling complex-associated protein. *BICRA* mutations have been previously implicated in Coffin-Siris Syndrome. In addition, we identified a *de novo* variant in *BARHL1* in a CHH patient. *BARHL1* is a homeobox gene, that is known to promote neurogenesis, neural survival and neuronal migration in developing brain and is a candidate gene for Joubert Syndrome. To investigate the potential involvement of these novel genes, we performed variant analysis in a cohort comprising 483 CHH patients. Remarkably, we identified 2 CHH patients harboring rare heterozygous variants in *BARHL1* and 1 CHH patient harboring a heterozygous rare variant in *BICRA* . **Conclusion**: Trio analysis has been fruitful to identify *de novo* variants in CHH using WGS. To validate these two putative candidate genes, further studies including functional analysis is required.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session I

PB3288 Trisomy of 4p and partial monosomy of 15q due to unbalanced translocation in an infant with multiple congenital anomalies - A case report.

Authors:

K. KARAKUS GULEN, N. Hara, X. Tang, C. Kumar; Westchester Med. Ctr., Valhalla, NY

Derivative chromosomes are often caused by unbalanced chromosome rearrangements that appear spontaneously or may be inherited from healthy parents carrying balanced reciprocal translocations. Trisomy 4p has been reported sporadically in the literature with variable phenotypes such as neurodevelopmental disorders, intellectual disability, poor growth, microcephaly, musculoskeletal anomalies and dysmorphic features. Deletion of 15q26 is a relatively rare chromosomal disorder, and it is described only in few cases. The severity of the 15q deletion phenotype depends on the size and location of the deletion that may include developmental delay, intellectual disability, dysmorphic features and skeletal abnormalities. Here we present a new case of a male infant born at 35 weeks gestation to a 25-year-old mother. The infant's birth weight was 1.67 kg (~3rd percentile) and birth length was 42.5 cm (~1st percentile) with multiple congenital anomalies, tracheomalacia VSD, PTO, bicuspid aorta and aortic coarctation, hypospadias, vertebral and digital anomalies. The chromosomal microarray and the cytogenetic investigations were indicated. Chromosome analysis was performed on GTG banded metaphases prepared from cultured lymphocytes that revealed a derivative chromosome 15 with an additional unidentified material at 15q region and the infant's karyotype as 46,XY,add(15)(q26.?1). The microarray analysis performed on whole blood showed a 48.8 megabase duplication from 4p16.3 to 4p11 and a 2.8 megabase terminal deletion of 15q26.3. The FISH studies using 4p and 15q subtelomere probes identified an unbalanced translocation between 4p and 15q regions. These findings confirmed that the additional material at 15q was from chromosome 4p representing the trisomy of the 4p region and the terminal deletion of 15q represented partial monosomy of 15q region. The duplicated interval from 4p16.3 to 4p11 involves 396 genes, 225 of that are protein-coding. The deleted interval at 15q26.3 involves 33 known genes, 20 of which are protein-coding. While none of these genes is individually known to have haploinsufficient phenotypes, this deletion may further modify the clinical phenotype of this case. Clinical correlation and genetic counseling were recommended. To the best of our knowledge, this unbalanced translocation between chromosome 4p and 15q is the first report of this combination of chromosomal abnormalities. Parental chromosome analyses were suggested to determine whether this abnormality is inherited or de novo in origin. In conclusion, this case demonstrates the effective use of cytogenetics, microarray and FISH studies for patient's diagnosis and care.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session II

PB3289 Unveiling the Cellular Landscape of Monogenic Genes of Kidney Stones: A Secondary Analysis of Single-cell RNA-seq Data.

Authors:

E. Eam[1,2], **J-e. Lo**[1], J. Chang[1,2], K. Lovrenert[1,2], D. Bodner[2], M. Baum[3,4], F. R. Schumacher[1], F. Hildebrandt[3,4], A. J. Majmundar[3,4], C-H. W. Wu[1,2]; [1]Case Western Reserve Univ., Cleveland, OH, [2]Univ. Hosp., Cleveland, OH, [3]Boston Children's Hosp., Boston, MA, [4]Harvard Med. Sch., Boston, MA

Kidney stones (nephrolithiasis) is an increasingly prevalent condition causing worldwide morbidity with a multitude of complex factors. In particular to the genetic factors, our previous review identified 57 causal monogenic genes; however, the mechanism of kidney stones formation is not completely understood.
Here, we performed a secondary analysis of kidney stone genes using the single-cell RNA sequencing (scRNA-seq) data from Park et. al (2018). We expanded the number of genes applied to their whole kidney scRNA-seq data from adult mice. The data matrix was downloaded from the Gene Expression Omnibus (accession number GSE107585), and we subset the 57 monogenic genes of kidney stones. The 16 distinct cell types, including kidney-specific and non-kidney-specific cells, were identified in alignment with the original authors.
The matrix was normalized, zero-adjusted, and z-scored. Unsupervised clustering of gene expression was performed by their z-scores using k-means. NbClust was used for clustering and optimization. ComplexHeatmap visualized the resultant expression matrix by a heatmap.
The results showed there were 8 distinct groups of genes that were expressed in different cell types. For example, SLC3A1 and SLC7A9, causal genes of cystine stones, were found to be exclusively overexpressed in the proximal tubule among the other kidney cell types. CLDN16 and CLDN19 were exclusively overexpressed in the ascending loop of Henle, whereas XDH and MOCS1 were overexpressed in neutrophils, a non-kidney-specific cell type.
The resultant heat map revealed that monogenic kidney stone genes are not universally expressed in the kidney but in specific cell types. This implies the pathogenesis started in that specific cell type. There were some genes showing overexpression in non-kidney-specific cell types, indicating that the primary pathogenesis is not from the kidney.
Our results suggest that future genetic research on kidney stones should identify the specific cell type where the gene is overexpressed. Functional studies investigating the novel gene mechanism or the drug development should target the specific cell types that greatly express the desired gene. In addition, our results reveal a new classification system for monogenic genes of kidney stones.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session III

PB3290 Unveiling the Significance of Copy Number Variations & Mitochondrial Mutations via a Single Next-Generation Sequencing Test

Authors:

P. Kamdar, T. Vaishnani, M. Haque, E. Jajodia, N. Arora, S. Chaudhary, J. Suthar; Unipath Speciality Lab. Ltd, Ahmedabad, India

Introduction Next-generation sequencing (NGS) has revolutionized our understanding of genetic variations, but detecting copy number variations (CNVs) and mitochondrial DNA mutations still remains a challenge and the incorporation of such variant detection may not be encompassed within a standard analytical framework. However, our study underscores the critical significance of incorporating the identification of copy number variation (CNV) and mitochondrial DNA mutations in standard next-generation sequencing (NGS) pipelines. **Methods** Samples referred to Unipath Speciality laboratory from March 2023 to May 2023 underwent exome sequencing using Sophia, enabling coverage of CNVs and mitochondrial DNA mutations. Library was prepared as per instruction manual of Sophia Exome Solution kit and sequencing was performed using Illumina NovaSeq 6000 platform. Variant calling, annotation and prioritization were carried out through machine learning approach used by Sophia DDM software. Relevant variants were reported based on ACMG 2015 guidelines. Maternal cell contamination was ruled out in case of prenatal samples. **Results** In our study, we found that 9 out of 230 patients exhibited either copy number variations or mitochondrial mutations. Our analysis unveiled pathogenic deletions encompassing a size range of 2.54 kilobases (affecting a single gene) to 15.4 megabases (involving over 100 genes). These deletions play a pivotal role in diagnosis of various disorders such as Duchenne muscular dystrophy, Methylmalonic aciduria, Turner syndrome, Krabbe's disease, osteosarcoma, DiGeorge syndrome, and Bartter syndrome. Furthermore, we identified a wide range of duplications, spanning from a few kilobases to entire chromosomes, such as Trisomy 18. Additionally, the identification of mutations in the mitochondrial DNA has proven instrumental in understanding and deciphering complex disease conditions like MELAS. **Conclusion** Our study highlights the significance of integrating CNV analysis and mitochondrial mutation assessment with SNV and small indel analysis to enhance the precision, effectiveness, and cost-efficiency of diagnostics. Unlike microarray, next-generation sequencing (NGS) has higher resolution to detect copy number variation lesser than 150 kb including a single exon.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session I

PB3291 † Update to the CLIA diagnostic procedure using optical genome mapping for the diagnosis of facioscapulohumeral dystrophy.

Authors:

J. Jang[1], H. Sung[1], J. Lee[1], S. Jo[1], J-S. Lee[1], M. Kim[2], M-W. Seong[1]; [1]Dept. of Lab. Med., Seoul Natl. Univ. Hosp., Seoul, Korea, Republic of, [2]Dept. of Genomic Med., Seoul Natl. Univ. Hosp., Seoul, Korea, Republic of

Background: Facioscapulohumeral muscular dystrophy (FSHD) is a rare progressive muscular dystrophy. The expression of the causative gene, DUX4 relies on the contraction or hypomethylation of the D4Z4 repeat sequence on chromosome 4q35 and a permissive haplotype. Currently, type 1 FSHD can be partially diagnosed in Korea. Cases with complex mechanisms such as hybrid or mosaicism, or type 2 FSHD is undiagnosable due to limitations in methods. Optical genome mapping (OGM) is an approach to detecting structural variations (SV) in whole genomes. As the base-specific fluorescent-labelled DNA flows through the nanochannel, the fluorescence pattern is captured by the camera. The imaged DNA is compared to a standard genome and SVs are detected by the different patterns. It can also be applied to detect contractions in repetitive sequences such as FSHD. The performance of OGM in simple FSHD caused by repetitive contractions is known, but not in FSHD involving complex mechanisms. We evaluate the diagnostic value of OGM in complex FSHD and propose an integrated diagnostic strategy.
Methods: We reviewed the test results of 218 patients with suspected FSHD at Seoul National University Hospital from 2017 to 2021, including 30 patients with complex FSHD. Complex FSHD includes hybrid, mosaic, 4 on 10, and type 2 FSHD that is not diagnosed by D4Z4 repeat contractions. OGM was performed in 26 patients, and the results were compared with SB using linear regression and the Blend-Altman method.
Results: A total of 226 alleles were identified for the 26 patients. The 10 hybrid cases showed the greatest discrepancy in the number of D4Z4 repeats between the two methods. Linear regression analysis showed that mosaics had an adjusted R^2 and p-value of 0.89 and < 0.01 respectively, whereas hybrids had -0.10 and 0.81, indicating a significant difference for hybrids. In the Blend-Altman analysis, the bias (confidence interval) was -0.096 (-1.632 to 1.441) for mosaics and 1.041 (-8.337 to 10.419) for hybrids, indicating a similar mean but a more variable distribution for hybrids.
Conclusions: We evaluated the diagnostic value of OGM for complex FSHD, which involves structural complexity beyond simple repeat contractions. Neither SB nor OGM alone could diagnose hybrid FSHD, and the combination of both was required. Here, we propose a modified Clinical Laboratory Improvement Amendments (CLIA) procedure in which OGM is performed first for suspected FSHD, followed by SB based on additional criteria, as a rapid and reliable method of diagnosing FSHD. However, both are repeat count methods, further studies using sequencing and epigenetic analysis are needed to better understand the pathogenesis of FSHD.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session II

PB3292 Vector copy number detection for CAR-T Therapy using a Droplet Digital PCR system.

Authors:

Y. Ding[1], Y. Wang[1], F. Tu[1], Z. He[2], H-J. He[2], Z. Zhu[1], S. Smirnov[1], A. Jadali[1], J-E. Lee[1], E. Stevenson[1], M. Sheldon[1], S. Nahas[1], R. Grimwood[1]; [1]Sampled, Piscataway, NJ, [2]Natl. Inst. of Standards and Technology, Gaithersburg, MD

Chimeric Antigen Receptor (CAR) T cell therapy has evolved and proven to be an effective regimen for individuals diagnosed with certain types of hematologic malignancies and solid tumors. Managing CAR-T cell therapy doses, monitoring safety and potency of CAR-T cell drug products and vector transduction rate is a challenge clinicians and CAR-T cell manufacturing facilities face. Vector Copy Number (VCN) detection by Droplet Digital PCR has been reported to be a valuable tool in assessing VCN and quality control of CAR-T cell products. Here, we report a method feasibility design and results using ddPCR system to detect VCN at Sampled. To do this, five DNA samples with well characterized lentiviral vector copy numbers extracted from transduced clonal Jurkat cell lines (provided by the NIST) were blinded at Sampled and run using the Bio-Rad QX200 ddPCR system, utilizing the Bio-Rad ddPCR Supermix (no dTUP) with assay on the target coupling with RPP30 (a house keeping gene) reference assay. Input DNA ranging from 0.005 ng to 20 ng per reaction were attempted with triplicates within same runs and repeated multiple times on the different days to estimate lowest possible input that could detect a target. Linearity between DNA input vs vector copy concentration was plotted. Only data with R^2 (coefficient of determination)>0.99 was accepted and used to estimate lowest input for target detection. Our results demonstrated that four samples detected VCN with expected concentrations (100% accuracy), and one showed no target detected, as expected. Limit of blank (LOB) was identified as 0 with 95% confidence. The lowest input for target detection was estimated as 4.54 copies per reaction at 95% confidence level. Our results demonstrated that a target was detected when input DNA was equal and higher than 0.005 ng per reaction. On average, with 10 ng input per reaction, CV% (Coefficient of variation) of intra-runs (repeatability) and inter-runs (reproducibility) of VCN concentration were 1.39% and 1.65% across samples respectively. In conclusion, VCN detection by the ddPCR system demonstrates high accuracy, precision and detection resolution applying low DNA input with less than 1ng per reaction. This workflow is suitable for a clinical laboratory to develop a LDT (Laboratory Developed Test) and to provide services for monitoring CAR-T production and clinical applications for individuals with CAR-T therapy.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session III

PB3293 WES CNV Filtering and Curation in Epilepsy and Obesity for Medically Reportable Secondary Findings per ACMG Guidelines

Authors:

S. Torkamandi, J. Glessner, H. Hakonarson; Children's Hosp. of Philadelphia, Philadelphia, PA

Copy number variants (CNVs) are important contributors to the genetic causes of epilepsy and obesity disorders. Whole-exome sequencing (WES) is a valuable tool for identifying CNVs associated with these conditions. This study aimed to enhance CNV calling accuracy in WES data using the XHMM protocol, while adhering to ACMG guidelines for medically reportable findings. To improve computational efficiency, we replaced the read depth calculation step with the mosdepth tool, yielding comparable results to the computationally intensive GATK. Normalization of read depth using PCA was crucial in reducing potential biases caused by the exome capture kit. XHMM CNV calling was performed on WES data from epilepsy (n=3,333) and obesity (n=9,533) patients. CNV calls were intersected with a curated list of 140 genes, including 65 CNV-specific genes. Out of 6,049 overlapping CNVs, 333 exhibited a 50% reciprocal overlap. Additionally, 18 CNVs intersected with reported CNV-specific gene locations. CNV counts per gene of interest and unique CNV numbers were generated for further analysis. Obesity CNVs and a syndromic list were processed across four cohorts. Filtering strategies were applied to 348,801 CNVs overlapping genes with a 50% query gene overlap, revealing distribution patterns based on length, genotype quality (GQ), and gene overlap frequency. A filtering strategy was proposed, restricting CNVs to deletions (DEL) or duplications (DUP) with length <10 Mb, GQ >500, and FILTER=PASS annotation. Applying these filters yielded 14,580 CNVs. Further applying a 50% reciprocal overlap criterion with entire gene boundaries and GQ >500 reduced the CNV count to 16. ClassifyCNV algorithm, following ACMG standards, categorized DEL or DUP rows. In conclusion, this study improved the accuracy and reliability of CNV calling in WES data using the XHMM protocol, following ACMG guidelines. The analysis identified CNVs associated with epilepsy and obesity disorders, highlighting specific genes of interest. The proposed filtering strategy allowed for refined CNV selection. This research provides valuable insights into the genetic etiology of these disorders and contributes to the understanding of CNVs in clinical applications.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session I

PB3294 WGS improves detection of Duchenne and Becker Muscular Dystrophy: Case studies

Authors:

S. Gessay[1], **E. Bend**[1], A. Gruber[1], S. Bianconi[2], A. Cooper[2], R. Peretz[3]; [1]PreventionGenetics LLC, Marshfield, WI, [2]Kaiser Permanente, San Diego, CA, [3]Harbor-UCLA Med. Ctr., Los Angeles, CA

Duchenne and Becker Muscular Dystrophy (DMD and BMD, respectively) are among the most common inherited neuromuscular conditions in children. The allelic disorders are caused by variants in *DMD*, located on Xp21.2, but efficient detection of variants has been difficult given the high percentage of causative intronic and copy number variants (CNVs). Traditional cytogenetic techniques, including karyotype and chromosomal microarray (CMA), lack sufficient resolution to identify sequence variants; and sequencing approaches, including Sanger and exome-based next-generation sequencing (NGS), offer limited detection of sequence variants and CNVs outside the exome. Whole genome sequencing (WGS), in contrast, covers both intronic and intergenic regions and is able to detect CNVs all with one test. We present three cases where relevant findings were discovered by WGS that escaped detection by another platform. 1) A 7-year-old male's clinical history was notable for Gower sign and progressive muscle weakness. Previous NGS panel analysis was inconclusive, leading the care team to pursue single gene testing on a WGS platform. The intronic variant c.5739+362A>G was detected and classified as likely pathogenic given descriptions from published case studies and its absence in population databases. 2) A 5-year-old male presented to clinic given his combined history of Gower sign, bilateral mid-calf hypertrophy, elevated CK levels, and positive muscle biopsy results. Genetic testing, including MLPA and unspecified sequencing analysis, were uninformative. Panel analysis performed on a WGS platform identified a complex rearrangement resulting in a duplication at 2q21.1q22.1 near the splice acceptor site of intron 23 in *DMD*. The location of the event led to a likely pathogenic classification given reports of large insertions altering splicing and disrupting gene expression. 3) A preadolescent biological male was evaluated for a years-long history of muscle pain and weakness. The patient's CK levels were significantly elevated, but results from sequencing analysis with del/dup were non-diagnostic. WGS run on his sample identified a hemizygous ~35.9 kb duplication in tandem encompassing exons 10 and 11 within *DMD*. Similar published case reports contributed to its pathogenic classification. These cases demonstrate that WGS can offer genetic confirmation in previously negative cases, which has the potential to provide patients with access to targeted therapeutics and inform recurrence risk for reproductive decisions.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session II

PB3295 Whole exome sequencing for prolonged neonatal/infantile intrahepatic cholestasis discovered a definitive molecular genetic diagnosis.

Authors:

S. Ito[1], T. Togawa[1], T. Tanimura[1], K. Imagawa[2], H. Hayashi[3], K. Ito[1], T. Endo[1], T. Sugiura[1], S. Saitoh[1]; [1]Nagoya City Univ., Nagoya, Japan, [2]Univ. of Tsukuba, Tsukuba, Japan, [3]The Univ. of Tokyo, Tokyo, Japan

Genetic liver diseases, such as cystic fibrosis, alpha-1 antitrypsin deficiency, and progressive familial intrahepatic cholestasis (PFIC), can cause intrahepatic cholestasis, leading to fibrosis and cirrhosis. The majority of these conditions manifest as direct hyperbilirubinemia, commencing in the neonatal or infantile period, commonly referred to as neonatal/infantile intrahepatic cholestasis (NIIC). Despite advancements in liver transplantation and the advent of emerging pharmaceutical agents, clinicians encounter challenges in establishing a precise diagnosis due to overlapping and extensive clinical manifestations. In a previous publication, we documented a conclusive molecular diagnosis in 27% of the patients enrolled in our targeted 61-gene panel study, utilizing the Ion Personal Genome Machine system (Thermo Fisher Scientific) in conjunction with bioinformatics pipelines. Nevertheless, approximately 70% of the patients remained with no definitive genetic diagnosis. Consequently, we have resolved to perform whole exome sequencing (WES) to enhance the diagnostic yield for cholestatic patients. Of the NIIC patients who were examined with the 61-gene panel between May 2013 and March 2023 and had prolonged cholestasis with no definitive genetic diagnosis, 18 probands, and their parents were included in the study. We performed WES with trio analysis and bioinformatic pipelines on the Illumina platform. Five of 18 (27.8%) probands received a definitive molecular genetic diagnosis; they carried causative variants in *ABCB11* (PFIC type2), *SLC25A13* (neonatal intrahepatic cholestasis due to citrin deficiency), *SLCO1B1* and *SLCO1B3* (Rotor syndrome), *MYO5B* (PFIC type 10), or *MTM1* (severe X-linked myotubular myopathy) in each. Out of the 6 identified genes, *ABCB11, SLC25A13, SLCO1B1*, and *SLCO1B3* were included within our 61-gene panel. We had already found one pathogenic variant on a single allele of *ABCB11* and *SLC25A13* in an autosomal recessive manner. Because WES discovered a wider range of the targeted region than that of our previous targeted 61-gene sequencing, we additionally detected one more causative pathogenic variant on the other allele of the 2 genes. Regarding *SLCO1B1* and *SLCO1B3* in a digenic recessive manner, we newly detected 4 pathogenic variants due to the same reason. Consequently, WES possessed the capability to uncover variants not only in genes that were previously unexamined but also in genes that were previously confirmed through panel analysis.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session III

PB3296 Whole exome/genome sequencing with rapid automated judgment algorithm for diagnosis of difficult-to-diagnose hereditary hearing impairment patients

Authors:

C-Y. Tsai[1,2], Y-S. Lu[2], Y-T. CHIANG[1,2], C-J. Hsu[3,2], J. Hsu[1], C-C. Wu[2], P-L. Chen[1,4,5]; [1]Graduate Inst. of Med. Genomics and Proteomics, Natl. Taiwan Univ. Coll. of Med., Taipei, Taiwan, [2]Dept. of Otolaryngology, Natl. Taiwan Univ. Hosp., Taipei, Taiwan, [3]Dept. of Otolaryngology, Taichung Tzu Chi Hosp., Taichung, Taiwan, [4]Dept. of Med. Genetics, Natl. Taiwan Univ. Hosp., Taipei, Taiwan, [5]Inst. of Molecular Med., Natl. Taiwan Univ. Coll. of Med., Taipei, Taiwan

Background: Hereditary hearing impairment (HHI) is a common inherited disorder characterized by genetic heterogeneity. However, the genetic etiologies of many unconfirmed HHI families remain elusive. Next-generation sequencing (NGS), specifically whole genome/exome region sequencing (WGS/WES), provides a powerful solution to explore broader genomic regions for hidden pathogenic variants. To effectively apply NGS in the clinical genetics of HHI, it is crucial to develop and package an expanded collection of HHI-related genetic information and an efficient method to identify causative variants from the vast amount of WGS/WES data.Method:
We employed WGS/WES assays based on the capture-based NGS approach using the Illumina platform. In addition, we developed a "virtual panel" consisting of 728 deafness genes, which integrates our NTUH targeted NGS panel (216 genes) and four other academic databases (Hereditary Hearing Loss, Deafness Variation database, clinical genomic database and PanelApp). Furthermore, we designed a Python-driven algorithm that rapidly and automatically evaluates NGS data based on multiple metrics, including the reported inheritance model of targeted genes, the genotype of each variant, and the cosegregation analysis of multiple samples. This algorithm generated all possible causative genotypes and identified the genetic etiologies based on the ACMG guideline.**Result:**We recruited a total of 46 multiplex families for the WGS/WES assays. The automated judgment process, which output all candidate genotypes from the annotated variant file of each family, took less than one minute. We identified a total of 31 pathogenic variants in 22 families (diagnostic rate = 47.8%), including 15 novel variants accounting for 41.9% of all confirmed variants. Among the 31 variants, 15 (48.4%) were categorized as null variants. Additionally, two intronic variants (*GSDME*: c.1883+5G>A and *MITF*: c.1013+5G>A), which exhibited high predicted scores for abnormal splicing, were considered potentially deleterious based on cosegregation analysis in the affected members within multiplex families.**Conclusion:**Compared to conventional panel-based NGS assays, WGS/WES assays coupled with the virtual panel of 728 HHI-related genes demonstrated a high diagnostic rate for causative variants in challenging multiplex HHI families. Furthermore, the development of a rapid algorithm for automated judgment of causative HHI variants is feasible for further clinical applications, alleviating the time-consuming and laborious manual pathogenicity assessment pipeline.

Session Title: Molecular and Cytogenetic Diagnostics Poster Session I

PB3297 Whole Genome Sequencing in Adults with Chronic Kidney Disease or End-Stage Kidney Disease in the Million Veteran Program

Authors:

J. Triozzi, H-C. Chen, R. Tao, A. Hung; Vanderbilt Univ. Med. Ctr., Nashville, TN

Background: As the utilization of diagnostic genetic testing in nephrology increases, genetic causes of kidney disease are identified at significant rates in the adult population. A molecular diagnosis has numerous advantages for patient care, including facilitating treatment decision-making, enabling familial testing, providing information for safe kidney allograft donation, as well as uncovering insights into the pathogenesis of kidney diseases. In this study, we report whole-genome sequencing results among patients with chronic kidney disease or end-stage kidney disease.

Methods:

We selected genes associated with Mendelian kidney and genitourinary disorders compiled in a previous study by Rasouly et al (PMID: 30476936). We performed a secondary analysis of whole genome sequencing data of patients with chronic kidney disease (CKD) or end-stage kidney disease (ESKD) available from the Department of Veteran Affairs Million Veteran Program (MVP) pilot study. CKD was defined as eGFR < 60 mL/min/1.73 m2 on two occasions 90 days apart. ESKD was defined by diagnosis or procedure codes for chronic dialysis, kidney transplant or eGFR < 15 mL/min on two occasions 90 days apart. The prevalence of genetic variants for CKD or ESKD was tabulated by category.

Results:

A total of 4236 patients met criteria for CKD (n = 3411, 81%) and ESKD (n = 825, 19%) with a mean age of 68 (SD 9.56) years. There were 37% African, 57% European, 3% Hispanic, and <1% Asian ancestry individuals. There was a high prevalence of diabetes (63%) and hypertension (99%). There were urinary markers of kidney disease including hematuria (n=542, %), proteinuria >1+ (N=1,013, %), and proteinuria >=2+ (N=602, %). A total of 431 genetic variants related to kidney disease were identified. These genetic variants were related to cystic and tubulointerstitial kidney disease (24%, PKD1, PKD2, PKHD1), complement-related diseases (17%, CFH, C3, CFI, ADAMTS13), tubular disorders (6%, HBB), and glomerular diseases (4%, COL4A3, COL4A4, CUBN). Notably, we identified multiple rare genetic causes of nephrolithiasis, such as cystinuria, primary hyperoxaluria, and hypophosphatemic rickets.

Conclusion:

Genetic variants for actionable disorders were identified in this cohort of patients with CKD or ESKD in the MVP. In contrast to previous reports, our findings identified rare nephrolithiasis variants among patients decreased kidney function, highlighting their clinical significance and potential impact on outcomes. More detailed phenotyping of our cohort is underway to better understand the clinical implications of these variants.

Session Title: Omics Technologies Poster Session I

PB3298 A clinical and molecular approach for the classification of *KANSL1* variants as a strategy to solve pitfalls of genome-wide investigations.

Authors:

G. Marangi[1], F. F. L'Erario[1], S. Maietta[1], S. Frangella[1], D. Pasquetti[1], A. G. Renzi[1], M. Carapelle[1], A. Panfili[1], A. Gazzellone[1], D. Orteschi[2], S. Bigoni[3], G. Zampino[2], M. Zollino[2]; [1]Università Cattolica Del Sacro Cuore, Rome, Italy, [2]Università Cattolica del Sacro Cuore, Rome, Italy, [3]Azienda Ospedaliero-Univ.ria di Ferrara, Ferrara, Italy

The development of high-throughput sequencing (HTS) methods has led to an extensive implementation of a genome-wide approach in routine genetic investigations. As a result, this has had a massive impact on the diagnostic yield, also enabling the rapid identification of new Mendelian disease genes. The dramatic increase in the number of variants that need to be interpreted constitutes the main drawback of the use of HTS as first-tier genetic test in diagnostic settings. Koolen-de Vries syndrome (KdVS) can be considered a paradigmatic example of the effect of the use of genome-wide approach in medical genetics. The recurrent 17q21.31 deletion has been firstly identified by means of array-CGH, while exome HTS in non-deleted patients allowed for the definition of *KANSL1* as the causative gene. Thereafter, potential loss-of-function variants in *KANSL1* detected in patients with syndromic intellectual disability have been routinely classified as pathogenic. However, common duplication polymorphisms involving the first two exons of the NM_015443.4 “MANE Select” transcript of *KANSL1* are found in about the 40 % of alleles in European populations. Hence, it is crucial to consider these polymorphisms in the evaluation of genetic variants. We propose a gene-specific strategy for *KANSL1* variant interpretation based on the architecture of the 17q21.31 region, the evaluation of consequences at mRNA level and a precise clinical characterization of patient phenotype.We report five genomic abnormalities of *KANSL1*: one case of exons 1-2 deletion detected by array-CGH, and four cases of variants affecting exon 2 (c.540delA and c.985_986delTT, in three cases) identified by exome sequencing. Our clinical and molecular evaluation included: 1) deep clinical evaluation of patients; 2) parental analysis; 3) MLPA of *KANSL1*, specifically devised to distinguish the benign duplication polymorphism form other CNVs; 4) cDNA sequencing of *KANSL1*. While we could classify as pathogenic the c.540delA variant found in a patient with a phenotype fully consistent for KdVS, we categorized both the exon 1-2 deletion and all the three c.985_986delTT variants as benign, since they affected the duplicated, non-functional copy of exons 1-2. Nonetheless, this classification cannot always be assumed as true, since the validation strategy we propose should be applied for any single patient carrying a potentially pathogenic variant in the first two exons of *KANSL1*. This report highlights the pivotal role of clinical genetics in precise diagnosis of syndromic neurodevelopmental disorders.

Session Title: Omics Technologies Poster Session II

PB3299 A cloud-based pipeline for genome-wide short tandem repeat detection in genome sequencing and its application to orofacial clefting.

Authors:

N. Kurtas[1], V. Jalili[2], B. Weisburd[3], K. Robinson[4], T. H. Beaty[5], A. Butali[6], C. J. Buxó[7], J. T. Hecht[8], G. M. Shaw[9], L. M. Moreno[10], J. C. Murray[11], S. M. Weinberg[12], M. L. Marazita[12], M. E. Talkowski[1], E. J. Leslie[4], H. Brand[1]; [1]Ctr. for Genomic Med., Massachusetts Gen. Hosp., Boston, MA, [2]Data Sci. Platform, Broad Inst. of MIT and Harvard, Cambridge, MA, [3]Program in Med. and Population Genetics, Broad Inst. of MIT and Harvard, Cambridge, MA, [4]Dept. of Human Genetics, Emory Univ., Atlanta, GA, [5]Dept. of Epidemiology, Johns Hopkins Univ., Baltimore, MD, [6]Dept. of Oral Biology, Radiology and Med., Univ. of Iowa, Iowa City, IA, [7]Sch. of Dental Med., Univ. of Puerto Rico, San Juan, Puerto Rico, [8]Dept. of Pediatrics, McGovern Med. Sch., Univ. of Texas Hlth.at Houston, Houston, TX, [9]Dept. of Pediatrics, Stanford Univ., Stanford, CA, [10]Dept. of Orthodontics & The Iowa Inst. for Oral Hlth.Res., Univ. of Iowa, Iowa City, IA, [11]Dept. of Pediatrics, Univ. of Iowa, Iowa City, IA, [12]Dept. of Oral and Craniofacial Sci., Univ. of Pittsburgh, Pittsburgh, PA

Short tandem repeats (STRs) are a class of structural variation characterized by the expansion or contraction of repeat motifs of 1-6 bp. STRs have been associated with a number of diseases and recent bioinformatic advances have facilitated accurate genotyping of known disease STRs in short-read whole genome sequencing data (srWGS). Despite these advances, there has been little success in discovering pathogenic STR loci outside these predefined lists due to technical challenges in novel STR detection from srWGS. Taking advantage of the current genotyping strength of STRs in srWGS, we developed a novel genome-wide STR discovery approach using a cloud-based pipeline for the Expansion Hunter (EH) tool. We applied the pipeline to genome wide STRs derived from application of TandemRepeatFinder on GRCh38 obtaining all STRs with 3 or more repeats that spanned ≥ 9bp in the reference (n = 3,281,006). We reduce genotyping cost by focusing on plausible functional regions by overlapping at least one of the following annotations: genic, TAD boundaries, non-coding constrained regions, VISTA enhancers, ultraconserved elements, HACER enhancers, human accelerated regions or constrained STR loci (Gymrek et al. 2017) leaving a total of 2.1M sites. We benchmarked this pipeline against 3 samples from the 1000 Genomes Project with matched srWGS and long-read sequencing data (Pacbio HiFi, STR calling with TRGT) and found an average of concordance of 86.9%.

We applied our STR pipeline to a cohort of 437 case-parent trios with probands affected with isolated cleft palate (CP), performing the first assessment of the effect of genome-wide STR variation on CP risk. Taking advantage of our trio structure, we primarily focused on the influence of *de novo* variants and found an average of 236 *de novo* variants per proband across 62,420 total sites. Upon checking for enrichment of *de novo* STR alleles in CP cases across the functional relevant annotations described above, we observed two significant categories (HACER enhancers - p = 9.7e-05; constrained STR loci - p=1.1e-61). Among the constrained STR, we observed an interesting locus with a single TTTG expansion in an intronic repeat of *PIEZO2* (chr18:11010608). This gene is associated with Marden-Walker Syndrome, which often features a cleft or high-arched palate. Interestingly, we found an additional intronic STR upstream in *PIEZO2* (chr18:10759203) that had a *de novo* expansion in 3 CP cases, providing further evidence that STR variation may influence this gene. Further functional studies are necessary to confirm these findings, but these results demonstrate the promise of investigating STRs in CP and potentially other birth defects.

Session Title: Omics Technologies Poster Session III

PB3300 A comprehensive single-cell RNA sequencing analysis evaluating the role of T-lymphocytes in idiopathic and genetic Parkinson's disease.

Authors:

M. Rosen[1], O. Narcis[1], A. Allan[1], T. Khashan[1], E. Mejia[1], C. Argyru[1], S. Frucht[2], R. Saunders-Pullman[3,1], G. Riboldi[2], T. Raj[1]; [1]Icahn Sch. of Med. at Mount Sinai, New York, NY, [2]New York Univ. Langone Hlth., New York, NY, [3]Mount Sinai Beth Israel, New York, NY

Parkinson's disease (PD) is a devastating neurodegenerative disease characterized by a loss of dopaminergic neurons in the substantia nigra, movement abnormalities, and in some cases the accumulation of α-synuclein (α-syn) as fibrils or Lewy bodies. It is well established that T-cells are involved in PD pathology, shaping the inflammatory cascade from the periphery into the brain. Moreover, genetic studies have linked the human leukocyte antigen (HLA) class II and the risk of developing PD, which suggests that CD4 T-cells may have a crucial role in their interaction with antigen-presenting cells. We explore the role of T-cells in PD by rigorously characterizing their transcriptional profile and T-cell receptor (TCR) repertoire during disease progression.Our aims are to (i) characterize immune cell expression and repertoire in PD patients, prodromal patients (subjects with elevated risk of developing PD, including GBA/LRRK2 genetic carriers and patients with rapid eye movement (REM) sleep behavior disorder) and control donors; (ii) assess the clinical relevance and antigen specificity of clonal T-cells; and (iii) explore the functional impact of human leukocyte antigen (HLA) risk and protective alleles on T-cell biology and PD pathogenesis. We are performing a large-scale, comprehensive single-cell RNA-seq analysis on a well-characterized cohort (n=150) to control for biological heterogeneity. We are planning to obtain gene expression, V(D)J immune receptor gene expression and genetic information for each donor. We perform standard single-cell data processing using an in-house pipeline utilizing the Seurat package. We perform differential expression analysis using both single-cell and pseudobulk methods. Our preliminary results (n=24) demonstrate differences in CD4 and CD8 T-cell proportions between PD patients and control donors (p=0.05 and p=0.1, respectively). Furthermore, we have observed distinct V(D)J gene expression patterns indicating clonal expansion in CD8 activated T-cells in PD. These novel insights shed light on the involvement of T-cells and their potential impact on PD development and progression.We are currently increasing our sample size to reach our target cohort. Further, we are processing genotype data so that we can impute HLA class II alleles. We will investigate the association between HLA alleles and TCR gene usage and the hypervariable complementarity determining region 3. Overall, the results of these experiments will further our understanding of the role of T-cells in PD, lead to mechanistic discovery and determine the potential for targeted immune-based therapeutic interventions or biomarker capabilities.

Session Title: Omics Technologies Poster Session I

PB3301 A consensus algorithm for genetic demultiplexing of scRNA-seq from pooled iPSC cultures of multiple subjects for cohort-scale analysis

Authors:

M. Fiorini, R. Thomas, S. Farhan, S. Amiri; McGill Univ., Montreal, QC, Canada

Induced pluripotent stem cells (iPSCs) and single-cell RNA sequencing (scRNA-seq) can be used to model the transcriptome of unique cell types in neurodegenerative disease *in vitro*. However, iPSC-based scRNA-seq remains costly and low throughput, restricting the feasibility of cohort-scale study designs, which is required for therapeutic exploration. Multiplexing iPSC lines bypasses the costly parallel growth of individual iPSC lines by pooling batches of lines in a single dish. Demultiplexing algorithms can then resolve the sample identity of each cell based on genetic differences. However, the accuracy of available demultiplexing software for cohort-scale study designs remains questionable, as benchmarking has been predominantly based on small sample sizes. To facilitate cohort-scale iPSC-based scRNA-seq, we are developing a consensus demultiplexing software that assigns sample identities to individual cells based on the predictions of four currently available demultiplexing algorithms. We are testing our consensus algorithm on multiple synthetically pooled iPSC-derived dopaminergic neuron scRNA-seq transcriptomes from control and Parkinson's disease (PD) patients generated by the FOUNDIN-PD project. In silico pools ranging from 4 to 80 iPSC lines will be produced and tested in our pipeline to ensure our algorithm remains accurate for cohort-scale cell batches. Preliminary results have shown that our consensus algorithm outperforms all four other demultiplexing algorithms for cell pools comprising three different iPSC lines, and we expect this improvement to increase for larger pools as individual algorithms struggle to maintain accuracy. Eventually, as proof of principle, we will harness our consensus demultiplexing algorithm to test the hypothesis that multiplexed iPSC lines and scRNA-seq can be used to screen for transcriptional responses to chemical compounds in dopaminergic neurons in PD patients. Our novel demultiplexing algorithm will facilitate the application of cohort-scale iPSC-based scRNA-seq across neurodegenerative diseases, allowing for cell-type specific transcriptional responses to chemical compounds to be explored and elucidate novel therapeutics.

Session Title: Omics Technologies Poster Session II

PB3302 A deep learning model for predicting probeset performance of genotyping microarrays

Authors:

F. Lin[1], A. Mittal[1], J. Gollub[2]; [1]Thermo Fisher, SANTA CLARA, CA, [2]Thermo Fisher Scientific, Santa Clara, CA

Genotyping microarrays are a high-throughput tool for measuring genetic variation across the genome. Accurate prediction of probeset performance is crucial for optimizing the array design strategy and maximizing accuracy and call rates. Probeset performance is impacted by various factors including hybridization specificity, secondary structure propensity, sequence similarity or identity to off-target loci, and sequence complexity. Deep learning algorithms have proven effective in identifying complex local patterns within the sequences. In this work, we built a deep learning model to predict probeset performance using probe sequence to capture the sequence composition and local sequence patterns and additional features, such as sequence similarity and the number of copies of the probeset on the array that can impact the robustness of signal intensities. The model is comprised of a bi-directional recurrent neural network fused with a deep neural network to incorporate sequence and numerical features. The model was trained using the performance information of over 3 million probesets. A binary probeset performance label was assigned based on the frequency with which the probeset passed the stringent Axiom genotyping QC metrics. The performance data was split into training, validation and testing set in an 8:1:1 ratio. The training and validation performance are similar, around an AUC-ROC of 0.75. The testing performance slightly decreased, with an AUC-ROC of 0.74. We also evaluated the contribution of each feature to the predictions and found both probe sequence and the two numeric features impacted the performance prediction, with sequence homology exhibiting the highest contribution to performance prediction.

Session Title: Omics Technologies Poster Session III

PB3303 A graph clustering algorithm for detection and genotyping of structural variants from long reads.

Authors:

J. Duitama, N. Gaitán; Univ. de los Andes, Bogotá, Colombia

Structural variants (SV) are genomic polymorphisms defined by their length (more than 50 bp). The usual types of SVs are deletions, insertions, translocations, inversions, and copy number variants. SV detection and genotyping is fundamental given the role of SVs in phenomena such as phenotypic variation and evolutionary events. Thus, methods to identify SVs using long-read sequencing data have been recently developed. We present an accurate and efficient algorithm to predict SVs from long-read sequencing data. The algorithm starts collecting evidence (Signatures) of SVs from read alignments. Then, signatures are clustered based on a Euclidean graph with coordinates calculated from lengths and genomic positions. Clustering is performed by the DBSCAN algorithm, which provides the advantage of delimiting clusters with high resolution. Clusters are transformed into SVs and a Bayesian model allows to precisely genotype SVs based on their supporting evidence. This algorithm is integrated into the single sample variants detector of the Next Generation Sequencing Experience Platform (NGSEP), which facilitates the integration with other functionalities for genomics analysis. Benchmark experiments show that our approach outperformed state-of-the-art tools on SV calling and genotyping, especially at 20x depth, on a simulated dataset with an F-score of 99.1%, and on the HG002 GIAB PacBio HiFi sample with an F-score of 94.5% and 96% GT-Accuracy. These experiments show that our algorithm reliably detects and genotypes most SV events. We believe this work significantly contributes to the development of bioinformatic strategies to maximize the use of long-read sequencing technologies.

Session Title: Omics Technologies Poster Session I

PB3304 A haplotype-aware framework for detecting differentially methylated regions using Oxford Nanopore sequencing.

Authors:

M. Ahsan, K. Wang; Children's Hosp. of Philadelphia, Philadelphia, PA

DNA methylation is an important biological marker that affects a range of physiological processes, such as aging and carcinogenesis, and cellular processes, such as differentiation and pluripotency. Analysis of differentially methylated regions can provide crucial insights into how certain genomic regions or biological pathways are affected by different biological or medical conditions. For instance, detection of differentially methylated regions between cancer-normal paired samples can lead to more accurate cancer subtyping and identification of therapeutic targets. Similarly, identification of differentially methylated regions between haplotypes of the same sample can lead to identification of novel imprinted regions, which can provide a better understanding of imprinted disorders. In recent years, Oxford Nanopore Technologies (ONT) sequencing has emerged as a reliable and direct technology in DNA methylation detection through the analysis of ionic current signal. However, recent studies aimed at differential methylation analysis from ONT sequencing use methods and tools designed for short-read bisulfite sequencing, such as DSS or Methylkit. To enable the use of short-read methods, these studies reduce all methylation probabilities from a single molecule into statistically independent counts of cytosines and thymidines. As a result, such analyses ignore the rich long-range information provided by Nanopore reads that cover multiple CpG sites whose methylation levels are linked together. This can reduce the statistical power and sensitivity of such analyses. Here we present a haplotype-aware framework for detecting differentially methylated regions using ONT sequencing. Our method exploits the dependence between methylation levels of CpG sites on the same reads to more accurately detect regional differences between samples. We evaluated our framework on paired-tissue samples and HG002 genome from Oxford Nanopore open datasets and found that incorporating haplotype information improves differential methylation detection.

Session Title: Omics Technologies Poster Session II

PB3305 † A high-throughput, automated ATAC-seq reveals quantitative regulatory effects of transcription factor dosage.

Authors:

B. B. Liu[1], S. Vermeulen[1], S. H. Kim[1], T. Clark[2], N. Iremadze[2], D. Lipson[2], Z. Shipony[2], W. J. Greenleaf[1]; [1]Stanford Univ., Stanford, CA, [2]Ultima Genomics, Newark, CA

Transcription factors (TFs) regulate gene expression by binding to accessible DNA regulatory regions. However, a quantitative understanding of the effects of TF dosage on the chromatin accessibility and gene expression landscape for most TFs remain unclear. A systematic investigation of the effects of varying TF levels on accessibility and gene expression is challenging to carry out across a large panel of TFs as current accessibility assays are time, labor, and cost-intensive and often lack reproducibility when manually repeated across small batches. To address this challenge, we developed roboATAC, an automated, scalable plate-based ATAC-seq platform that replaces rate-limiting centrifugation steps with bead-based solutions and validated roboATAC against the gold-standard omniATAC method. Using roboATAC and RNAseq, we assayed genome-wide chromatin accessibility and gene expression changes induced by variable overexpression of individual and pairs of five TFs (GATA1, GATA2, TAL1, KLF1, SPI1) in HEK293T cells across a total of 88 samples. We sequenced both ATAC and RNA libraries with Ultima Genomics. Our analysis showed significant enrichment of TF motifs associated with overexpressed TFs within accessible regions. We also see dosage-dependent effects on accessibility that are TF specific, with increasing TF footprint depths and increasing accessibility flanking TF binding motifs as a function of increasing TF expression. In addition, we identified thousands of differential genes with increased expression in each condition, enriching for pathways relevant to the TF function (e.g. SPI1 induced expression of genes involved in T cell activation and lymphocyte differentiation absent in wild type HEK293T cells) and exhibiting dosage-dependent effects consistent with those observed in the open chromatin. Furthermore, our data revealed diverse dynamics in TF dosage response, with the majority of motif-containing peaks exhibiting limited sensitivity to TF dosage, while a small percentage (e.g. 8% in SPI1 samples) showed linear or hill dynamics. We also observed that peaks with higher TF binding affinity reached accessibility and linked gene expression saturation at earlier dosages, highlighting the role of TF-DNA interactions in governing transcriptional outcomes. Overall, this dataset is a valuable resource for generating mechanistic hypotheses on how transcription factor level regulates gene expression, as well as developing strategies to utilize those mechanisms for reversing disease states.

Session Title: Omics Technologies Poster Session III

PB3306 A joint model of cell-type-specific RNA and protein expressions in the human brain.

Authors:

D. Bai[1], P. Emani[2], M. Gerstein[3]; [1]Yale Univ., New haven, CT, [2]Yale Univ., New Haven, CT, [3]Yale Univ, New Haven, CT

Recent advances in single-cell technologies have provided unprecedented insights into our understanding of transcriptional diversity at the single-cell level. However, despite the importance of quantifying proteins at a cellular level, there are no proteomic amplification methods to do so at the same resolution as RNA. To address this, we have leveraged information from single-cell RNA-seq (scRNA-seq) to aid in the identification of cell-type-specific proteomic signatures by incorporating the stochastic nature of gene expression for better understanding of the biological variation between mRNA and protein distributions in human brain tissue samples. First, we introduce a zero-inflated Beta Poisson distribution to model cell-type specific mRNA distributions from scRNA-seq data. This model allows us to capture information about underlying transcriptional parameters that give rise to stochastic gene expression. To separate technical noise from biological components, we incorporated a binomial distribution noise model. Next, we applied deconvolution of tissue-level RNA-seq to obtain cell type fractions within each tissue sample. Finally, we used matching proteomics data from the same individuals, as well as mathematical relationships between the expected protein levels, translational rates and transcriptional rates to impute cell-type specific protein signatures. We subsequently identified genes and cell types that show significant variation between mRNA and protein levels. We compared our results to experiments on RNA and protein variations. In summary, our framework circumvents the unavailability of single-cell proteomics by using models of stochastic gene and protein expression, thus offering better insights into the biological discrepancies between cell-type-specific mRNA and protein levels.

Session Title: Omics Technologies Poster Session I

PB3307 A new hybrid method provides robust analysis for clustered data, with applications to single-cell transcriptomics data.

Authors:

M. Du, Z. Yu, V. Berrocal, X. Xu; Univ. of California, Irvine, CA

Technological developments in quantifying molecular profiling characteristics, such as transcriptomics, proteomics, microbiome data, have gone through multiple breakthroughs. As a result, we can now measure the whole gene profiles of each individual cell. Identifying differentially expressed genes across various conditions remains challenging due to small sample sizes, non-normal distributions of transcriptomic counts, and repeated measurements from the same biological unit. Existing methods often focus on only a subset of these issues. Here we propose a novel computationally feasible hybrid method that takes advantage of the robustness of rank-based methods and the statistical efficiency of parametric methods for small sample sizes. Simulation studies suggest that our method achieves optimal statistical properties compared to existing methods. Importantly, our new method effectively controls Type I error rate at the nominal level while achieving a good statistical power of detecting differentially expressed genes. Furthermore, our method readily extends to differential analysis involving multiple conditions and enables adjustment for potential covariates. We applied our method to a spatial transcriptomics dataset of Alzhimer’s disease mouse model. The differentially expressed genes identified through our method align with known genes associated with Alzheimer's disease, further validating our approach. In conclusion, our novel hybrid method effectively tackles key challenges simultaneously and offers a new computationally and statistically efficient framework for analyzing complex molecular profiling data.

Session Title: Omics Technologies Poster Session II

PB3308 A new tool for enhancing the interpretability of machine learning prediction models leveraging '-omics' data.

Authors:

X. Loffree[1,2], H. Yun[1,3], J. Li[4,5], L. Liang[1]; [1]Dept. of Biostatistics, Harvard T.H. Chan Sch. of Publ. Hlth., Boston, MA, [2]Dept. of Statistics and Actuarial Sci., Univ. of Waterloo, Waterloo, ON, Canada, [3]Dept. of Epidemiology, Harvard T.H. Chan Sch. of Publ. Hlth., Boston, MA, [4]Div. of Preventive Med., Dept. of Med., Brigham and Women's Hosp. and Harvard Med. Sch., Boston, MA, [5]Dept. of Nutrition, Harvard T.H. Chan Sch. of Publ. Hlth., Boston, MA

Background: When using high-dimensional data in prediction analysis, the trade-off between the predictive power and interpretability of machine learning models is important. Greater predictive power comes at the expense of interpretability. The model interpretability is pivotal in medical settings, especially for healthcare providers, clinical consulting, and pharmaceutical discovery. To objectively assess the interpretability for prediction models based on large-scale data, we have designed a Python tool to enhance the interpretability of machine learning models using omics data. Methods: We examined three interpretability metrics: Permutation Feature Importance (PFI), Local Interpretable Model-agnostic Explanations (LIME), and SHapley Additive exPlanations (SHAP). Each employs a different mathematical principle to determine the individual contribution of a feature to the prediction of a black box model. Our Python tool applies these metrics to a machine learning model, evaluates the distributions of these metrics, and identifies the most influential features in determining predictions. Additionally, we propose the Robust Interpretability Metric (RIM) - a new interpretability metric calculated by aggregating results from LIME, PFI, and SHAP. In a previous study, we established metabolomic signatures (MS) for 8 dietary patterns including the DASH diet, using elastic net regression in the Lifestyle Validation Study (LVS, n=1134). As a case study, we applied our tool to facilitate the interpretations of these MS. Results: Applying our tool to interpret the MS models, we identified the most influential metabolites in determining dietary index alignment. Using the DASH diet as an example, among the 286 metabolites used to train the model, RIM values revealed 4 glycerophospholipids as the most important feature for prediction: lyso-phosphatidylcholine (PC) 22:5 (RIM=5.61), PC 34:0 (RIM=5.09), PC 38:3 (RIM=4.49), and phophatidylethanolamine 18:0 (RIM=3.94). The distribution of the 286 RIM values was skewed slightly to the right and most were close to zero. In general, only a small number of metabolites contributed greatly to the model prediction. The RIM values of the 4 aforementioned metabolites accounted for 20.92% of the sum of the total RIM value from all 286 metabolites. We further demonstrated that, the RIM, together with LIME, PFI, and SHAP, identified robust metabolite predictors for the MS of other dietary patterns, thus prioritizing top predictors that could greatly facilitate the biological interpretation of the MS.

Session Title: Omics Technologies Poster Session III

PB3309 A novel method for hepatocyte isolation combined with cell size agnostic single cell RNA sequencing enables scalable analysis of liver cells

Authors:

C. Emery[1], M. Gallina[1], D. Loh[1], B. Davis[2], A. Sapre[2]; [1]Miltenyi Biotec Inc., San Diego, CA, [2]Parse BioSci.s, Seattle, WA

The successful application of single cell RNA sequencing (scRNA-seq) heavily relies on the isolation of high-quality, viable single cells from heterogeneous tissues. The difficulty of obtaining viable single cells in some tissues is so great that sequencing nuclei has become the default strategy since it is traditionally a faster, easier process. Single nuclei RNA sequencing (snRNA-Seq) is also the default for especially large cells such as hepatocytes and cardiomyocytes, which are often not compatible with droplet based single cell sequencing instruments. While snRNA-Seq is a well-accepted substitute for scRNA-Seq, it results in the loss of cytoplasmic transcripts and cellular context for binucleated cells, which is commonly seen in hepatocytes. Dissociation of liver tissue into a viable single cell suspension poses challenges due to the unique characteristics of hepatocytes, the main functional cells of the liver. Current methods for the isolation of viable hepatocytes typically require two-step manual perfusion and suffer from limitations such as low throughput, a lack of standardization, and complexity, thereby compromising the scale, reproducibility, and quality of scRNA-seq data for liver. To overcome these challenges, we developed a semi-automated approach for ex vivo liver perfusion of up to 8 rodent livers in parallel that yields intact, viable liver cells suitable for cell culture, flow cytometry, and scRNA-seq analysis. To validate the performance of our method, we conducted flow cytometry and scRNA-seq on isolated liver cells using Parse Biosciences Evercode™ Whole-Transcriptome technology. This instrument free approach based on split-pool combinatorial barcoding is cell size agnostic which allowed us to successfully capture the transcriptomic diversity within the hepatic cell and non-parenchymal cell populations. The resulting scRNA-seq data demonstrated high-quality gene expression profiles, enabling the identification of known liver cell types.This semi-automated method for liver cell isolation complemented with Evercode cell fixation could be scaled up to process dozens of rodent livers in a single day resulting in scRNA-seq data from up to 96 samples and a million cells. In conclusion, we present a novel, scalable method for the gentle, rapid isolation of liver cells ideal for scRNA-seq applications. This method holds great promise for furthering our understanding liver diseases, drug discovery, metabolism, and biodistribution.

Session Title: Omics Technologies Poster Session I

PB3310 A novel method for nucleic acid extraction from FFPE tissue: high compatibility and enhanced convenience with NGS applications.

Authors:

L. Torres Benito[1], T. Schughart[1], I. Rose[2], T. Hess[1]; [1]BioEcho Life Sci. GmbH, Cologne, Germany, [2]Reference Med. Inc., Phoenix, AZ

Formalin-fixed paraffin-embedded (FFPE) tissue samples are invaluable resources for retrospective studies as well as routine diagnostic procedures. However, the extraction of high-quality nucleic acids from FFPE tissues poses significant challenges due to the chemical modifications and degradation caused by formalin fixation. Here, we introduce a novel method for DNA and RNA extraction from FFPE tissues, designed to improve convenience and prove compatibility with downstream applications commonly used in diagnostic and research settings. The EchoLUTION technology addresses key issues encountered during the process. After tissue decrosslinking, paraffin and detergent removal, the DNA is extracted in a single step. The DNA passes through the purification matrix without further interaction while impurities are held back and thereby removed. The workflow leads to a protocol that is up to 70 % faster than with established kits. Further, the novel method does not use xylene or other organic solvents like ethanol, thereby reducing hazardous reagents to a minimum. The performance of the technology was validated by extracting nucleic acids from a spectrum of standardized human FFPE samples. Results determined that high DNA yield, and suitable integrity were obtained with the new method. Additionally, we assessed the compatibility of the nucleic acids extracted with the new method via sequencing downstream applications. For this, we selected tumor-normal sample pairs from brain, breast, colon, and kidney, as well as a commercial FFPE reference standard. We performed targeted gene sequencing using an in-house tumor panel comprising a conventional hybridization capture covering 40 onco- and tumor suppressor genes. Next-generation sequencing (NGS) confirmed all variants expected to be found in the reference standard. Furthermore, somatic driver mutations were identified in all tumor-normal sample pairs, which could be further validated using Sanger sequencing. In summary, our results demonstrate that the new extraction technology offers improved convenience in handling FFPE samples while maintaining compatibility with downstream applications. The streamlined workflow reduces processing time and provides high-quality nucleic acids. The availability of high-quality DNA from FFPE tissues enables comprehensive genomic analyses, facilitating discoveries in molecular pathology, personalized medicine, and beyond.

Session Title: Omics Technologies Poster Session II

PB3311 A novel single-cell *in situ* WGS library preparation method is compatible with target enrichment.

Authors:

K. Zobeck, H. W. Richards, S. F. Sanchez, D. O. Vasquez, A. Campos, X. C. Chen, M. R. Malinas, B. A. Novak, J. D. Wells; Factorial Biotechnologies, San Carlos, CA

Single-cell sequencing has been dominated by methods for detecting RNA expression. In cancer, the source for these expression changes remains difficult to assess with existing single-cell technologies. Two predominant methods currently exist to detect genomic variants in single cells: directly assaying the genome using a single-cell amplicon method and computationally mining variants from single-cell RNAseq data. Both methods have inherent limitations: 1) amplicon panels are limited is the number of targets they can interrogate; 2) current amplicon methods have minimal cell recovery; 3) genomic variant detection in RNA data is limited to coding regions, missing important variants in promoters and introns; 4) expression differences and 3'-bias can exacerbate variant detection limits in RNA data. We describe an *in situ* DNA library prep method that can be combined with existing single-cell technologies to sequence the whole genome. On its own, and with sufficient sequencing, our in situ whole genome sequencing (WGS) library prep could be used to detect large CNVs, structural rearrangements, and SNPs. However, the amount of sequencing needed for sufficient coverage of human WGS data at single-cell resolution is excessively high. We therefore developed our in situ WGS library prep method for compatibility with existing target enrichment technologies. Here, we show how the output of our single-cell WGS library prep can be used with hybrid capture to detect variants in specific target regions of each cell. This method will provide a scalable approach for identifying genomic variants in single cells.

Session Title: Omics Technologies Poster Session III

PB3312 A Protein Signatures-Based Decoy Framework For Improved Protein Identification In Target-Decoy Search Approach

Authors:

T. Bensellak[1], A. Moussa[2,3]; [1]UAE, Tetouan, Morocco, [2]ENSATg, Tangier, Morocco, [3]UAE, Tangier, Morocco

The primary objective of the target-decoy strategy is to estimate the false discovery rate (FDR) for reported peptide or protein matches during a protein database search. Various strategies, such as decoy database generation methods and search engine combinations, have been explored. Earlier research investigated the influence of decoy construction models and showed stochastic/statistic methods to be more promising and accurate than basic sequence reversal or shuffling methods.In this paper, we propose a novel decoy creation framework based on proteins' significant biological signatures, patterns, and profiles, using stochastic models such as Markov. As part of the proposed approach's flexibility, decoy sequence generation can be adapted to digestion sites and be peptide-based or protein-based. For comparison and benchmark purposes, we investigated a standard MS/MS data set of two well-known protein pools based on E. coli peptide fragments to compare the proposed approach to standard methods by assessing the false discovery rate and identification correctness.When compared to default methods, the false discovery rate was quite high. The imbalanced number of discovered patterns in the two pools has resulted in an improved accuracy and specificity for sequences with the most signatures. For certain examined samples, the proposed method improved the correct and incorrect identification ratios by 12.3 percent and 7.7 percent, respectively.

Session Title: Omics Technologies Poster Session I

PB3313 A sequence variation graph-based approach to genotype immunoglobulin genes using short read sequencing data

Authors:

D. Duchen, S. H. Kleinstein; Yale Univ. Sch. of Med., New Haven, CT

Despite the important role our antibodies play in the adaptive immune response to infection, genetic variation across the genes encoding their components remains poorly characterized as most are located in highly repetitive loci with an underappreciated amount of structural variation. For example, within the human immunoglobulin heavy chain locus (IGH) on chromosome 14 there are over 140 functional or pseudogene immunoglobulin heavy chain variable region genes (IGHV), of which many have high sequence similarities, lie within structural variants, or are highly polymorphic with many potential alleles. The characterization of these loci is therefore difficult, especially when using widely available short-read sequencing data. As a result, the genetic variability across the IGH locus is often excluded from host genome-wide association studies. Building upon the limited number of tools able to genotype IGH-related genes at allelic resolution using high-throughput sequencing data, we developed an approach which leverages a whole genome sequence variation graph containing sequences from the International ImMunoGeneTics (IMGT) and other databases for read mapping and alignment. We then infer the observed alleles across the IGH genes using a graph-based approach originally designed to estimate the relative abundance of individual haplotypes/strains from a viral quasispecies. Leveraging available sequencing data from 1000 Genomes Project individuals with a fully-resolved IGH locus for benchmarking, we find our graph-based approach successfully infers the correct IGHV alleles for the majority of IGHV genes using solely short-read Illumina sequencing data.

Session Title: Omics Technologies Poster Session II

PB3314 A Statistical Testing Framework for Structural Difference and Similarity of Multiple Proteomics Networks based on Gaussian Graphical Models

Authors:

J. Chen[1], J. Murabito[2], K. Lunetta[3]; [1]Boston Univ. Sch. of Publ. Hlth., Boston, MA, [2]Boston Univ. Sch. of Med., Framingham, MA, [3]Boston Univ SPH Crosstown Ctr, Rm 313, Boston, MA

Proteomic networks, representing interactions among proteins, play a crucial role in understanding biological pathways and their regulation under different conditions. Differential network analysis aims to identify differences among network components for multiple disease or phenotype states. The Gaussian Graphical Model (GGM) is a statistical network model approach that represents the partial correlations among components. There are several methods for estimating group differences in sparse GGMs, but most apply only to comparisons between two groups, and statistical tests that can rigorously examine similar and differential edges across groups are still underdeveloped: the challenging problem of multiple testing related to the differential and similar structures of multiple GGMs persists. Thus, we propose a new two-step statistical framework that controls the false discovery rate (FDR) in testing structural differences and similarities for multiple groups. We conduct simulations considering GGMs under several common network structures and different proportions of samples across groups to evaluate our proposed framework's performance and compare it with existing methods. We find that our method outperforms existing methods for identifying both group differences and similarities under several network structures such as random and hub graphs. Differential edges in band-structure graphs are more difficult to detect for all methods; for this structure, the joint graphical lasso approach performs better than our proposed method and moderated network models. Based on our simulations we suggest modeling strategies tailored to different network structures. We apply our method to 68 inflammatory biomarkers and three cognitive domain scores measured on a sample of 708 individuals from the Framingham Heart Study categorized into three groups based on Apolipoprotein E (APOE) genotypes. We find 374 similar edges across all the APOE subgroups and 2 edges that are significantly different across the groups. All edges between cognitive scores and proteins were similar across the three APOE groups. Two differential edges: edges between AXIN1 and STAMBP and between CXCL10 and STAMBP, are identified as potentially differential pathways related to neuroinflammation, in that APOE E2 shows positive relationships between CXCL10 and STAMBP whereas negative relationships for APOE E4 and E3E3 subgroups, and different negative strengths between AXIN1 and STAMBP were observed across groups. By applying our framework to high-throughput protein data, we provide a valuable tool for joint comparisons of multiple proteomics networks.

Session Title: Omics Technologies Poster Session III

PB3315 A transformer model for accurate genotype imputation.

Authors:

M. Mowlaei[1], C. Li[1], J. Chen[2], B. Jamialahmadi[3], S. Kumar[4], T. Rebbeck[5], X. Shi[6]; [1]Temple Univ., Philadelphia, PA, [2]Harbin Inst. of Technology, Shenzhen, China, [3]Univ. of Waterloo, Waterloo, ON, Canada, [4]Temple Univ, Philadelphia, PA, [5]Harvard T. H. Chan Sch. of Publ. Hlth., Boston, MA, [6]Temple Univ., Philadephia, PA

Recent advances in Next-Generation Sequencing enable researchers to sequence a whole genome within a day. However, differences in sequencing platforms, assay failures, and genotype calling errors can lead to untyped and missing genotypes. This negatively affects the accuracy of downstream analysis in disease association studies and causal inference. Genotype imputation addresses this problem by predicting missing genotypes based on the patterns found in the reference samples.

Here, we present a transformer model for imputation (STI), which can accurately predict bi-allelic and multi-allelic missing genotypes. Transformers, such as Generative Pre-trained Transformer 4 (GPT4), have shown outstanding performance in many natural language processing and computer vision tasks in recent years. Considering each genotype sequence can be treated similar to a sentence in natural language, we decided to develop STI to tackle genotype imputation. Specifically, our STI model breaks down a sequence into smaller windows and forms a separate deep neural network composed of convolutional and transformer blocks for each window. In STI, we apply convolution blocks to conserve locality in SNP-SNP correlations that create linkage disequilibrium blocks. Although transformers use the attention mechanism to attend to important segments of the input at a global scale, they are costly to apply to long sequences as in human genomes. To address the memory cost of implementing transformers, we used a splitting strategy, like Minimac4 and many other tools, to make the model scalable and efficient in memory and computation cost. The combination of transformers and convolution blocks thus enables STI to outperform other deep learning and classical imputation methods.

We use the yeast and the human 1000 Genomes Project datasets in different settings to extensively evaluate the STI model. Our experimental results showed that STI is competitive with or better compared to the existing imputation models, including Minimac4 and other deep learning methods, in terms of accuracy and imputation quality score metrics. Additionally, our model can impute multi-allelic SNVs with higher accuracy than existing deep learning methods. In summary, we develop a new deep learning-based model for genotype imputation, with potential to be applied to predict missing genotypes in disease association studies and meta-analysis.

Session Title: Omics Technologies Poster Session I

PB3316 Advancing the pace of omics research and discovery

Authors:

E. Floden; Seqera, Barcelona, Spain

Increasingly, multiple omics disciplines are being combined to study biological processes in a more comprehensive manner. Multi-omics techniques are used in everything from personalized medicine and dosing control to population-level studies to the development of new therapeutics. Traditionally, techniques used for large-scale genomic analysis in research and clinical settings have been plagued with technical and procedural challenges, including complexity, reproducibility, auditability, ensuring data provenance, and integrating with public datasets. Bioinformaticians also face a variety of practical challenges, such as marshaling and managing necessary compute resources for analysis, monitoring pipeline execution, managing costs, and maximizing the use of various sequencing platforms. In this talk, the author will discuss his organization's efforts to address many of these large-scale analysis problems. The author will discuss Nextflow, an open-source project for multi-omics data analysis pipelines, and introduce nf-core, a community effort to build a high-quality set of curated analysis pipelines. The author will provide specific examples of pipelines and how they address a wide range of omics disciplines and provide specific examples of how pharmaceutical, biotech companies, and public health authorities are using these open science pipelines and techniques to automate end-to-end analysis. Attendees will learn how they can advance their research and improve productivity and efficiency by leveraging reusable open science pipelines and modules.

Session Title: Omics Technologies Poster Session II

PB3317 † Age-related transcriptional regulation in blood revealed by long-read sequencing

Authors:

M. Tomlinson, K. Small, A. Hodgkinson; King's Coll. London, London, United Kingdom

Ageing is characterised by physiological decline and a growing susceptibility to disease. While the cellular and molecular hallmarks of ageing are well known, emerging evidence suggests altered transcriptional processes play a vital role in cellular ageing. Here, we utilised long-read Oxford Nanopore Technology in a longitudinal design to sequence the transcriptomes of 19 females at two time points approximately ten years apart, enabling high-resolution analysis. Specifically, we focused on two age groups: middle-aged (age at first visit: 52.04±2.19) and older adult (age at first visit: 73.52±2.67). Full-length cDNA sequencing was performed, followed by bioinformatic processing using the 'nanoseq' pipeline, which allowed detailed characterisation of the changing transcriptional landscape during ageing. Comparing transcriptomes between time points, we found decreased transcript usage across samples, accompanied by a corresponding decline in the relative abundance of alternative transcripts. Using paired samples, we found significant differences in transcript usage (FDR 5%) for 55 genes between time points. These genes were enriched in cellular stress response pathways (adjusted $p < 0.05$). We found 122 differential alternative splicing events that were enriched in genes involved in immune processes including negative regulation of NIK/NF-kappa B signalling (adjusted $p = 0.01$) and T-cell receptor signalling (adjusted $p = 0.01$). Moreover, we observed 7 differential intron retention events, which predominantly occurred in genes implicated in reduced natural killer cell count ($p = 0.0064$). Notably, we identified 1104 genes between time points and 735 genes between age groups with significant changes in transcript diversity, measured by Laplace entropy ($p < 0.05$). Functional enrichment analysis revealed these genes were associated with age-related processes, including neurodegenerative pathways ($p = 0.00039$). Importantly, we observed a genome-wide decrease in transcript diversity with age and this effect was most pronounced in older adults ($p = 0.035$). In summary, this study utilising long-read sequencing data provides enhanced resolution and comprehensive profiling of the transcriptional changes associated with ageing in blood. These findings shed light on the dynamic alterations occurring in the transcriptome during the ageing process and the role of alternative splicing.

Session Title: Omics Technologies Poster Session III

PB3318 Alternative Therapeutics for Breast Cancer: An Assessment of Promising Anticarcinogenic Plants Derived from Traditional Medicinal Use; Persimmon Leaf Extract (*Diospiros kaki*), Marjoram Extract (*Origanum majorana*) and Thyme Extract (*Thymus vulgaris*) on *In Vitro* and *In Vivo* systems

Authors:

A. Mattingly[1], D. Ivankovic[2], Z. Vickery[1], J. Spencer[1], L. Boccuto[1], A. Broom[3], HCG SON Laboratory Undergraduate Student Aids; [1]Clemson Univ., Clemson, SC, [2]Clemson Univ., Clemson, SC, SC, [3]Anderson Univ., Anderson, SC

Overview Despite the advancements in targeted radiation and chemotherapy, these treatments still have limitations and pose risks to healthy tissue. However, alternative therapies, including the use of persimmon leaves and Mediterranean herbs like marjoram and thyme, offer potential solutions in the fight against breast cancer. Through our observations, these natural remedies have demonstrated a range of beneficial properties, including anticarcinogenic, apoptotic, pro-metabolite, and antioxidant effects.
Methodology Preparation of Persimmon, Marjoram, and Thyme: The extraction and preparation of alternative therapeutics was completed using microwave-assisted ethanol extraction (ETHOS-X) and subsequent rotary evaporation, followed by a drying period. Cell Culturing and Selection: Cell cultures were maintained according to ATCC guidelines with evidence-based modifications for healthy breast MCF-12A and MCF-12F as control lines, triple-negative HTB-26 and HCC-38, hormonal positive BT-483, and MCF7. Metabolomics: Biolog plates PMM1-PMM8 (Phenotype Mammalian Microarray) were used to collect metabolomics data in the form of normalized spectra and absorbance readings. This measures the amount of NADH produced in the presence of different molecules such as metabolic effectors, amino acids, hormones, ions, and various others. Enrichment, Pathway, and Metabolic Analysis: Used to validate pathway interactions in alternative therapeutics, human cancer cells, and zebrafish biological systems. Metabolism and metabolic analysis are useful to determine which components are causal to therapeutic effect. MTS assay: Conducted to evaluate cell viability, cytotoxicity, and proliferation after treatment with all 3 alternative therapeutics at set concentrations. *In vivo* Zebrafish Injections and Treatment: 5D, Tg(fli1:EGFP), zebrafish (*Danio rerio*) are raised, bred, and kept according to ZIRC and Clemson IACUC husbandry standards. Six breast cancer cell lines are fluorescently labeled in culture, incubated, prepared, and subsequently injected into zebrafish embryos. The embryos are treated to remove pigmentation, allowing for visualization and quantification of metastasis, tumor growth, vasculature, and response to treatment. Gene expression profiles are created using tail digests with detection via qRTPCR and the pathways of interest are investigated.
Future Direction Our preliminary findings indicate high potential for the use of natural compounds in treatment protocols for breast cancer: future studies on mammalian *in vivo* models and human subjects are necessary to validate and expand these results.

Session Title: Omics Technologies Poster Session I

PB3319 Amplicon based NGS using mostly natural sequencing-by-synthesis.

Authors:

J. Sinson[1], M. Pena[1], S. Li[1], S. Zhao[1], P. Liu[1,2]; [1]Baylor Coll. of Med., Houston, TX, [2]Baylor Genetics, Houston, TX

The recent advancement in next-generation sequencing (NGS) technology produced unique platforms that aim to lower the cost and time of sequencing while improving quality. These advances led to an increase in the use of whole genome sequencing (WGS) and whole transcriptome analysis with total RNA sequencing (RNASeq). While we are moving towards a time where WGS is the first diagnostic test to be performed, most clinical laboratories are currently offering targeted NGS panels or whole exome sequencing (WES) along with WGS. Here, we aim to highlight the importance of other NGS assays using one of the newly developed sequencing platforms.Utilizing the UG 100 sequencer created by Ultima Genomics, we were able to develop and sequence our own amplicon based NGS assay that accompanies our WGS. The amplicon based NGS assay was developed as an identification tool to confirm the identity of the WGS samples. It was designed to enable genotyping of 53 SNP loci across the human genome. The 53 loci consist of 49 autosomal and 4 allosomal SNPs with redundancy to ensure the gender of the sample if allelic drop out were to occur. All 53 primer pairs were designed to produce a 600 base pair amplicon and were pooled accordingly to set up in one multiplex polymerase chain reaction (PCR). The addition of the Ultima NGS barcodes to the amplicons was done through a secondary PCR that targets a linker sequence added at the end of each primer. The assay was tested on 12 Coriell samples which consist of 9 lymphoblastoid cell lines and 3 fibroblasts. Using our identification assay, we were able to match the 53 SNPs found in our assay against the same SNPs in the WGS data. Using a similar approach, we developed another assay that verifies low level somatic variants by PCR amplicon sequencing.As the focus in the industry shifts towards generating large amount of data in a short period of time, targeted NGS testing remains a necessity for most clinical laboratories to implement in their operations. Here we highlighted the importance and feasibility of sequencing targeted NGS assays in a newly developed sequencing platform.

Session Title: Omics Technologies Poster Session II

PB3320 An integration of common disease GWAS and omics to fuel rare disease target discovery.

Authors:

Y. Wu, S. Berasi, L. Shang, S. Li, H. Yang, Y-C. Cheng, M. Tadin-Strapps, K. Loomis; Pfizer, Cambridge, MA

Autosomal dominant polycystic kidney disease (ADPKD) is a rare renal disease which leads to end-stage renal disease following enlargement of the kidneys due to the growth and expansion of fluid-filled cysts. ADPKD affects over 300,000 patients in the US. Tolvaptan is the only approved therapy, but the marginal efficacy and hepatotoxicity lead to a large remaining unmet medical need. Human genetics is recognized as an 'experiment of nature' model for drug development. Drug targets with human genetic support are estimated to be nearly twice as likely to result in approved drugs. ADPKD is a ciliopathy caused by mutations in the *PKD1/2* genes; however, the causal genes remain undruggable. To identify genetic modifiers as potential therapeutic targets while being unable to utilize ADPKD specific genome-wide association studies (GWAS) as a starting point, we focused on the ~800 candidate genes possibly associated with renal function from published eGFR GWAS (CKDGen, MGI-HUNT, and Kidney biobank). In addition, as cilia function is a mechanism relevant to ADPKD, we next merged ~950 known/predicted ciliary genes cataloged in CiliaCarta, SysCilia and CilioGenics, and identified 34 overlapping genes. The omics integration further prioritized 16 genes with colocalized eQTLs or SNP-CpG -gene associations by kidney meQTL and eQTM. Eight are predicted targetable using small molecules or antibodies based on tractability assessment. *CDKL5* was identified as a target of interest given a robust eGFR-associated signal (p=1.4x10^{-9}). The same SNP also led the association with blood urea nitrogen (BUN) in a meta-analysis of BioBankJapan, UK Biobank and FinnGen. A kidney-specific eQTL for *CDKL5* was colocalized with the GWAS locus (COLOC PP4=0.96) in GTEx kidney cortex (p=0.0012; not significant in other ~50 tissues) and NEPTUNE tubulointerstitium (p=6.9x10^{-12}). The risk allele for reduced eGFR was associated with higher *CDKL5* mRNA expression, indicating antagonism as a potential direction of treatment. Further supporting our hypothesis, a previously published phosphoproteomic screen demonstrated that CDKL5 kinase substrates are enriched in protein clusters modulating cilium function. Given the GWAS and omics support, target validation experiments evaluated the biological role of CDKL5 in ADPKD using *in vitro* and *in vivo* models. In summary, our multi-omics integrative approach translated GWAS of common diseases/related phenotypes into a novel target hypothesis for a rare kidney disease following the incorporation of known disease biology. This example provides a therapeutic target identification paradigm for rare diseases where direct GWAS are not feasible.

Session Title: Omics Technologies Poster Session III

PB3321 An integrative approach to detect potential blood-based biomarkers for frailty.

Authors:

M. Suganuma[1], M. Furutani[2,1], T. Hosoyama[3], S. Akiyama[1], R. Mitsumori[1], R. Otsuka[4], M. Takemura[5], Y. Matsui[5], Y. Nakano[2], S. Niida[6], K. Ozaki[1,2,7], S. Satake[8], D. Shigemizu[1,7]; [1]Med. Genome Ctr., Natl. Ctr. for Geriatrics and Gerontology, Obu, Japan, [2]Dept. of Cardiovascular Med., Hiroshima Univ. Graduate Sch. of BioMed. and Hlth.Sci., Hiroshima, Japan, [3]GeroSci. Res. Ctr., Natl. Ctr. for Geriatrics and Gerontology, Obu, Japan, [4]Ctr. for Gerontology and Social Sci., Natl. Ctr. for Geriatrics and Gerontology, Obu, Japan, [5]Ctr. for Frailty and Locomotive Syndrome, Natl. Ctr. for Geriatrics and Gerontology, Obu, Japan, [6]Core Facility Admin., Natl. Ctr. for Geriatrics and Gerontology, Obu, Japan, [7]7RIKEN Ctr. for Integrative Med. Sci., Yokohama, Japan, [8]Dept. of Frailty Res., Natl. Ctr. for Geriatrics and Gerontology, Obu, Japan

Frailty is a common geriatric syndrome characterized by age-associated declines in function across multiple physiologic systems and cognitive reserves. The number of people with frailty is increasing with the current rapid population aging, especially in developed counties. A recent meta-analysis reported estimated incidences of frailty and prefrailty of 43.4 and 150.6 new cases per 1000 person-years, respectively. Frailty is associated with a high risk for adverse health outcomes such as physical disabilities and hospitalizations and mortality, but it is reversible and preventable with appropriate interventions. Thus, promising novel biomarkers for the diagnosis of frailty are urgently required. Here, we performed a comprehensive analysis of clinical data, messenger RNA sequencing (RNA-seq), and aging-related factors to detect candidate biomarkers of frailty by using a total of 104 older adults aged between 65 and 90 years, representing 61 frail subjects and 43 robust subjects. We identified two candidate biomarkers of frailty from the clinical data analysis, nine from the RNA-seq analysis, and six from the aging-related factors analysis by logistic regression analysis adjusting for age, sex, and body mass index (BMI). Risk prediction models were then constructed by using combinations of the candidate biomarkers and clinical information (age, sex, and BMI) with a random forest classifier. As a result, the best models used combinations of five biomarkers achieved a high area under the curve (AUC) of 0.95 in an independent validation cohort (95% confidence interval: 0.79-0.97). Our risk prediction models afforded significantly higher AUC than did models constructed using only basic clinical information (Welch's *t*-test $p < 0.001$). All five biomarkers also showed statistically significant correlations with components of the Japanese version of the Cardiovascular Health Study (J-CHS) frailty diagnostic criteria. We believe that further refinement may lead to future clinical use.

Session Title: Omics Technologies Poster Session I

PB3322 Analysis of public long-read transcriptome datasets reveals a putative SINE-VNTR-Alu-intrinsic antisense promoter.

Authors:

A. Lorenzetti[1], T. A. Evans[1,2], J. A. Erwin[1,2], A. Paquola[1,2]; [1]Lieber Inst. for Brain Dev., Baltimore, MD, [2]Johns Hopkins Univ. Sch. of Med., Baltimore, MD

SINE-VNTR-Alu elements (SVAs) are non-autonomous hominid-specific retrotransposons present at approximately 2,700 copies within the human genome. SVAs greatly contribute to the estimated GC-rich tandem repeat differences between humans and other hominids and are considered a potential source underlying divergence of some human-specific traits. Here, we aim to gain insights into their function and impact by analyzing SVA-derived transcripts in long-read sequencing data. SVAs generally comprise five distinct regions, such as i) a 5' (CCCTCT)n tandem repeat, ii) an inverted Alu-like region, iii) a variable number tandem repeat, iv) a SINE-R region, and v) a Poly-A tail. SVA insertions may be pathogenic, causing several diseases, which include but are not limited to Fukuyama congenital muscular dystrophy, autosomal recessive hypercholesterolemia, and X-linked dystonia-parkinsonism (XDP). In XDP, a neurodegenerative movement disorder, the 5' tandem repeat length has been associated with the age at onset and age at death. Due to their repetitive nature, it has been difficult to accurately characterize the endogenous transcripts originating from SVA loci and identify SVA intrinsic promoters. Limitations inherent in short-read sequencing techniques for their transcriptional profiling include ambiguity in read mapping and inability of reads to fully span tandem repeats. Conversely, long-read sequencing technologies may overcome these issues by providing full-length or near full-length transcripts that result in fewer ambiguous alignments and successfully span low complexity regions. With the increasing number of public datasets, we analyzed genome-wide SVA expression and identified a novel putative antisense promoter intrinsic to the 5' tandem repeats. Moreover, we found a subset of SVAs that primarily contribute to the observed overall expression signal. Further investigation into the specific context of this subset of SVA sequences may provide insights into the dynamic behavior of SVAs, their influences on gene expression, and consequently how these elements may contribute to disease-associated phenotypes.

Session Title: Omics Technologies Poster Session II

PB3323 Antibodies and Aptamers: A Comprehensive Evaluation of the Current State-of-the-Art Technologies in High-Throughput Proteomics**.**

Authors:

P. Schreiner[1,2], C. Yandava[3], D. Panyard[4], N. Duro[5,3], D. Dochtermann[3], C. Gauthier[1,3], P. Devineni[3], M. Francis[3,6], B. Gorman[6,3], P. Tsao[7,8], S. Pyarajan[9,3]; [1]Booz Allen Hamilton, Lexington, MA, [2]VA Boston Hlth.care System, Boston,, MA, [3]VA Boston Hlth.care System, Boston, MA, [4]Stanford Univ., Sunnyvale, CA, [5]Booz Allen Hamilton, Tampa Bay, FL, [6]Booz Allen Hamilton, Boston, MA, [7]VAPAHCS/Stanford Univ., Palo Alto, CA, [8]Stanford Univ. Sch. of Med., Stanford, CA, [9]Brigham Women's Hosp., Boston, MA

Proteins are the functional unit of diverse cellular functions whose detection and quantification can be used to inform diagnostic, prognostic, and therapeutic applications. Genetic, epigenetic, transcriptional, and translational abnormalities can all influence protein expression and function, which can directly or indirectly cause disease. Traditionally, these proteins have been measured using methods such as mass spectrometry, enzyme-linked immunosorbent assay (ELISA), or multiplex bead array assays, but new platforms are emerging to detect protein abundance with higher throughput. We profile aptamer- and antibody-based high-throughput proteomics platforms in the Million Veteran Program (MVP), alongside other high-throughput omics data, which presents a unique opportunity to assess platform performance and associate protein levels with disease in a diverse cohort.
In this MVP pilot study, we observe the performance of 8,761 total proteins (6,386 unique) in a common cohort of 829 MVP participants of European, African, and Hispanic ancestry using Olink Explore 1536 (OE1536) and Somascan v4 (SSv4). We discuss considerations and recommendations for best practices regarding data quality control and sample inclusion by platform. Metrics including platform-specific data properties, sample-level and batch variability metrics, technical reproducibility, and concordance of commonly profiled proteins are reported for each platform. We identify 865 statistically significant protein-disease associations with abdominal aortic aneurysm (AAA), 116 significant associations with lung cancer, and 2 significant associations with Alzheimer's disease. To assess the validity of identified protein-disease associations, the top significant protein-disease associations were subjected to review in accepted, peer-reviewed literature. We also perform a genome-wide protein quantitative trait loci (pQTL) analysis to identify 4,449 statistically significant genetic variant-protein associations from 147 unique loci. Finally, we assess the efficacy of proteomics data to predict known covariates - like age, gender, and ethnicity - using machine learning. Our results demonstrate a direct comparison of high-throughput proteomics platforms which provide unprecedented insights into the current state of proteogenomic applications.

Session Title: Omics Technologies Poster Session III

PB3324 Assessing the stability of differential gene expression signatures across assemblies.

Authors:

A. Lund[1], L. Liharska[2], R. Thompson[2], Y-C. Wang[2], A. W. Charney[2], N. D. Beckmann[1]; [1]Icahn Sch. of Med., New York, NY, [2]Icahn Sch. of Med. at Mount Sinai, New York, NY

Reproducibility is a pillar of modern science. Consequently, many studies of complex traits rely on prior work to establish that their results are consistent with what is already known. With constant improvement of genome assemblies, an underlying assumption is that, while massive differences in gene expression can be explained by assembly version, these differences do not significantly impact the gene expression traits that are identified as associated with disease. Here, we present what is to our knowledge, the first systematic attempt to test this underlying assumption. To test this assumption, we utilized a large transcriptomics dataset from Alzheimer's disease (AD) patients and controls (ROSMAP). We aligned the raw RNA-seq reads to 3 genome assemblies that capture the breadth of what has been used in the literature to study that complex trait: hg18, hg19, and hg38. We also aligned to 2 hg38 transcriptome assemblies (GRCh38.p12 (release 30) and GRCh38.p13 (release 43)). To establish a baseline of differences due to assembly version, differential expressed genes (DEGs) were identified between each pair of assembly. Additionally, AD DEGs were compared across assemblies. As expected, we found that for all pairs of assemblies, between 98.8% and 99.3% of expressed genes were identified as significantly differentially expressed (DE). Interestingly, HLA-DRA was found as one of the top DEGs between hg19 and GRCH38.p13. This gene has been notoriously challenging to map, thereby exemplifying gene expression changes due to technical effects that are identified across assemblies. An in depth characterization of the changes of gene expression between assembly could be important to interpret changes in DEG associations to specific traits of interest. AD DEGs (FDR < 0.05) were found to be highly similar across assembly. Unsurprisingly, the greatest difference was between the oldest and newest assemblies (i.e., hg18 and GRCh38.p13) Included in those genes were a set of genes that have been repeatedly shown to be important to AD. Our results show that, while almost all genes are identified as DEGs between assemblies, AD DEGs remain mostly stable across them. While these findings were not unexpected, they nonetheless provide the first proof that the underlying assumption that complex trait DEGs are reproducible across assemblies is valid. It is important to note that, for a subset of specific AD genes, this was not the case, underlying the importance of realigning previous datasets to newer assemblies to ensure the most accurate results and to identify new genes associated to disease that may have been previously missed due to the use of contemporary yet inaccurate assemblies.

Session Title: Omics Technologies Poster Session I

PB3325 Assessing viral variant detection algorithms to improve characterization of gene therapy products and biologics.

Authors:

H. Desai, Y. Zhou, B. Hasson, A. Overgard, A. Bridgeland, W. Dolan; milliporesigma, Rockville, MD

Identity testing is an evolving requirement of regulatory agencies to characterize biologics intended for human therapeutic products such as vaccines and gene therapy. The use of Next generation sequencing (NGS) is increasingly used in the biosafety testing sphere to characterize and confirm the identity of viral seed stocks and expression vectors used in vaccine or gene therapy products. The most common approaches for gene therapy involve using a viral vector to deliver genetic material to patient cells. It is important to characterize viral vectors to ensure they do not contain variants which can negatively impact patient outcomes. Downstream bioinformatics analysis must capture true positive variants and limit spurious results to establish sequence identity and purity of the expression vectors. Advancements in best practices and standards for viral variant detection is critical to ensure the safety of patients and meet the expectations of regulatory guidance.

For identity testing, a diverse range of computational approaches is utilized to analyze 12 open-source and commercial variant callers to examine viral variant calling. In this presentation we will discuss how evaluation of these 12 open-source variant callers including: bcfTools, mutect2, varscan2, freebayes, deepvariant, vardict, variant callers affect sensitivity and specificity of viral variant detection. Specificity, sensitivity, and reproducibility is especially important for patient care which starts with determining the safety of biologics. We use 189 in silico and synthetic viral datasets representing different variants including snps, indels, large insertions and large deletions to conduct studies to determine false positives, false negatives and recommend optimal methods for viral variant detection. By improving the sensitivity and specificity for detecting variants within biologics we can improve the confidence in the results of clinical trials which can benefit clinical outcomes for patients requiring novel approaches to treating genetic conditions or cancer.

Results demonstrate that NGS can detect 1% variant frequencies, integrating multiple variant calling applications is essential for detecting different types of variants, and marking duplicate reads arising from PCR amplification can improve analysis turnaround time while maintaining sensitivity and specificity. This presentation will also demonstrate how utilizing paired-end inward oriented reads for analysis can improve confidence in variant calling by depleting false positive variants.

Session Title: Omics Technologies Poster Session II

PB3326 Automated localization and quantification of RNA transcripts and protein clusters in single cells

Authors:

B. Hospelhorn, B. K. Kesler, G. Neuert; Vanderbilt Univ., Nashville, TN

The regulation of gene expression is a major factor in cells' ability to respond to environmental stimuli and differentiate during the development of multi-cellular organisms. However, a transcriptional program is not only defined by which genes are upregulated and downregulated. Understanding the finer kinetics of transcriptional tuning is important for fully understanding transcription as a process and how its regulation impacts cellular phenotype. Because variability increases with length of time and the number of cells measured in a population, high resolution spatiotemporal localization of individual RNA molecules in single cells is important for studying these dynamics at high precision. RNA fluorescent in situ hybridization (RNA-FISH) is a frequently used technique for visualizing RNA in fixed cells using fluorescent probes. Automated processing of the resulting images is essential for large datasets that may include many different transcripts and time points. Here we demonstrate that our MATLAB based RNA-FISH image processing pipeline is a useful tool for automatically detecting the 3D locations of cell boundaries and RNA transcripts at single molecule resolution in an RNA-FISH image stack. In particular, this tool is effective for facilitating quantitative analyses of FISH data such as determining the colocalization of multiple transcripts or the relative amount of RNA in various subcellular compartments. Our tool also performs well on images with immunofluorescent (IF) and GFP tagged protein targets. We have developed a tool capable of performing analysis for a variety of imaging-based approaches to studying biomolecular dynamics. We will use this tool to advance understanding of eukaryotic transcription regulation mechanisms by studying the quantities and behaviors of transcripts, regulatory RNA, and other targets at discrete time points.

Session Title: Omics Technologies Poster Session III

PB3327 Automating next-generation sequencing: Ultra-long library preparation using a digital microfluidic system.

Authors:

J. Hosea[1], W. Timp[1], C-C. Lee[2], F. Christodoulou[2]; [1]Johns Hopkins Univ., Baltimore, MD, [2]INTEGRA BioSci.s, San Francisco, CA

Introduction: Single molecule sequencing has reached an inflection point where the quality and cost of data generation is now scalable for human studies. However, significant roadblocks preclude wide adoption, specifically because of the difficulty of DNA extraction and library preparation on high-molecule weight DNA. More careful DNA handling through automation would provide greater consistency between libraries and even potentially reduce the cost of prep reagents through a reduction in volume. Here we have implemented nanopore sequencing library preparation on the Miro Canvas, a digital microfluidics platform that allows for automation of a variety of complex NGS sample-prep protocols. The smaller volumes and gentle liquid handling could prove ideal for HMW library preparation.
Methods: We extracted HG002 (GIAB lymphoblast) HMW DNA using NEB Monarch HMW DNA Extraction Kit for Cells & Blood, then size selected using PacBio's SRE XL Kit. We used the Miro Canvas to generate an ultra-long library of HMW HG002 DNA using ONT's LSK110 library prep reagents. The DNA library was then sequenced on the ONT GridION instrument using a R9.4.1 flow cell. Quantification of library yield, sequencing data, and hands-on time were analyzed to assess the Miro Canvas' performance.
Results: The ONT ultra-long protocol automated on the Miro Canvas yielded a final DNA library concentration of 3.68 ng/uL in a volume of ~20 uL from an input of 1,000 ng DNA in a 12 uL volume. On average, the Miro Canvas used 4x less reagents than manual library preparation. The ONT sequencing run done using a MinION flow cell had a yield of 3.09 Gb and generated 174.14 k reads. The estimated N50 for the sample was 43.38 kb. The hands-on time for setting up the reagents for loading on the Miro Canvas was ~30 minutes, while the actual protocol took ~3 hours and 15 minutes to run on the Miro Canvas.
Conclusions: Our preliminary results suggest that the Miro Canvas can produce a high quality ultra-long library for sequencing on the ONT GridION instrument. Raw sequencing data is similar to that we have previously gotten for HG002 cells with manual library preparation. Further experiments and data analysis will be done to generate more consistent high-quality libraries to show Miro Canvas is a satisfactory replacement for manual preparation. Minimal hands-on time and less reagents usage will provide a more time and cost-effective option for NGS.

Session Title: Omics Technologies Poster Session I

PB3328 Bacterial diversity in the COVID-19 Metagenome: An AI-aided approach

Authors:

N. En Najih[1], O. Awe[2,3], B. Onyango[4], E. J. Tettevi[5,6,7], T. E. Lulamba[8], S. L. Msipa[9], F. Annassiri[1], F. Omar[10], A. MOUSSA[1]; [1]Systems and data engineering team, The Natl. school of applied Sci.s, Abdelmalek Essaadi Univ., Tangier, Morocco, [2]Univ. of Ibadan, Ibadan, Nigeria, [3]African Society for Bioinformatics and Computational Biology, South Africa, South Africa, [4]Pwani Univ., , Intl. Ctr. of Insect Physiology and Ecology, Nairobi, Kenya, [5]BioMed. and Publ. Hlth.Res. Unit, Council for Scientific and Industrial Res.-WRI, Accra, Ghana, [6]Dept. of Biochemistry, Cell and Molecular Biology, Sch. of Biological Sci., Univ. of Ghana, Legon, Ghana, [7]West African Ctr. for Cell Biology of Infectious Pathogens, Sch. of Biological Sci., Univ. of Ghana, Legon, Ghana, [8]South African Natl. Bioinformatics Inst., Univ. of the Western Cape, Cape Town, South Africa, [9]Dept. of Hlth.Sci., Univ. of Cape Town, Cape Town, South Africa, [10]Inst. for Biotechnology Res., Jomo Kenyatta Univ. of Agriculture and Technology, Nairobi, Kenya

Background: The microbiomes within the human body play a crucial role in establishing and maintaining immune homeostasis, which is essential for overall human well-being. COVID-19, a respiratory tract infection caused by severe acute respiratory syndrome, enters the body through the oropharynx as one of its main entry points. Metagenomics offers powerful tools for studying the genomes of diverse microbial communities that cannot be easily cultured, allowing for the identification of novel or unexpected pathogens in an unbiased manner. Machine learning (ML) can help overcome challenges in handling large metagenomics data by leveraging advancements in computing power. Methods: A metagenomics dataset from China containing 16S rRNA data was retrieved for the purpose of developing ML tools. Data quality was checked using the DADA2 quality control pipeline, followed by filtering and trimming. Single-end reads were replicated and chimeras were removed. Amplicon Sequence Variant (ASV) tables were generated, which were then used in developing three ML models: LightGBM, Random Forest, and XGBoost. These models were evaluated based on AUC and prediction accuracy for identifying COVID-19 microbial biomarkers using human gut metagenomics dataset. Consensus biomarkers were generated as potential differentially abundant microbes. Results: Our study provides insights into the gut microbiome composition in COVID-19 patients and suggests a possible link between respiratory and gastrointestinal health. The identification of Streptococcus, Bacteroides, Pseudomonas, and TM7x species by all three ML models may indicate an association with COVID-19 or other diseases with similar symptoms. The accuracy rates of the LightGBM, Random Forest, and XGBoost algorithms are 1, 0.935, and 0.978, respectively. These findings have important implications for the potential use of these models in future research related to COVID-19 and related diseases. Conclusion: We have successfully created three machine learning models that can quickly diagnose COVID-19 infection using metagenomics data from the human gut microbiome. This methodology has the potential to be applied to other diseases as well.

Session Title: Omics Technologies Poster Session II

PB3329 Benchmarking enzymatic library preparation workflows in a PCR-free whole genome sequencing context

Authors:

P. Benson, J. Haimes, Z. Jaafar, T. Harrison, S. Mishkin, A. Gaines, D. Wendel, K. Giorda, M. Ranik, B. Kudlow; Watchmaker Genomics, Boulder, CO

PCR-free libraries are considered the gold standard for whole genome sequencing (WGS) with minimal bias. These libraries should contain appropriate insert lengths to make efficient use of 2x150bp read structure; however, size selection must retain yields that will efficiently cluster on the sequencer. Clinical samples are often scarce, thus balancing yield with size selection is an important consideration when selecting a WGS workflow. This study describes the ideal parameters for PCR-free WGS library preparation using eight commercially available library prep solutions, and assesses the sequencing data quality of each one.

PCR-free libraries for WGS were prepared from NA12878 using seven enzymatic fragmentation chemistries and one sonication-based workflow. Libraries were created from 300ng of input, varying size selection parameters to make libraries of similar size. An additional Watchmaker library was prepared from 75ng of input, to demonstrate the workflow's scalability and to determine the minimum viable input. Libraries were generated using both single and double-sided size selection to assess the benefits of removing long library molecules. The libraries were sequenced by 2x150 reads on a NovaSeq 6000, targeting 30X coverage.

Out of all chemistries surveyed, Watchmaker had the highest post-size selection yield by a factor of almost 2. The 75ng input into Watchmaker produced yields comparable to 300ng inputs into commercially available enzymatic fragmentation products. The Watchmaker libraries minimized adapter sequences, indels, softclips, chimeras, improper pairs, and hairpin artifacts identified in the sequencing data. Indicators of variant calling accuracy were consistent with the highest performance workflows. Watchmaker libraries had the highest coverage at difficult-to-cover promoter regions (Ross et al., 2013).

The Watchmaker DNA library preparation workflow offers significantly improved PCR-free yields while minimizing sequencing biases and artifacts. This increases the number of clinically relevant samples that can achieve the desired coverage depth for WGS.

Session Title: Omics Technologies Poster Session III

PB3330 Benchmarking genotype-based demultiplexing algorithms in multimodal single-cell data

Authors:

T. Li[1], M. Alvarez[2], C. Liu[1], K. Abuhanna[3], Y. Sun[1], J. Ernst[2], K. Plath[1], B. Balliu[2], C. Luo[1], N. Zaitlen[2]; [1]Univ. of California, Los Angeles, Los Angeles, CA, [2]UCLA, Los Angeles, CA, [3]Dept. of Human Genetics, David Geffen Sch. of Med. at UCLA, Los Angeles, CA

Multiplexed single-cell sequencing has become a popular experimental design choice to reduce technical variation and sequencing costs. Genotype-based demultiplexing algorithms are a class of methods developed to infer cell-level individual identity and remove doublets without further sequencing. These methods have been benchmarked in single-cell RNA-seq data based on their ability to discriminate doublets. However, these methods are not well understood in other single-cell sequencing assays, such as ATAC-seq and 10X Genomics' joint RNA/ATAC-seq, nor have they accounted for ambient molecules in their modeling. Furthermore, it remains unclear if leveraging variant information from both modalities leads to more accurate deconvolution. To address these challenges, we developed ambisim, a novel read simulator that generates data similar to a 10X Multiome experiment based on user-input proportions of singlets, doublets, and ambient molecules, which allows us to control contamination rates and establish a ground truth for cell assignments. In addition, we develop a novel metric, variant consistency, to estimate lower bounds of ambient molecule contamination in cells. We then evaluated the performance of five widely-used demultiplexing algorithms in both simulated and real 10X Multiome data from a fibroblast to iPSC reprogramming trajectory, multiplexed across four samples.In simulations, we find that all demultiplexing methods cannot accurately classify droplet status in RNA-seq (min AUC = 0.55, max AUC = 0.62) and ATAC-seq (min AUC = 0.56, max AUC = 0.65) based on realistic proportions of singlets, doublets, and ambient molecules. In real data, the agreement of individual assignments between demultiplexing methods within and across modalities is low (average 75%), with most discordant calls arising from singlet vs. doublet cell classification. Finally, our variant consistency metric shows that deeper sequenced variants from ATAC-seq have upwards of 5-fold less contamination than RNA-seq, highlighting ATAC-seq's utility for demultiplexing. Taken together, our results suggest that all existing algorithms perform poorly in multimodal single-cell data and that new ambient molecule-aware demultiplexing algorithms are necessary to recover higher quality cells.

Session Title: Omics Technologies Poster Session I

PB3331 Best practices for perturbation MPRA-a comprehensive evaluation of sequence design strategies

Authors:

J. Liu[1], T. Ashuach[2], F. Inoue[3], N. Ahituv[2], N. Yosef[4], A. Kreimer[1]; [1]Rutgers Univ., Piscataway, NJ, [2]UCSF, San Francisco, CA, [3]Kyoto Univ., Kyoto, Japan, [4]UC Berkeley, Berkeley, CA

Perturbation-based massively parallel reporter assays have enabled the identification of functional elements within non-coding genome regions. However, the gold standard of perturbation sequence design methods remains scant. Here, we utilize a publicly available dataset we recently generated of both wild type (WT) and perturbed (PERT) sequences. The perturbed sequences were carefully designed to target transcription factor (TF) binding motifs in regulatory regions, using three different perturbation approaches. The first two perturbation approaches replace the motif with a constant non-motif sequence while the third approach randomly scrambles the motif's nucleotides. The MPRA readout of each perturbation sequence consists of a) functional regulatory sites (FRS) identification: a multiclass label indicating whether is a functional or non-functional site; b) regulatory effect (Log2FC): a numeric variable.

We first define metrics to assess and compare the effect of the perturbation in each method. We next extract sequence based features of WT and PERT sequences, and subtract the corresponding WT features from each of the PERT features. We use these feature differences as input for several prediction models with FRS identity labels and Log2FC as target values in classification and regression models, respectively.

We find that the third perturbation approach has higher specificity for targeting the motif and the lowest number of net motif change. Moreover, we find that the non-linear models exhibit high robustness for classification and regression in the three perturbation approaches, with the third perturbation outperforming the first two.

Overall, our results suggest that modeling perturbation-based MPRA output is robust and that the random nucleotide shuffling model can serve as a framework for such assessments, leading to improvements in MPRA experimental design.

Session Title: Omics Technologies Poster Session II

PB3332 Bioinformatics investigation of the MAGERS schizophrenia cohort data.

Authors:

S. Rogic[1], G. Poirier-Morency[1], N. Libby[1], P. Carrion[2], M. Lisonek[2], J. Jun[2], N. Verzosa[2], H. Williams[2], P. Laghaei[2], M. Hamer[2], M. Guron[3], A. Elbert[4], C. Cook[5], K. Borle[2], C. Tyson[6], A. Minor[7], M. Hrynchak[7], A. Mungall[8], P. Sullivan[9], W. Honer[10], R. Stowe[11], P. Pavlidis[12]; [1]Dept. of Psychiatry and Michael Smith Lab., Univ. of British Columbia, Vancouver, BC, Canada, [2]Dept. of Psychiatry, Univ. of British Columbia, Vancouver, BC, Canada, [3]Michael Smith Lab., Univ. of British Columbia, Vancouver, BC, Canada, [4]Dept. of Med. Genetics, Univ. of British Columbia, Vancouver, BC, Canada, [5]BC Children's Hosp. Res. Inst. and Dept. of Med. Genetics, Univ. of British Columbia, Vancouver, BC, Canada, [6]BC Children's Hosp. Res. Inst. and Dept. of Pathology and Lab. Med., Univ. of British Columbia, Vancouver, BC, Canada, [7]Royal Columbian Hosp., New Westminster, BC, Canada, [8]Canada's Michael Smith Genome Sci. Ctr., Vancouver, BC, Canada, [9]Ctr. for Psychiatric Genomics and Dept.s of Genetics, Psychiatry, and Epidemiology, Univ. of North Carolina, Chapel Hill, NC, [10]Dept. of Psychiatry and Ctr. for Brain Hlth., Univ. of British Columbia, Vancouver, BC, Canada, [11]Dept. of Psychiatry and Neurology (Med.) and Ctr. for Brain Hlth., Univ. of British Columbia, Vancouver, BC, Canada, [12]Dept. of Psychiatry, Michael Smith Lab., and Ctr. for Brain Hlth., Univ. of British Columbia, Vancouver, BC, Canada

The Metabolic & Genetic Explorations in Refractory Schizophrenia (MAGERS) study is a pilot clinical, -omics and genetic counseling research project conducted in 50 participants with severe, treatment-resistant schizophrenia or schizoaffective disorder. As individuals with severe psychosis are more likely to harbor rare variants of large effect, we set out to identify candidate genetic drivers with the potential precision medicine implications.
All participants underwent deep medical and psychiatric phenotyping, neurological and dysmorphological exams, extensive biochemical/metabolic screening, and a cytokine and inflammatory marker panel; and clinical chromosomal microarrays (CMA). In the first phase of the project, we performed 10x Genomics linked-read whole genome and Illumina short read whole blood transcriptome sequencing on the first 25 participants. Here, we report a comprehensive bioinformatics investigation of the available data.
We developed a variant annotation and prioritization pipeline, which incorporates data from multiple genomic databases and leverages existing tools, such as GATK, Ensembl VEP and Long Ranger. The initial variant prioritization was done using a comprehensive set of criteria, aligned with current industry standards. SNVs and indels were validated using RNAseq data and visually inspected in the IGV browser. Structural variants were compared with CMA calls and validated visually in the Loupe browser. STRs were called with ExpansionHunter. We used both WGS and RNAseq data to compute allele-specific expression, by generating personalized phased haplotypes to which RNAseq reads were aligned. Significant allelic imbalances were inferred using ANEVA-DOT and GTEx whole blood expression data as background.
Finally, we used RNAseq data to search for expression and splicing outliers within our cohort. We used a combination of simple statistical approaches and existing tools, such as FRASER, to identify the most likely candidates for further investigation.
While results prioritization was challenging due to the lack of parental data, and the fact that very few genes have solid genetic evidence for conferring risk, our comprehensive bioinformatics investigation, aided by tools including SCHEMA, Franklin, and Varsome and intensive genotype-phenotype correlation, led to the identification of multiple candidate genes potentially relevant to schizophrenia neurobiology
(including *SETD1A*, *FOXP1*, *AIFM1*, *MDGA1*, *TENM1*, *TENM2*, and *CNOT1*). Overall, 6/25 (24%) of participants harbored a pathogenic/likely pathogenic sequence variant, 2 more had pathogenic CNVs, and all cases harbored at least one intriguing VUS.

Session Title: Omics Technologies Poster Session III

PB3333 Biometric, Information Management and Sharing Strategy to success a Clinical Trial

Authors:

A. Diarra, O. Guindo, M. COULIBALY, B. MARIKO; Univ. Clinical Res. Ctr., Bamako, Mali

Background: Persistent clinical trial data management problems are slowing trials, sparking U.S. Food and Drug Administration (FDA) audit findings and 483s, and putting patients at risk. Many of these problems can be mitigated from the beginning, when a patient attends a clinical research trial for the first time. Objective: Biometric enrollment and verification of patients along the trial, visits management and automatic mobile real time dashboard reporting to ensure the patient safety and trial success. Methods: Centralized Information Management Solution utilizing biometrics to link clinical test subjects to their data at each stage of study, scheduled visits management and real time dashboard reporting for a better adherence of participants on scheduled visits, resulting in improved data management and more complete data. Scheduled visit management is key factor of success but can be relatively challenging with growing numbers of participants, multiplication of sites and duration of the trial. The system provides an extremely straightforward, secure, and accurate way to manage visits, share across study team in compliance with regulations on computer, smart phone. The system provides a dynamic study dashboard, a real time powerful Business Intelligent tool which helps streamline trial operations and improve efficiency in research, it provides also a fantastic report summarizing the number of screened, eligible, non-eligible and participants waiting into the circuit and gives the real time visit completion report by providing for each scheduled visit name, the number of completed, pending, and missed visits.
Conclusion: Our system has also many more useful management tools components such as: Workload prediction in different units during the planning visits, within the scheduled visits, in the sample management. With more accurate, reliable, and completed data, this innovation will result in more efficient development of vaccines for infectious diseases, as well as increased confidence in clinical trial findings.

Session Title: Omics Technologies Poster Session I

PB3334 † Building the spectrum of ground truth genetic variation in a four-generation 21-member CEPH family.

Authors:

Z. Kronenberg[1], K. Munson[2], D. Porubsky[2], C. Nolan[1], W. Rowell[1], B. Pedersen[3], C. Fanslow[1], C. Lambert[1], N. Koundinya[2], W. Harvey[2], K. Hoekzema[2], J. Knuth[2], G. Garcia[2], T. Mokveld[1], E. Dolzhenko[1], S. Watkins[3], D. Neklason[3], A. Quinlan[3], L. Jorde[3], E. Eichler[2], M. Eberle[1]; [1]PacBio, Menlo Park, CA, [2]Dept. of Genome Sci., Univ. of Washington Sch. of Med., Seattle, WA, [3]Univ. of Utah, Salt Lake City, UT

Highly accurate long-read sequencing characterizes the full spectrum of genetic variation across the genome, but variant calling software is still catching up to the sequencing technologies. To develop long-read methods for calling difficult variants and variants in the genomic dark regions, it is important to have a comprehensive ground truth dataset for benchmarking. Until now, most benchmarking datasets were primarily built using short-read technologies that are limited to the easily characterized parts of the genome. We are developing a comprehensive truth set by utilizing the power of genetic inheritance within a four-generation family (CEPH pedigree 1463 plus a newly collected fourth generation) characterized with multiple sequencing technologies (PacBio, ONT, Illumina and Strand-seq) from blood-derived DNA. Large kinship pedigrees provide greater power to establish inheritance patterns when compared to trios; allowing us to adjudicate variant calls across the entire genome and not just in well behaved regions.

As a first demonstration of these data, we identified the chromosomal transmission from the parents (NA12877 & NA12878) to eight consented children. Combining the variant calls made by short and long reads, we identified 6,133,436 small variants (5,121,742 SNVs and 1,011,656 indels) that segregated consistently with the chromosomal transmission. Compared to current GiAB or Platinum Genomes benchmarking data we added an additional 18% and 10% small variants in NA12878, respectively. Our NA12878 long-read specific variants (245,997) were annotated with a high percentage of complex regions including 18% in segmental duplications, 35% in centromeric regions, 40% in repeats, and 7% unannotated. Similarly, in the diverse MHC region we note 43% (13,123 vs. 9,152) more SNVs in our truth set compared to previous work.

We identified that the long-read variants that failed our inheritance validation test mostly fell in centromeric and telomeric regions, where recent work by the T2T consortium is improving our knowledge of the genome. To understand the impact of a more accurate reference for long-read variant calling, we are repeating this analysis using the T2T-CHM13 reference genome. We will present on this ever-evolving benchmarking database including more complex variants such as an analysis of over 1M tandem repeats and an assessment of *de novo* variation where false positive and false negative rates can be assessed using the fourth generation of this pedigree. The sequence data, variant calls, and software required to rapidly benchmark new computational variant discovery tools will be freely available to the community.

Session Title: Omics Technologies Poster Session II

PB3335 Cell-free DNA metagenomic sequencing for pathogen detection and organ dysfunction monitoring.

Authors:

O. Mzava[1], A. P. Cheng[1], A. Chang[1], S. Smalling[1], L-A. Kounatse Djomnang[1], J. S. Lenz[1], R. Longman[2], A. Steadman[3], L. G. Gómez-Escobar[2], M. Salvatore[2], M. J. Satlin[2], M. Suthanthiran[2], J. R. Lee[2], C. E. Mason[2], D. Dadhania[2], E. J. Schenck[2], I. De Vlaminck[1]; [1]Cornell Univ., Ithaca, NY, [2]Weill Cornell Med., New York City, NY, [3]Global Hlth.Labs, Bellevue, WA

There is a significant interest in the use of metagenomic DNA sequencing to screen for infection. Yet, DNA contamination poses a major obstacle, as even small amounts of contaminant DNA can complicate the interpretation of the sequencing results. We have developed Sample-intrinsic microbial DNA Found by Tagging and Sequencing (SIFT-seq), a metagenomic sequencing assay that is robust against contamination. SIFT-seq tags sample-intrinsic DNA directly in plasma and urine with a chemical label that can be recorded via DNA sequencing. Contaminating DNA that is introduced in the sample after this initial tagging step can then be identified and eliminated. For tagging, we implemented deamination of unmethylated cytosines via bisulfite salt treatment of DNA. We applied SIFT-seq to screen for infections from microorganisms with low burden in blood and urine, to identify COVID-19 co-infection, to characterize the urinary microbiome, and to identify microbial DNA signatures of sepsis and inflammatory bowel disease in blood. Moreover, considering SIFT-seq uses bisulfite salt that is routinely used in standard Whole Genome Bisulfite sequencing (WGBS) to perform methylome analysis, we have shown that by leveraging the tissue and cell-type specificity of methylation profiles, SIFT-seq can be used to deconvolve the tissues/cell-types of origin of DNA molecules, enabling the inference of relative contributions of each tissue/cell-type within a DNA mixture. We have tested the utility of this assay in identifying sepsis-causing pathogens and monitor organ dysfunction in a cohort of sepsis patients (n=110) where timely and accurate identification of the sepsis-causing pathogens and precise monitoring of organ failure are critical for improved patient outcomes. We found a strong agreement between microbes detected by SIFT-seq and blood culture. Additionally, our analysis revealed a contribution of DNA from a diverse range of tissues and cell types, with the majority originating from Granulocytes. Furthermore, we observed signatures indicating solid organ damage in sepsis patients. Altogether we have demonstrated an assay that addresses the issue of environmental DNA contamination while simultaneously enabling monitoring of organ damage in patients.

Session Title: Omics Technologies Poster Session III

PB3336 Characterization of cell type RNA isoform composition and alternative splicing events in the adult human cortex.

Authors:

Y. Hadas[1], X. Lin[1], B. Valdebenito Maturana[2], H. van Bakel[2,3,4], D. Pinto[1,2,3]; [1]Dept. of Psychiatry, Icahn Sch. of Med. at Mount Sinai, New York, NY, [2]Dept. of Genetics and Genomics Sci., Icahn Sch. of Med. at Mount Sinai, New York, NY, [3]Icahn Genomics Inst., Icahn Sch. of Med. at Mount Sinai, New York, NY, [4]Dept. of Microbiol., Icahn Sch. of Med. at Mount Sinai, New York, NY

Genetic changes in regulatory regions play an important role in the etiology of neuropsychiatric disorders, resulting in dysregulation of gene expression and splicing. We have demonstrated that isoform-specific expression alterations and isoform networks in postmortem cortical tissue exhibit great effect sizes in psychiatric disorders, capture the largest genetic enrichments, and demonstrate substantial disease and cell type specificity. We have used long-read RNA sequencing (ISO-seq) to construct the highest resolution reference of full-length isoforms in the human prefrontal cortex (PFC) to date (IsoHuB), revealing thousands of novel isoforms at the tissue level. We further expand on this work by mapping isoform expression in the same brain region at a single-nucleus resolution. **Methods:** Full-length RNA isoforms of nuclei isolated from 13 postmortem PFC of control donors were profiled using four complementary sequencing methods including single nuclei RNA-seq (snRNA-seq, 9 donors), single nuclei ISO-seq (snISO-seq, 5 donors), multiplexed arrays isoform sequencing (MAS-ISO-seq; 4 donors) and switching mechanism at 5′ end of RNA template sequencing (SMART-seq-Pro, 4 donors). We added to our analysis publicly available cortical datasets of 8 snISO-seq and 4 SMART-seq-v4 controls. All the resulting sequences were aligned to our IsoHuB reference supplemented with GENCODE 41. The snRNA-seq data were used to quantify gene expression for cell type assignment, while long-read snISO-seq and MAS-ISO-seq data were used for isoform characterization and quantification. SMART-seq data were used for both. **Results:** A total of 106K nuclei across 25 donors passed QC. All 27 cortical cell subtypes, as defined by the PsychENCODE Consortium, were identified using each of the sequencing approaches, including excitatory and inhibitory neurons, glia, and immune cells. More than 160K distinct full-length isoforms were characterized and quantified using snISO-seq and MAS-ISO-seq data, and more than 300K isoforms were quantified using SMART-seq data. We further identified widespread differences in isoform expression and usage between PFC cell types, including isoforms with combinations of distant splice sites that cannot be resolved by short-read RNA-seq data alone. **Discussion:** We identified thousands of cell type-specific isoforms, many of which include novel isoforms in neuropsychiatric risk genes. Our results show that combining short-read and long-read RNA-seq at a single nuclei resolution is an efficient approach for characterizing cell type-specific isoform usage in tissues relevant to neuropsychiatric conditions.

Session Title: Omics Technologies Poster Session I

PB3337 Characterization of optimal protein quantification strategies for sensitive pQTL analyses in human blood plasma samples

Authors:

G. Venkataraman[1], Y. Huang[1], B. Lacar[1], A. Stukalov[1], J. Yao[1], L. Cantrell[1], J. Wang[1], H. Guturu[1], K. Motamedchaboki[1], M. K. R. Donovan[1], M. B. De Geus[2], S. Das[2], P. Kivisäkk[2], S. E. Arnold[2], A. Siddiqui[1], S. Batzoglou[1]; [1]Seer, Inc., Redwood City, CA, [2]Massachusetts Gen. Hosp., Boston, MA

Genome-wide association studies (GWAS) have associated millions of genetic variants with thousands of complex traits that likely act by modifying protein regulation. Protein quantitative trait loci (pQTL) analyses explicitly linking genetic variants to protein expression levels may offer insights into the mechanistic impact of genetic variation on complex traits. However, the large dynamic concentration range present in sample types like plasma makes it challenging to achieve the simultaneous coverage, throughput, and statistical power required for these analyses. Recent advancements in mass spectrometry (MS) proteomics (e.g. the development of the nanoparticle-based Proteograph™ XT [Seer, Inc.]) have facilitated deep, broad proteome analyses. As the application of nanoparticle-based MS proteomics to pQTL analysis is a relatively new advancement, comprehensive assessment of the impact of protein quantification strategy on pQTL results is required. Here, we explore multiple strategies and their impact on pQTL sensitivity.

In this study, we utilized the Proteograph™ XT workflow to identify and characterize pQTLs using blood plasma from 150 subjects (80 Alzheimer's Disease patients, 70 healthy controls). We systematically assessed pQTL sensitivity by surveying results from 121 protein quantification strategies. pQTL analysis was performed using PLINK v2.00a with age, sex, batch, and the first ten genetic principal components as covariates. Protein quantification strategies spanned various methods, including different protein grouping algorithms, filters for proteotypic peptides, normalizations across the protein quantifications, and summarizations of protein abundance from multiple nanoparticle measurements.

Our results showed that some quantification strategies yielded significantly improved pQTL sensitivity over others. Specifically, we found more pQTLs when data for each nanoparticle suspension was considered separately as opposed to summarized. Additionally, we found that using the "Top-N" approach to quantify proteins resulted in improved sensitivity over the "MaxLFQ" approach. There was a range (1-60) in the number of pQTLs found throughout the quantifications, emphasizing the importance of quantification strategy for downstream pQTL analyses; however, 224 of the 279 unique pQTLs found were replicated across quantifications. These results suggest that with the optimal quantification strategy, the application of the new, large-scale Proteograph™ XT workflow enables sensitive pQTL analyses for studying the genetic basis of protein regulation and its implications in human health and disease.

Session Title: Omics Technologies Poster Session II

PB3338 CLCLSA: cross-omics linked embedding with contrastive learning and self Attention for multi-omics integration using incomplete multi-omics data

Authors:

C. Zhao[1], A. Liu[2], X. Zhang[2], X. Cao[3], Z. Ding[2], Q. Sha[4], H. Shen[2], H-W. Deng[2]; [1]Michigan Technological Univ., Houghton, MI, [2]Tulane Univ., New Orleans, LA, [3]Michigan Technological Univ - Houghton, MI, Houghton, MI, [4]Michgan Technological Univ., Houghton, MI

Integrating heterogeneous and high-dimensional multi-omics data is of increasing importance in understanding genetic information. However, a major challenge in multi-omics data integration is the presence of unpaired data, resulting from variations in instrument sensitivity and cost. In this study, we propose Cross-omics Linked unified embedding with Contrastive Learning and Self Attention (CLCLSA), a deep learning method to integrate incomplete multi-omics data for multi-omics data classification. By utilizing complete multi-omics data as supervision, our model employs cross-omics autoencoders to learn feature representations across different omics data. Prior to latent feature concatenation, we leverage multi-omics contrastive learning to maximize the mutual information between different omics types. Additionally, we incorporate feature-level self-attention and omics-level self-attention mechanisms to dynamically identify the most informative features for multi-omics data integration. We conducted extensive experiments on four public multi-omics datasets to evaluate the proposed CLCLSA approach, including, ROSMAP, LGG, BRCA and KIPAN datasets with mRNA expression data, miRNA expression data and DNA methylation data. The experimental results indicate that our proposed CLCLSA produces promising results in multi-omics data classification. Our findings highlight the effectiveness of CLCLSA in addressing the challenges posed by incomplete multi-omics datasets, paving the way for improved understanding of complex genetic mechanisms.

Session Title: Omics Technologies Poster Session III

PB3339 Clinical Validation of AION: A Regulatory-Compliant AI platform for Variant Interpretation in Rare Diseases

Authors:

K. Ibañez[1], A. Bertana[1], E. Verdura[1], M. Schelker[1], J. Rayner[1], I. Lehtinen[1], R. Genomics England[2], D. Salvado[1], D. Neville[1], R. Acuna-Hidalgo[1]; [1]Nostos Genomics GmbH, Berlin, Germany, [2]Genomics England Res. Consortium, London, United Kingdom

Introduction

Genetic variant interpretation is critical in diagnosing rare diseases and guiding personalized treatments. AION is an AI-based software medical device designed to classify and prioritize genetic variants for automated variant interpretation. Regulatory compliance ensures medical software meets quality, safety, and reliability standards, building trust and leading to improved healthcare outcomes. To this purpose, we developed a streamlined regulatory-compliant evaluation pipeline for AION, our variant interpretation platform. Furthermore, we have conducted an external clinical performance validation study of AION using a sample from the 100,000 Genomes Project (100kGP) cohort.

Methods

We established validAION, a regulatory-compliant method for continuous assessment of AION's clinical, analytical and technical performance. Furthermore, our external clinical validation study included 318 individuals (57 singletons, 261 trios) with monogenic diseases and validated causative variant(s), diagnosed by a NHS Genomic Medicine Centre in the United Kingdom, with research consent and WGS data in GRCh37. Individuals with other types of causative variants (CNV, SV, STR, etc) were excluded. AION's sensitivity (detection of causative variants) and variant ranking were assessed.

Results

validAION enabled continuous validation and compliant implementation of changes in AION through rigorous evaluation processes across 1283 cases, including technical and clinical cohort evaluation, end-to-end production cohort assessment, and external clinical validation. In the external clinical performance validation, causative variants were detected in 91.3% of cases (95% CI=88-94), rising to 93.1% for trios (95% CI=89.4-95.6). Higher sensitivity was observed in the pediatric cohort (94%, 95% CI=89.5-96.9) and within the intellectual disability category (96.8%, 95% CI=91.0-99.3). In most cases analyzed, the top-ranked variant by AION was the diagnostic variant reported by clinicians across GMCs.

Conclusions

AION, a web-based variant interpretation software, utilizes genetic and phenotypic information to provide ranked molecular and clinical diagnoses. Our study emphasizes the importance of continuous assessment of AION's clinical, analytical and technical performance through a compliant and streamlined quality assurance process. The successful validation of AION using the 100kGP cohort demonstrates its potential to revolutionize variant interpretation and enhance patient care in the genetics field.

Session Title: Omics Technologies Poster Session I

PB3340 Closing the gap: Targeting complex medically relevant regions of the human genome

Authors:

M. Mahmoud[1,2], H. Corbitt[3], J. Harting[4], X. Chen[4], S. Jhangiani[1], H. Doddapaneni[1], Q. Meng[1], C. Lambert[4], S. Zhang[4], P. Baybayan[4], Z. Kronenberg[4], T. Han[3], C. Riegler[3], J. Farek[1], G. Metcalf[1], R. A. Gibbs[1], M. A. Eberle[4], S. Kingan[4], C. Morrison[3], G. Henno[4], D. Muzny[1], F. J. Sedlazeck[1,2,5]; [1]Baylor Coll. of Med., Human Genome Sequencing Ctr., Houston, TX, [2]Dept. of Molecular and Human Genetics, Baylor Coll. of Med., Houston, TX, [3]Twist BioSci., South San Francisco, CA, [4]PacBio, Menlo Park, CA, [5]Dept. of Computer Sci., Rice Univ., Houston, TX

Next-generation sequencing (NGS) has revolutionized personalized medicine. However, NGS still faces challenges in resolving complex genomic regions characterized by low mappability, extensive repeats, and large duplications. To meet these challenges, we present a cost-effective and scalable approach that combines DNA capture with long-read sequencing. This approach ensures highly accurate genomic analysis. We have developed a comprehensive panel of complex medically relevant genes to be coupled with the Twist target enrichment system. Additionally, we have developed an analysis workflow that leverages the advantages of multiplexed capture, HiFi sequencing, advanced mapping, assembly, and variant calling methods. This workflow also incorporates new, specialized tools. By combining these advancements, we achieved a higher accuracy in single nucleotide variant (SNV) and structural variant (SV) calling. The curated panel comprises 389 complete genes, including *LPA, HLA,* and *SMN1/2*. These genes have been associated with various medical conditions such as cardiomyopathies, neuropathies, and cancer. This framework can potentially enhance clinical applications for personalized medicine, leading to improved patient outcomes, while reducing costs. The approach was first assessed using 12 DNA reference samples, including the well-known HG002, all sequenced with a single Revio SMRT cell. Remarkably, the approach achieved an average gene coverage that was 1.4 times higher than that of HiFi whole-genome sequencing (WGS). This provided an approximate coverage of 38.34x per gene, compared to 26.78x for HiFi WGS. Additionally, our approach provided 1.10 Gbp of HiFi data covering 94% of the target regions with at least 8-fold coverage (HG002). The workflow provided high-quality variant calling even in these highly complex medically relevant regions, achieving 97% accuracy for substitutions, compared to WGS HiFi (97.53%). Regarding indels, we achieved an accuracy of 88% for deletions and 80% for insertions. Additionally, our SV calling demonstrated a high precision of 96% and an F1-score of 94%. We achieved an average gene coverage of up to 67x for the remaining benchmarked samples, accompanied by accuracy rates of 99% for substitutions and 95% for indels. These results highlight the reliability of this approach to accurately detecting variation in medically important genes. We will present data from HiFi WGS, Illumina WGS, and our panel from additional samples obtained from two programs: the HeartCare cardiovascular disease risk program and the GREGoR Mendelian disease discovery program.

Session Title: Omics Technologies Poster Session II

PB3341 Colocalization of proteomics and blood pressure GWAS signals identifies potential novel drug targets for treatment of hypertension

Authors:

A. Pozarickij[1], I. Millwood[1], N. Wright[1], L. Li[2], R. Clarke[1], Z. Chen[1], R. Walters[1]; [1]Univ. of Oxford, Oxford, United Kingdom, [2]Peking Univ., Beijing, China

Proteomics in diverse ancestries provides new opportunities to discover proteins related to human biology and disease, including potential drug targets. We assessed the relationship of the cis-pQTLs with blood pressure (BP) traits in the context of the China Kadoorie Biobank (CKB). We conducted genetic association analyses of 7,301 SomaScan aptamers in 3,965 CKB participants, identifying pQTLs for 4,958 (68%) aptamers targeting 4,256 (66%) proteins. Of these pQTLs, 1,487 (30%) were cis-pQTLs closer than 0.5 Mbp to the corresponding structural gene. We also conducted GWAS in 100,000 CKB participants of four BP traits - systolic and diastolic BP (SBP, DBP), pulse pressure (PP), and mean arterial pressure (MAP) - both with and without adjustment for BMI. We identified 18 protein cis-pQTL loci which included variants associated with at least one BP trait at P<5x10-6. The ARL3 and EFNA1 loci were associated with all four BP traits; conversely, seven cis-pQTL loci were BP trait-specific. We performed finemapping and colocalization analyses for each of these protein-trait pair using SuSiE and Coloc respectively and identified 41 shared signals (posterior probability >0.8) indicating potential causal relationships between protein levels and BP. These shared signals were further evaluated using two-sample bidirectional Mendelian randomization, using either the lead cis-pQTL variants or a genetic score derived from separate published BP GWAS performed in Biobank Japan. In some cases (e.g. GBA), we concluded that elevated BP caused higher protein levels, but not vice versa. For some other proteins (e.g. EFNA1) there appeared to be bidirectional causal effects. Notably, we identified several proteins with evidence for a unidirectional causal effect on one or more BP traits, including some proteins (e.g. ENPEP) with previous evidence for an association with hypertension; these joint signals may indicate novel potential drug targets for control of hypertension.

Session Title: Omics Technologies Poster Session III

PB3342 Combined whole-genome and Sanger sequencing analyses of *GBA* revealed a pitfall depending on the reference human genome sequence.

Authors:

K. Orimo[1], J. Mitsui[1], T. Matsukawa[1], M. Tanaka[2], J. Nomoto[2], H. Ishiura[1,3], Y. Omae[4], Y. Kawai[4], K. Tokunaga[4], T. Toda[1], S. Tsuji[1,2]; [1]The Univ. of Tokyo, Tokyo, Japan, [2]Intl. Univ. of Hlth.and Welfare, Chiba, Japan, [3]Okayama Univ. Graduate Sch. of Med., Dentistry and Pharmaceutical Sci., Okayama, Japan, [4]Natl. Ctr. for Global Hlth.and Med., Tokyo, Japan

Introduction: *GBA*, encoding glucocerebrosidase, is the causative gene for Gaucher disease. Furthermore, *GBA* variants that are causative for Gaucher disease are also associated with an increased risk of Parkinson's disease and multiple system atrophy (MSA). *GBA* locates near the highly homologous pseudogene (*GBAP1*), and gene-pseudogene rearrangements are not infrequently observed in patients with Gaucher disease, hence, identification of *GBA* variants by short-read sequencing is challenging. **Methods:** Genomic DNAs obtained from 502 cases of MSA were included in this study. We searched for *GBA* variants in these cases employing two pipelines, including short-read whole genome sequence analysis (NovaSeq 6000, Illumina) and Sanger sequencing. For Sanger sequencing, genomic DNAs were amplified using three primer pairs designed to selectively amplify exons 1-5, exons 5-8, and exons 9-11 of *GBA* but not *GBAP1* followed by direct nucleotide sequence analysis. In 170 cases, the reads were aligned to the reference GRCh38/hg38 (hg38) using BWA-MEM, and variant calling was performed using GATK4. In 332 cases, the mapping to hg38 and variant calling were performed using Clara Parabricks, a GPU-accelerated program that provides equivalent functionality to the BWA-MEM and GATK. Long read sequencing analysis employing PacBio Sequel II with circular consensus sequencing reads were further conducted to discriminate variants located in *GBA* and *GBAP1*. **Results:** While eleven kinds of variants were called by either of the pipelines, there were inconsistencies with the four variants. c.115+1 G>A (NM_00157.4) was called only by short-read sequencing. Long read sequencing analysis identified a structural variant in *GBAP1*, where exons 2-11 of *GBA* is inserted in *GBAP1*, resulting in calling c.115+1 G>A in *GBAP1* erroneously as a variant in *GBA*. p.K505K and p.L483P were only detected by Sanger sequencing, whereas p.A495P/p.V499V, which are located on the same allele, were only detected by short-read sequencing. These inconsistencies were due to the rare variants registered in *GBAP1* of hg38, which are identical to those of *GBA*. When hg38 is used in the mapping pipeline, inconsistencies in calling the variants of *GBA* were observed. Correct alignment was accomplished when hg19 was used as the reference genome. After validating the results, 3/502 (0.00598) cases carried p.L483P, which is higher than the previously reported incidence in control subjects (2/1509, 0.0013). **Conclusions:** We demonstrated the accuracy of variant calling may depend on the reference genome used in particular for genes with highly homologous paralogs and on the structural variants involving these genes.

Session Title: Omics Technologies Poster Session I

PB3343 Comparative analysis of computational algorithms for spatial transcriptomics using REVEAL

Authors:

C. Bragdon, K. Sharma, M. Colosimo, J. Zahiri, Z. Pitluk, S. Sarangi; Paradigm4, Waltham, MA

Single cell multiomics sequencing has made impressive advancements in the past decade with rise of multiple techniques to profile the genome, transcriptome, and proteome. However, certain challenges persist - inducing stress responses, death, and aggregation in cells during the extraction process and loss of the spatial relationships (inter-cellular, cell-matrix, and cell clusters) due to spatial information not being captured. To address these challenges, spatial transcriptomics (ST) has emerged as a groundbreaking solution.Depending on the methodology, ST captures the transcriptional profile of individual cells or small groups of cells along with their spatial positions in relation to the tissue slice being analyzed. The spatial context provides invaluable insights into unraveling cellular phenotypes/cell states, novel cell types, physiological connections between cells, and the differences in physiology between healthy and diseased states. We had previously demonstrated a bioinformatics platform called REVEAL for storing, querying, and analyzing spatial transcriptomics datasets. Here, we extend that work by demonstrating a workflow in REVEAL for evaluating various analytical methods using numerous publicly available ST datasets. We tested an optimized version of the function AddModuleScore from the popular Seurat package. This function can be used to check expression of gene signatures, for example, a set of genes that characterize a particular cell state. Next, we performed a comparative analysis between AddModuleScore and another gene signature scoring algorithm called UCell. Finally, we show results from evaluation of SpaGCN - an algorithm that integrates gene expression, spatial location and histology to identify spatial domains and spatially variange genes by graph convolutional network.Advancement in computational tools and methods play a pivotal role in extracting meaningful biological insights from complex datasets such as ST. However, one of the challenges is often the ease of use for executing these different algorithms at scale across multiple datasets. Through this work, we highlight the application of a user-friendly workflow in REVEAL for evaluating multiple analytical methods.

Session Title: Omics Technologies Poster Session II

PB3344 Comparative analysis of multiple single-cell RNASeq platforms

Authors:

C. Kao[1], H. Qu[1], C. Hou[1], M. Hufnell[1], D. Slater[1], F. Wang[1], K. Wang[1], H. Hakonarson[1,2]; [1]Children's Hosp. of Philadelphia, Philadelphia, PA, [2]Univ. of Pennsylvania, Perelman Sch. of Med., Philadelphia, PA

Single cell transcriptomics, or scRNASeq, has experienced transformative progress especially within the last ~5 years, in large part driven by the accessibility of droplet microfluidics technology to partition each cell robustly into uniquely barcoded, individual gel-beads and encapsulating the cell/bead pair in an emulsion containing the requisite cDNA-generating reagents. The impact of scRNASeq on basic and translational studies in genomics and cellular biology has been exponential. However, a microfluidics platform requires the use of separate, specialized instrumentation which can be a bottleneck from both cost and technical perspectives. Recently, alternative scRNASeq platforms have become commercially available based on either successive split-pool barcoding techniques or proprietary emulsion reagents that generate single-cell/hydrogel droplets via self-assembly by mixing/vortexing. We performed scRNASeq using the following assays/platforms: 1) Parse Biosciences split-pool barcoding assay, 2) Scale Biosciences split-pool barcoding assay, 3) Particle-templated instant partition sequencing (PIP-Seq) from Fluent BioSciences, 4) 10x Genomics (10xG) Chromium single-cell 3’ gene-expression assay, v3.1, and 5) 10xG Chromium single-cell flex assay (v1). About 2,500-5,000 cells from 4 peripheral blood mononuclear cell (PBMC) samples derived from pediatric subjects with hypoplastic left heart syndrome (HLHS) were processed on each platform. Fixed cells were used for all assays except the standard 10xG 3’ gene-expression assay. The cDNA libraries generated from each type of assay were processed further according to manufacturers' instructions to generate Illumina-compatible sequencing libraries, then subsequently sequenced on the NovaSeq. For 4 of the 5 platforms, full-length cDNA libraries were generated, with the exception of the 10xG flex assay (which uses a probe-based method instead of priming and extension from 3’ poly-A tails). Thus, an aliquot from each full-length cDNA library was also processed and sequenced on Oxford Nanopore to generate long-read transcriptome datasets. The quality of both the short- and long-read transcriptome datasets were analyzed and compared, using the 10xG 3’ gene-expression assay as the benchmark. Additionally, the advantages and disadvantages/limitations of each platform with regards to processing and capture efficiency/input cell numbers were also assessed.

Session Title: Omics Technologies Poster Session III

PB3345 Comparing HG38 to T2T-CHM13 and the need for new methods of annotation: non-syntenic genes, additional gene matches, and WASHC1 gone wild? Genomics. #itscomplicated

Authors:

M. Page, M. Wadsworth, B. Aguzzoli Heberle, J. Miller, M. Ebbert; Univ. of Kentucky, Lexington, KY

Background: In 2022, the T2T consortium released the first truly complete human genome, CHM13, that is already impacting genomics research—including resolving errors in hg38. For most genomics work however, a reference is only as insightful as its annotation. As CHM13 grows in popularity, understanding both structure and annotation differences between CHM13 and hg38 is critical to truly understanding what we have lost, gained, and changed with CHM13. Here, we highlight important differences between the two (including mistakes and oversights) not to criticize the excellent work already accomplished, but to demonstrate the need for improved methods and importance of accurate annotations. Here's the truth: Genomics. #itscomplicated, and we aim to help improve the human reference genomes and their annotations.
Methods: We compared annotations between Ensembl hg38 v101 & CHM13 UCSC GENCODEv35 CAT/Liftoff v2, identifying genes that were not strictly syntenic and chose one category to interrogate: non-syntenic genes (NSG). We BLAT'ed NSG sequences from hg38 to both the full CHM13 reference and back to hg38. We also compared to the newer CHM13 JHU RefSeqv110 + Liftoff v4 annotation after manually adding Ensembl IDs. Finally, we compared with long-read RNASeq.
Results: We identified 1,193 genes that were not strictly syntenic between hg38 & CHM13, including 68 NSGs (i.e., changed chromosomes). Of the 68 NSGs we verified only 39 matched CHM13 v2 annotations (requiring >90% identity); using liftover, however, only 9 matched. We also identified 229 additional unannotated BLAT matches from 9 NSGs, representing potentially 229 additional copies of these gene bodies. 6 NSGs were annotated >10kb larger in hg38 than CHM13. Comparing v2 with v4 annotations, 23,093 gene Ensembl IDs were missing in v4—likely because of compatibility challenges. Of the 68 NSGs, only 18 were found in the v4 annotation with gene ids. Some NSGs remained as annotated in CHM13 v2, some reverted to their original hg38 locus, some changed chromosome again, while others reverted to their original chromosome but were also duplicated elsewhere. *WASHC1* is an interesting example, where based on long-read RNASeq, none of the annotations appear accurate.
Conclusions: As the first complete genome, CHM13 brings powerful insights, but also highlights we still have much to learn. Specifically, our results suggest that there remain many inaccurate annotations across both references and liftovers can be wildly inaccurate, highlighting the need for better annotations and methods. Genomics requires more than just a reference—accurate annotations are vital to interpretation. In short, Genomics: #itscomplicated

Session Title: Omics Technologies Poster Session I

PB3346 Comparison of 5' and 3' capture techniques for performing pooled CRISPR screens in differentiated neurons.

Authors:

A. Anderson, B. B. Rogers, K. Trausch-Lowther, J. N. Cochran, L. F. Rizzardi, R. M. Myers; HudsonAlpha Inst. for Biotechnology, Huntsville, AL

Pooled CRISPR screens with single-cell sequencing, such as Perturb-seq and CROP-seq, have improved our ability to profile genomic perturbations and assess their individual effects on gene expression on a tremendous scale. The method of capture and barcoding of transcripts are a fundamental component of such assays that can have a significant impact on the interpreted results. 10x Genomics offers two methods for performing single-cell CRISPR screens. The first is their more established 3’ CRISPR assay, and the second is the newer and currently recommended 5’ CRISPR assay. We set out to identify any technical differences and confounding factors between using the 5’ and 3’ reagent sets for our future studies. We compared the performance of both kits under identical experimental conditions applied to CRISPRi screens in excitatory neurons (KOLF2.1J NGN2) with a pool of 130 targeted gRNAs and four safe harbor gRNAs profiling 16,472 and 20,750 cells in the 5’ and 3’ kits, respectively. We identified over 500 genes that were differentially measured between the two approaches. Additionally, we compared target genes that were significantly knocked down when the gRNA was present and found that the DEGs were often significant in both approaches. We observed similar rates of gRNA detection between the two approaches; however, the 5' kit exhibited significantly higher numbers of gRNA UMIs per cell than did the 3’ kit, increasing confidence in gRNA calls. Our findings provide a comprehensive comparison of the two approaches under controlled conditions and identified significant technical differences. These differences can be taken into consideration when making direct comparisons between previously published 3’ studies and future 5’ work. We demonstrate that despite these differences, the identification of knocked down target genes remains unaffected, ensuring the reliability of data generated using both kits.

Session Title: Omics Technologies Poster Session II

PB3347 Comparison of structural variant calls from Oxford Nanopore haplotype resolved and telomere-to-telomere genome assemblies

Authors:

S. Mckenzie[1], S. Nurk[2], A. Dawson[2], S. Juul[3], P. Rescheneder[2]; [1]Oxford Nanopore Technologies, Spokane, WA, [2]Oxford Nanopore Technologies, Oxford, United Kingdom, [3]Oxford Nanopore Technologies, New York, NY

Structural variants (SVs), including large insertions and deletions, duplications, inversions, and translocations, represent a major source of genetic variation in human populations. These variants have historically been difficult to detect with array and short-read based approaches, however, hindering efforts to understand their importance in human health and disease. Modern long-read sequencing technologies have drastically improved our ability to identify and characterize SVs, recently culminating in complete "telomere-to-telomere" (T2T) *de novo* genome assemblies which can represent the full range of human sequence variation. Generating T2T genomes remains challenging and costly and requires a combination of multiple sequencing data types. It is therefore important to compare T2T assembly variant calling to other available methods - including raw read alignment and other haplotype-resolved assembly-based methods - to determine if and when T2T assembly is necessary.
Here we compare SV calls from alignment based, haplotype-resolved *de novo* assembly, and T2T *de novo* assembly approaches using Oxford Nanopore simplex and duplex whole genome sequencing data. We validate our calls against both Genome in a Bottle (GIAB) HG002 SV benchmarks and calls derived from the Telomere-to-Telomere Consortium's multi-technology HG002 assembly. We demonstrate very high accuracy SV calls from all approaches. Variant discovery accuracy within Genome in a Bottle high-confidence regions was comparable across the different approaches. However, sequence-resolved accuracy was higher for *de novo* assembly-based methods than current raw read alignment methods. *De novo* assembly approaches also showed higher concordance with Telomere-to-Telomere Consortium validation variants in challenging genomic contexts outside the GIAB high-confidence regions. T2T *de novo* assembly outperformed other *de novo* assembly methods, though the gains were marginal for the majority of SVs in most genomic contexts, therefore haplotype-resolved *de novo* assembly may strike the best balance between cost and accuracy for the majority of SV detection applications

Session Title: Omics Technologies Poster Session III

PB3348 Comprehensive characterization of V(D)J recombination using long-read transcriptome data

Authors:

K. Hu[1,2], Z. Chong[1,2]; [1]Dept. of Genetics, Heersink Sch. of Med., Univ. of Alabama at Birmingham, Birmingham, AL, [2]Informatics Inst., Heersink Sch. of Med., Univ. of Alabama at Birmingham, Birmingham, AL

V(D)J recombination, a genetic process in developing B and T cells, is crucial for generating the diversity of the immune repertoire. Novel long-read transcriptome sequencing technologies, such as PacBio Iso-Seq and Nanopore Direct RNA-seq, deliver full-length transcript sequencing reads. Unlike short reads, long reads can capture the entire V(D)J sequence in a single read, thereby enhancing the potential accuracy of V(D)J detection. This study introduces VDJCrafter, a robust pipeline specifically developed for in-depth analysis of V(D)J recombination in long-read transcriptomes. The pipeline begins by aligning long-read Iso-Seq data to the latest human GRCh38 reference genome using Minimap2. Following this, a bed file comprising V(D)J chromosome locations is created using GENCODE annotation. Subsequently, candidate long reads overlapping the V(D)J regions are extracted using the bed file, and then are aligned to the IMGT V(D)J database via BLAST, yielding a V(D)J gene list complete with sequences and annotations. Finally, CDR1 and CDR2 are generated on germline V genes by IMGT CDR consensus sequences and CDR3 through V(D)J assembly, and then are translated into respective amino acid sequences. The results of VDJCrafter performance on 12 Iso-seq datasets from HGSVC are compared with TRUST4, a short-read V(D)J identification tool. VDJCrafter successfully identified full-length V(D)J gene sequences that exhibited up to 100% alignment with known V(D)J sequences in the IMGT database. V gene subclasses from long reads showed approximately 60.74% overlap with short reads, while D gene and J gene detection demonstrated overlaps of roughly 96.42% and 100%, respectively, between long and short reads from the same sample. When examining the top enrichments of light chain recombination of immunoglobulin, 8 out of the top 10 recombinations were of the same type, with closely aligned percentages in long and short reads. These observations are supported by validation against short reads and the application of the standard immune repertoire from IMGT database. We also found that V(D)J genes in cancer samples displayed distinct types of enrichment compared to healthy samples, indicating that disease samples are likely to exhibit unique V(D)J recombination patterns. In conclusion, our findings suggest that long reads outperform short reads in V(D)J recombination identification with higher accuracy. The results above indicate that long-read sequencing provides more robust information to elucidate V(D)J recombination, potentially discovering novel genes and recombinations, thus enhancing our understanding of the immune repertoire in diseases.

Session Title: Omics Technologies Poster Session I

PB3349 Comprehensive, integrated long-read genomic, epigenomic, and transcriptomic characterization for solving the molecular basis of Mendelian conditions.

Authors:

M. Vollger[1], J. Korlach[2], K. Eldred[1], E. Swanson[1], J. Underwood[3], J. Ranchalis[1], Y. Mao[1], E. Blue[1], U. Schwarze[4], C. Saunders[5], A. Wenger[6], I. McLaughlin[7], A. Souppe[6], C. Lambert[6], A. Allworth[1], S. Chanprasert[1], I. Glass[4], M. Horike-Pyne[1], K. Leppig[8], E. Rosenthal[1], S. Sheppeard[1], S. Strohbehn[1], P. Byers[1], M. Bamshad[1], F. Hisama[4], G. Jarvik[9], T. Reh[1], Y. Sancak[1], K. Dipple[10], A. Stergachis[1]; [1]Univ. of Washington, Seattle, WA, [2]PacBio, Camas, CA, [3]PacBio, Seattle, WA, [4]Univ of Washington, Seattle, WA, [5]PacBio, San Diego, CA, [6]PacBio, Menlo Park, CA, [7]Pacific BioSci.s of California, Inc., Menlo Park, CA, [8]Kaiser Permanente of Washington, Seattle, WA, [9]Univ Washington Med Ctr., Seattle, WA, [10]Seattle Children s Hosp. and Univ. of Wash, Seattle, WA

Highly accurate long-read sequencing has the potential to dramatically enhance our understanding of the genetic basis of Mendelian conditions through the improved detection of genetic variation (e.g., HiFi sequencing), as well as the elucidation of the functional impact of genetic variation on RNA splicing (e.g., Iso-Seq), and gene regulation (e.g., mCpG detection and Fiber-Seq). To test the utility of these long-read approaches for evaluating individuals with Mendelian conditions, we combined HiFi sequencing, bulk MAS-Seq, and Fiber-seq to enable the simultaneous long-read multi-omic profiling (i.e., genome, methylome, chromatin epigenome, and transcriptome) of patient samples using a single sequencing run.

First, we developed a machine learning approach for Fiber-seq data to accurately assemble single-molecule chromatin accessibility, nucleosome positioning, and transcription factor occupancy using the N6-methyladenine stenciled-patterns produced during the Fiber-seq reaction. Application of this approach to GM12878 and HG002 cells permits the genome-wide identification of haplotype-specific chromatin architectures. In addition, Fiber-seq data retains highly accurate variant calling (SNV F1 99.94%) and the diploid de novo genome assembly capabilities from a single PacBio Revio sequencing run (contig N50s 36.3/28.8 Mb).

We next applied long-read multi-omic profiling to fibroblasts from a participant within the Undiagnosed Diseases Network (UDN) with unexplained developmental delay, polymicrogyria, sensorineural hearing loss, and bilateral retinoblastomas, with a known chromosome X-13 balanced translocation of uncertain significance. Long-read genomic data identified that the translocation causes NBEA haploinsufficiency, explaining her developmental delay. Long-read transcriptomic data identified that the translocation also creates a PDK3-MAB21L1 fusion transcript that is presumed to result in PDK3 gain-of-function, which inhibits pyruvate dehydrogenase activity, explaining her features associated with mitochondrial and pyruvate dehydrogenase dysfunction. Finally, long-read epigenomic data identified that the translocated chromosome containing both XIST and RB1 is subjected to X chromosome inactivation (XCI) in ~5% of cells, resulting in the silencing of the RB1 locus that is located 13 Mb away from the translocation breakpoint, thus representing the “first hit” for predisposing her to retinoblastomas. Overall, we demonstrate the utility of long-read multi-omic profiling for resolving the mechanisms underlying unsolved rare conditions.

Session Title: Omics Technologies Poster Session II

PB3350 ContextSV: A novel computational method for calling structural variants and integrating information across sequencing platforms.

Authors:

J. Perdomo[1], K. Wang[2]; [1]Children's Hosp. of Philadelphia, Philadelphia, PA, [2]CHOP, Philadelphia, PA

Structural variants (SVs) are genomic alterations generally larger than 50 bp which form the largest source of human genome variation. SV identification and characterization are important for gaining insight into the genetic basis of diseases: Identifying SVs associated with specific clinical phenotypes empowers clinical diagnoses and enables researchers to identify potential molecular mechanisms for these genomic alterations. Emerging long read sequencing technologies such as the Oxford Nanopore (ONT) and Pacific Biosciences (PacBio) platforms provide the resolution required to resolve larger and more complex SVs. Nevertheless, variable error rates in these technologies possibly result in a high false positive rate and low robustness for SVs detected using only long read sequencing data. The rich repertoire of available technologies, such as Illumina short read sequencing and Bionano optical mapping, can be leveraged to resolve these limitations. Here we introduce ContextSV, a novel structural variant calling method which uses a hybrid approach to achieve high-accuracy, robust SV calling: Long read sequencing data is used to identify SV candidates, while short read sequencing provides high-accuracy sites for resolving breakpoints in complex SVs, and optical maps are used as long-range scaffolds for high-quality read assembly prior to running SV detection algorithms. To improve accuracy, we train a binary classification model which is used to score candidate SVs based on coverage and genomic context, which are key features for SV validation. We then filter likely false positives by thresholding SVs based on score. Finally, we plan to incorporate support for pangenome graph reference formats in ContextSV: Pangenome graph structures can better represent common haplotypes in the human population relative to a single linear reference genome, and thus they would form a more comprehensive reference for identifying SVs. Large efforts led by projects such as the Human Pangenome Reference Consortium (HPRC) aim to release a pangenome graph reference representing a large, diverse set of human genome sequences, and thus there is a growing importance for future SV callers to provide support for these formats. In summary, our ContextSV method enables capturing large, complex SVs with high accuracy and robustness by leveraging information across multiple genome analysis technologies and using a machine learning model to compute confidence scores, while providing support for future developments in pangenome graph reference formats.

Session Title: Omics Technologies Poster Session III

PB3351 Cost-effective, high-resolution whole transcriptome sequencing for gene fusion detection applications

Authors:

J. Pavlica[1], J. Haimes[1], P. Benson[1], K. Reed[1], T. Harrison[1], N. Iremadze[2], D. Wendel[1], T. Clark[2], A. Kathiria[2], A. Jaimovich[3], B. Kudlow[1]; [1]Watchmaker Genomics, Boulder, CO, [2]Ultima Genomics, Newark, CA, [3]Ultima Genomics, Menlo Park, CA

As per-base sequencing costs continue to decline, there is less of a need to restrict reads to pre-identified regions of interest through targeted sequencing. This is particularly relevant for RNA analysis, where cost-effective sequencing allows for a hypothesis-free whole transcriptome sequencing (WTS) approach - useful for identifying new features such as gene fusions. We optimized the Watchmaker RNA Library Prep Kit with Polaris Depletion for sequencing on the Ultima Genomics platform, an ultra-high capacity, single-end read sequencing system, and evaluated the workflow for gene fusion detection via WTS.

The Watchmaker RNA-seq solution generates high-complexity libraries in under 4.5 hours, including rRNA depletion, and maintains stranded information via uracil incorporation during the synthesis of the second strand. During library amplification, the polymerase stalls at uracils resulting in amplification of only the first cDNA strand. To create a workflow compatible with Ultima, however, an alternate method for second strand removal is necessary. We optimized a uracil DNA glycosylase incubation process that works in tandem with the adapter ligation reaction, effectively degrading the second strand without increasing workflow time. Several workflow variables such as RNA mass, template quality, library insert size, and sequencing depth were evaluated to determine their impact on data quality and interpretation.

The UG 100 system can handle longer (300 bp and greater) reads, which is advantageous for accurately mapping split reads important for fusion detection. We evaluated various library insert sizes with both intact and FFPE RNA to understand the performance impact of longer inserts versus including shorter molecules to increase library complexity. We also evaluated the effect of insert length on sequencing quality metrics from the Ultima system, which employs emulsion PCR for clonal templating. We performed an impact assessment of insert size on key Ultima sequencing metrics, as well as final library complexity and fusion detection sensitivity. Overall, we demonstrated robust performance across a wide input range from 1 to 1000 ng, as well as with FFPE control materials. Overall, results indicate excellent compatibility of the Watchmaker RNA Library Prep Kit with Polaris Depletion solution with the Ultima UG 100 system for whole transcriptome sequencing applications, including fusion detection.

Session Title: Omics Technologies Poster Session II

PB3353 CRAM compression: Practical across-technologies considerations for large-scale sequencing projects.

Authors:

A. Al Ali[1], P. Kandavel[1], H. Al Mabrazi[1], G. Carvalho[2], V. Kusuma[1], G. Katagi[1], S. Elavalli[1], A. Yousif[1], M. Akhter[1], J. Mafofo[1], T. Magalhaes[1], J. Quilez[1], F. Al Marzooqi[1], The Emirati Genome Program; [1]G42 Hlth.care, Abu Dhabi, United Arab Emirates, [2]UCLA, Los Angeles, CA

CRAM is an efficient format to store high-throughput sequencing data and it has been widely adopted. We thus plan to use CRAM for the Emirati Genome Program, which aims to sequence the genomes of ~1 million nationals in the United Arab Emirates using short- and long-read sequencing technologies (Illumina, MGI and Oxford Nanopore Sequencing). We conducted a pilot study on the three technologies before start using CRAM at scale. We found CRAM achieved 40-70% compression depending on the sequencing platform. As expected, CRAM compression was data lossless and did not alter variant calls. In our cloud, we observed compression speeds 0.7-1.4 GB per minute, varying on the sequencing platform too. This translates into ~1-2 hours using a single CPU to compress a ~30X human whole-genome sequencing sample. Despite its wide use, we found little publicly available information about CRAM compression rate, speed, losslessness and parallelization, especially across many sequencing platforms. This work will have direct application for Emirati Genome Program and provide practical considerations for other large-scale sequencing efforts.

Session Title: Omics Technologies Poster Session III

PB3354 CRISPR/Cas9 depletion of high expression genes in human fibroblast samples increases the diagnostic potential for rare diseases

Authors:

R. Corona[1], G. Carvalho[1], L-K. Wang[1], A. Y. Huang[1], D. Ramachandran[2], J. Diaz[2], M. F. Khalid[2], O. Abinader[2], S. Ranganathan[2], A. Siddique[2], K. Brown[2], J. R. Armstrong[2], S. F. Nelson[1]; [1]UCLA, Los Angeles, CA, [2]Jumpcode Genomics, San Diego, CA

Whole genome sequencing (WGS) improves identification of rare variants compared to whole exome sequencing (WES), but diagnostic rates for rare genetic disorders remain modest. To enhance diagnostic yield, whole transcriptome RNA sequencing (RNA-Seq) is combined with WGS to assess functional consequences of genetic variation. However, RNA-Seq data is most useful for observing splicing, mono-allelic expression, or outlier expression of more highly expressed genes. This limits the global utility of using any one tissue source for RNA-Seq to augment WGS data. To partially improve observation of more low expression genes using RNA-Seq, we present an approach to selectively remove highly expressed genes from RNA-Seq on large scale by CRISPR/Cas9 cleavage and size selection to enrich the observation of mRNAs of lower expression transcripts. As implemented here, this method targets 4450 expressed genes (TPM > 30) in RNA-Seq data from cultured fibroblasts using 353,628 different guides. Virtually all genes are reduced in abundance in the treated library relative to the untreated library, and on average there is a 5-fold increase reads that are not targeted (TPM < 30) in the sequenced library, and the depletion of library fragments from highly expressed genes increases the number of genes observed above a TPM of 30 by 4,477 at similar sequence depth. Thus, large scale library depletion can increase the proportion of an RNA-Seq library of less well observed genes and results in enhanced assessment of rare splicing aberrations and allelic expression alterations, increasing interpretability of WGS data. Because there is high gene expression correlation across technical replicates within both the targeted genes and non-targeted genes, the method, in effect, broadens the dynamic range of RNA-Seq, and can be designed for any tissue type. This advancement provides a valuable tool for understanding and diagnosing rare genetic disorders, unraveling disease mechanisms, and identifying potential biomarkers and therapeutic targets.

Session Title: Omics Technologies Poster Session I

PB3355 Cross-platform Hi-C meta-analysis identifies functional insulators that actively block enhancer-promoter interactions

Authors:

W. Xu[1,2], J. Cui[1,2], S. Zhang[1,2], L. Lu[1,2], X. Liu[1], Y. Li[1], F. Jin[1,3,4]; [1]Dept. of Genetics and Genome Sci., Sch. of Med., Case Western Reserve Univ., Cleveland, OH, [2]The BioMed. Sci. Training Program (BSTP), Sch. of Med., Case Western Reserve Univ., Cleveland, OH, [3]Dept. of Computer and Data Sci., Case Western Reserve Univ., Cleveland, OH, [4]Dept. of Population and Quantitative Hlth.Sci., Case Comprehensive Cancer Ctr., Case Western Reserve Univ., Cleveland, OH

Insulator protein CTCF controls genome architecture through forming thousands of cohesin-dependent structural loops. However, genome-wide studies only found mild transcriptional consequences upon acute CTCF-depletion, raising confusions about how CTCF regulates enhancer-promoter (E-P) interactions and gene expression. Here we reanalyze independent Hi-C, in situ Hi-C, and micro-C data in mouse embryonic stem cells upon acute CTCF-, RAD21-, and WAPL-depletion; DeepLoop is used to enable robust comparison of orthogonal Hi-C data at kb-resolution regardless of sequencing depth. All datasets show that most loops are lost upon CTCF depletion, but E-P interactions are enriched among the retained loops, and interestingly a small number of newly gained loops repressed by CTCF.From multiplatform Hi-C data, we identified several hundred recurrent events in which new E-P interactions form after the insulating CTCF loops disappear. We therefore define FINs (functional insulators) as CTCF sites that actively insulate their flanking sequences. In CTCF-depleted cells these newly gained E-P interactions require cohesin activity. WAPL-depletion causes relaxation of FIN loops and abolish insulator functions. Importantly, CTCF-repressed genes are enriched near FINs, but CTCF-dependent genes are enriched near TAD-boundaries. We also validated the transcription regulatory functions of several FINs with CTCF-blocking assays. Taken together, DeepLoop meta-analysis unifies multiplatform Hi-C data and demonstrated that FINs, but not TAD-boundaries, are bona fide insulators.

Session Title: Omics Technologies Poster Session III

PB3357 Deciphering cell functions with micrometer precision, leveraging ultra-high-resolution spatial transcriptomics datasets

Authors:

Y. Si[1], C. Lee[2], G. Jun[3], S. Zöllner[1], J. Lee[1], H. Kang[1]; [1]Univ. of Michigan, Ann Arbor, MI, [2]Genetics Dept. HMS, Boston, MA, [3]Univ. of Texas Hlth.Sci. Ctr. at Houst, Houston, TX

Recent advances in single-cell and spatial multi-omics technologies are rapidly improving the resolution of spatial transcriptomics. Microscopic-resolution (<1μm, compared to typical cell size 8~100μm) spatial transcriptomics technologies, such as Seq-Scope or Stereo-seq, assay transcriptomes over hundreds of millions of spatial barcodes per tissue section. Each spatial barcode, however, captures only a small number of genes, producing a notably sparse digital expression matrix. As a result, typical procedure in analyzing high-resolution spatial transcriptomes relied on fixed coarsely patterned segments or (nuclear) staining-based segmentations, compromising their microscopic resolution or expression information. To overcome these limitations, we developed an efficient computational method for cell function decoding at the level of individual spatial barcode (pixel) without requiring external, image-based cell segmentation. Our method annotates millions of unique pixels to cell functions or tissue conditions using stochastic variational inference under a model that resembles latent Dirichlet allocation.
We applied our method to the injured mouse colon dataset generated by Seq-scope and the mouse embryo dataset generated by Stereo-seq, which contain tens to a couple hundred millions of spatial barcodes spaced at ~0.5μm resolution. In mouse colon data, our pixel-level decoding method revealed the fine-scale architecture of colon tissue, profiling cellular heterogeneity in epithelial, connective tissue and smooth muscle layers responding to physical injury with remarkable preservation of tissue texture, such as smooth muscle orientation and colonic crypt organization. Different immune cell populations infiltrating into different tissue layers were also profiled at microscopic level. In the developing mouse embryo dataset, our method effectively captures spatially organized gene expression modules that represent cellular functions as well as developmental processes. It detects cell-level boundaries between as well as within tissues, and utilizes transcripts that were typically thrown out in conventional nuclear segmentation methods. Our method enables large scale inference of spatial transcriptomics fully realizing the potential of the microscopic-resolution of the cutting-edge technologies.

Session Title: Omics Technologies Poster Session I

PB3358 Deep coverage full-length isoform characterization and quantification using highly accurate long-read sequencing

Authors:

D. Wissel[1], M. Mehlferber[2], V. Pavelko[2], F. Reese[3], M. D. Robinson[1], G. Sheynkman[2]; [1]Univ. of Zurich, Zurich, Switzerland, [2]Univ. of Virginia, Charlottesville, VA, [3]Univ. of California, Irvine, Irvine, CA

Long-read RNA sequencing resolves transcript structures at high resolution. Recent studies (Glinos et al. 2022, Abood et al. 2023) have shown the utility of long-read RNA sequencing data in interpreting the molecular mechanisms of disease-associated splicing, for example, by connecting GWAS-colocalized sQTLs with the impacted protein isoforms. Despite these developments, the high sequencing cost and lower throughput on previous PacBio systems have hindered scalable isoform classification and quantification over large numbers of samples and broader use in functional genomics.

Recently, a methodological breakthrough of an amplicon concatenation approach—MAS-Seq (Al'Kafaji et al. 2023) - has enabled an order-of-magnitude higher throughput in sequencing coverage while maintaining high read accuracy (average Phred 35). PacBio has recently commercialized this approach for bulk isoform sequencing, and in collaboration, we obtained MAS-Seq bulk Iso-Seq data on the PacBio Revio system of the WTC-11 stem cell line, covering 6 different time points (Day 0 through 5) with replicates (11 samples). We included strategic spike-ins of multiple SIRV and ERCC mixes and generated over 600 million full-length reads ranging from 100-10,000 bp (average: 2.5 kb). The transcript length distributions were similar to traditional non-concatenation libraries, and had CAGE and PolyA signal support, indicating a level of comprehensiveness of sampling of full-length transcripts that has not been reported before.

With this dataset, we address open questions about required read depth for isoform discovery, quantification, and read saturation, including read-number equivalence to Illumina data on the same samples. We detected 22,201 annotated genes and a total of 119,891 isoform models that were common across all samples. We detected 30,614 (25.5%) novel transcripts not found in the GENCODE comprehensive annotation. Data were highly reproducible (gene-level Pearson's r ranged from 0.93 to 0.96 across replicates). Rarefaction curve analysis revealed complete saturation of gene and isoform identification at 10-20M reads per sample, indicating future opportunities for multiplexing.

These samples also were subjected to deep coverage MS, and we will describe novel ORF prediction pipelines for patient-specific proteomics, concurrently capturing amino acid variants (de Souza et al. 2022, unpublished) in addition to splicing, for complex disease and cancer immunotherapeutic applications. These studies and rigorous benchmarks lay the groundwork for the application of population-scale transcriptome analysis for functional genomics and disease splicing analysis.

Session Title: Omics Technologies Poster Session II

PB3359 † Deep long-read RNAseq in human brain identifies *de novo* RNA isoforms, novel gene bodies, and complex isoform diversity for Alzheimer's disease genes

Authors:

B. Aguzzoli Heberle[1], J. Brandon[1], M. Page[2], K. Nations[1], M. Wadsworth[1], K. I. Dikobe[1], S. Goodwin[3], E. Ghiban[4], R. Wappel[4], S. Mavruk Eskipehlivan[4], P. Nelson[1], J. Miller[1], J. Fryer[5], M. Ebbert[1]; [1]Univ. of Kentucky, Lexington, KY, [2]Univ. of Kentucky, Lexington, KS, [3]Cold Spring Harbor Lab., Cold Spring Harbor, NY, [4]Cold Spring Harbor Lab, Cold Spring Harbor, NY, [5]Mayo Clinic, Scottsdale, AZ

Intro: We often discuss genes as if they have a single function, yet human genes average 7 RNA isoforms, potentially resulting in 7 distinct proteins/functions. Due to limitations in short-read RNAseq, researchers are forced to collapse all isoforms into a single gene measurement—a major oversimplification. Long reads, however, span entire RNA molecules in a single read, allowing accurate quantification of all isoforms, including de novo isoforms. Accurately quantifying isoforms further enables differential isoform expression analysis and reveals isoform diversity within a single tissue/cell—a major advantage over short reads, enabling researchers to begin discerning specific functions for each isoform within a "single" gene (e.g., two BCL-X isoforms: XL = anti-apoptotic, XS = pro-apoptotic). Here, we perform deep long-read RNAseq in human brain to: (1) discover all isoforms across all genes (including de novo gene discovery); and (2) begin exploring functions for individual RNA isoforms in human health and disease. Methods: We sequenced 12 frontal cortex postmortem aged human brain samples (6 Alzheimer's (AD) cases & 6 controls, 50% female) using 1 Oxford Nanopore PromethION flow cell per sample (~40M mapped reads/sample). Analysis included pychopper, minimap2, bambu, and DESeq2. We only report high-confidence isoforms/genes. Results: We discovered 267 new high-confidence spliced gene bodies expressed with median Counts Per Million (CPM) > 1 and 437 new high-confidence RNA isoforms in annotated genes (9 new spliced mitochondrial isoforms w/ 4 PCR confirmed), where 53 are from medically relevant genes, including MAOB and POLB. We identified 7042 genes expressing multiple RNA isoforms in one tissue, including key AD genes: MAPT (4), TARDBP (4), APP (5), PSEN1 (5), and BIN1 (8), demonstrating the diversity of a "single" gene and the need to determine individual isoform function. As proof of concept, we identified 23 differentially expressed isoforms (FDR $p < 0.01$ & $|log2FC| > 1$) between AD cases and controls, where the gene (all isoforms collapsed as a single measurement) was not differentially expressed. Conclusions: We identified hundreds of new high-confidence RNA isoforms and gene bodies, demonstrating significant gaps remain in our understanding of RNA isoform diversity. More importantly, we began to explore RNA isoform function for every gene by quantifying individual isoform expression levels in human brain tissue. We demonstrate that performing differential gene-level expression is important, but insufficient, and suggest that deep long-read RNAseq is necessary to understand the full complexity of transcriptional changes during disease.

Session Title: Omics Technologies Poster Session III

PB3360 Deep-learning based bulk RNA-seq deconvolution method recovers cell-type specific signals.

Authors:

C. Tang[1], Q. Sun[2], J. Wen[3], Y. Li[4]; [1]UNC Chapel Hill, Chapel Hill, NC, [2]UNC-Chapel Hill, Chapel Hill, NC, [3]Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, [4]Univ North Carolina, Chapel Hill, NC

Bulk RNA-seq deconvolution is distinctly useful in revealing cell type specific (CTS) gene expression profiles by combining the advantages of bulk and single-cell(sc) or single-nuclei (sn) RNA-seq data. The former, available in target samples at tissue level, tend to have larger sample size and high signal-noise-ratio, while the latter, available as reference, are at single cell level resolution. Distinct from traditional cell-type-deconvolution methods that infer only cell type proportions for each target sample, we aim to estimate their CTS expression for every gene, by integrating with sc/sn RNA-seq reference. Here we propose DL-unmix, a deep-learning based method that utilizes sc/sn RNA-seq reference for model training and applies the trained model on target samples' bulk RNA-seq data to infer a sample by gene by cell type expression matrix. DL-unmix first constructs pseudo bulks from sc/snRNA-seq reference, and then builds a Deep Neural Network to predict its CTS expression without using the truth. After model convergence, bulk RNA-seq data with cell type proportions estimated using traditional deconvolution method will be treated as model input to obtain inferred CTS expression. Since it is challenging to accurately estimate CTS expression for all genes in each cell type, DL-unmix includes a pre-training process that solely utilizes the reference data to select gene-cell type combinations that can be "well inferred" to guide pre- and post-training filtering. For simulation-based evaluation of DL-unmix, we split mouse primary motor cortex snRNA-seq data as reference and target set, created pseudo bulks for the test set, and applied DL-unmixed and alternative methods (TCA and bMIND). With pretraining-guided filtering, DL-unmix demonstrates best performance: correlation between estimated and true CTS expression improves by up to 50%, compared to TCA/bMIND. We then applied DL-unmix to deconvolute ROSMAP real bulk RNA-seq data with ROSMAP 48-donor snRNA-seq data reference. As the downstream analysis, we identified differentially expressed genes (DEG) between Alzheimer's disease (AD) cases and controls for each cell type. We ran functional enrichment analysis for these CTS DE genes via g:Profiler. Among others, we found enrichment of the UDP-glucosyltransferase activity GO pathway for DEGs identified in astrocytes. The finding is consistent with previous studies which identified strong links between expression levels of glycosyltransferases and neurodegenerative diseases including AD. We believe DL-unmix will be a novel, valuable method for revealing CTS insights.

Session Title: Omics Technologies Poster Session I

PB3361 Detection and confirmation of structural variants in a human trio using three Next-Generation Sequencing technologies.

Authors:

K. Hilkert[1], T. Oleksyk[2], K. Shchubelka[1], W. Wolfsberger[1], S. O. Castro-Marquez[1]; [1]Oakland Univ., Rochester, MI, [2]Oakland Univ. - Biology, Rochester, MI

A structural variant (SV) is defined as a DNA rearrangement of more than 50 nucleotides such as inversions, translocations, insertions, and deletions. SVs, which can be called by long and short read sequencing, are a root cause and risk factor of various human diseases including autism, schizophrenia, and HIV. As a result, accurate detection is crucial for future diagnosis and treatment. Assessment of SV callers' reliability, sensitivity, and precision has revealed that they may fail to detect SVs and report false positives. Understanding the precision and accuracy of SV callers is important as a trade-off exists between the two for identifying SVs, for this reason, we require a more robust protocol for validation of such genomic events. Currently, there is no gold standard for how to call SVs from sequencing data. We hypothesize that some SVs will be falsely called by proprietary callers and the utilization of various technologies will highlight this low accuracy.

Our study utilizes whole genome sequencing and PCR product-based verification to evaluate the accuracy of SV callers by various sequencing and computational methods. A family trio was sequenced using several different technologies including long read Oxford Nanopore Technologies, along with that of Illumina and BGI short read approaches. Our efforts included combining data analyzed to detect and classify various SVs found in the trio. Raw sequencing data were interrogated for the presence of SVs, and the resulting dataset was then filtered to obtain variants of 50 nucleotides or greater in size. Next, the data from the various platforms were compared to evaluate their concordance. The resulting dataset of discordant SVs was verified using PCR for the purposes of independent validation.

Although concordance exists, a discrepancy is seen, highlighting the fact that we cannot rely on one technology for valid proof of an SV's existence. The accuracy of a single SV caller has inherent limitations, we found this roadblock is remedied by comparing SV callers combined with a PCR method of validation. This protocol has proven to be useful for increasing confidence to enable this methodology to be taken seriously in a future clinical setting.

Session Title: Omics Technologies Poster Session II

PB3362 Detection of *de novo* mutations in incomplete trio with missing parent information.

Authors:

S. Lee, M-h. Choi, J-W. Nam; Hanyang Univ., Seoul, Korea, Republic of

Trio-based (parent and child) whole genome sequencing studies have enabled the detection of de novo mutations (DNMs) by identifying mutations present in the child's genome but not in the parent's genome. However, the genomic information of both parents is often not available for various reasons. Here, we present a caller to identify DNMs when one or both parental genomes are missing. The unit of DNM calling in this situation is also a trio, but with the genomes of siblings replacing those of the missing parents, which are referred to as incomplete trios - "linear trio" takes into account two siblings instead of the missing parents, and "reverse trio" takes into account one sibling instead of one missing parent. For these scenarios, our method can call putative DNMs in genomic regions where both of a proband's haplotypes are found in other incomplete trio members. Hidden Markov Models are used to infer sibling inheritance states, and callable regions are classified based on a combination of inheritance states with careful filtering methods. Theoretically, DNMs can be predicted in 56.25% of a proband's genome for linear trios and in 50% of a proband's genome for reverse trios. Our method was validated on artificial incomplete trios from CEPH families consisting of multiple complete trios, showing an average precision of 75.3% and a recall of 89.2% for linear trios, and an average precision of 68.5% and a recall of 91.7% for reverse trios. The DNMs also showed a significant paternal age effect for both types of trios. We also estimated the sample size required in a case-control study to investigate the effect of a given exposure on the number of DNMs, which revealed that 92 linear or 110 reverse trios are required to detect a 5% difference in DNMs under current false positive rates. We believe that our work provides a practical means to detect disease-related DNMs in older probands, especially with sporadic late-onset disorders and in the absence of parental genomic information.

Session Title: Omics Technologies Poster Session III

PB3363 Detection of germline copy number variation from whole genome sequencing on UG100: A cost-effective and robust approach.

Authors:

T. Biniashvili, S. Pollock, D. Bogumil, G. Hornung, P. Winer, I. Soifer, O. Barad; Ultima Genomics, Newark, CA

Copy Number Variations (CNVs) play a pivotal role in genetic diversity and have been widely acknowledged for their clinical significance, particularly in relation to numerous genetic diseases. With the advent of whole-genome sequencing, there has been a significant breakthrough in uncovering copy number variants on a genome-wide scale, both in population studies and clinical applications.

In this study, we present a comprehensive pipeline, leveraging an optimized version of cn.mops, to accurately detect germline copy number variations (CNVs) using whole-genome sequencing data generated on the novel UG100 sequencing platform. To ensure reliable CNV calling for variants longer than 10Kb, we employed a cohort-based approach that effectively normalizes sequencing biases. To validate the accuracy of our pipeline, we conducted a thorough comparison of our CNV calls with those obtained from the NYGC cohort, consisting of samples from the 1000 Genomes Project. Notably, we achieved a high average precision and recall concordance of 98.5% for calls across all samples. It is worth mentioning that this concordance assessment was performed in an "apples-to-apples" manner, where the CNV calling tool was independently applied to each dataset, and the resulting CNVs were compared within the high-confidence regions.

Furthermore, we demonstrated the efficacy of our pipeline by successfully identifying known pathogenic CNVs in 16 clinical samples. These samples exhibited lengths ranging from 100Kb to 35Mb, as well as instances where entire chromosomes were affected. The associated diseases for these samples encompass a diverse spectrum, including DiGeorge syndrome, Miller-Dieker syndrome, Albinism, Angelman syndrome, Williams-Beuren syndrome, Smith-Magenis syndrome and Wolf-Hirschhorn syndrome. This validation underscores the robustness of our methodology and its potential in clinical settings.

Additionally, we showcased the versatility and cost-effectiveness of our approach by conducting analyses across a wide range of sequencing coverages, spanning from 5X to 47X per sample.

In summary, we present an optimized version of cn.mops using UG100 whole genome sequencing providing a robust and cost-effective method that holds promise for both research and clinical applications.

Session Title: Omics Technologies Poster Session I

PB3364 Different but similar: Systematic investigation of proteogenomic variation in men and women and relevance for human diseases

Authors:

M. Koprulu[1], E. Wheeler[1], N. Kerrison[1], S. Denaxas[2], H. Hemingway[3], N. Wareham[1], M. Pietzner[4], C. Langenberg[5]; [1]MRC Epidemiology Unit, Univ. of Cambridge, Cambridge, United Kingdom, [2]Inst. of Hlth.Informatics, UCL, London, United Kingdom, [3]UCL Inst. of Hlth.Informatics, London, United Kingdom, [4]Univ. of Cambridge, Cambridge, Germany, [5]Precision Hlth.care Univ. Res. Inst., Cambridge, United Kingdom

Sex is an important factor involved in almost all biological processes and profoundly affects human health and disease. The mechanisms underlying these differences remain largely unknown and have not been systematically studied. Here, we conducted a large-scale investigation of sex differences in the genetic regulation of the human plasma proteome across 4,979 protein targets assessed in 4,403 women and 3,945 men and their relevance for human disease.

Plasma levels of almost two-thirds (60.5%, n=3,105) of measured proteins significantly differed by sex, but genetic effects on protein targets were remarkably similar between men and women with very few (0.3%, n=15 of 4979) protein quantitative loci (pQTLs) showing evidence of statistically significant differences (heterogeneity p-value <1.01x10-11). Of the 15 identified sex-differential pQTLs, only one cis pQTL associated with PATE4 (prostate and testis expressed 4) showed sex dimorphic effects, i.e. different effect directions in men versus women. Heterogeneity of the most of the remaining 14 loci were driven by small differences in directionally consistent, significant effects in men and women. Systematic evaluation of associations between the identified sex-differential pQTLs and 1,312 disease outcomes did not reveal any significant heterogeneity between men and women (heterogeneity p-value < 3.8x10-5). Our results emphasise a conserved genetic regulation of the plasma proteome between sexes and suggest that differences in levels are not explained by germline genetic effects. With few exceptions, our results suggest that pQTLs identified in sex-combined analyses that provide greater statistical power can be utilized for downstream applications such as genetically guided drug discovery, validation and causal inference across sexes.

Session Title: Omics Technologies Poster Session II

PB3365 Dissecting cis-regulatory interactions in the Major Histocompatibility Locus using single cell CRISPR/dCas9-based regulatory element screening.

Authors:

L. Bounds[1], A. Barrera[1], M. ter Weele[1], E. Wu[2], S. Liu[1], S. Li[1], P. Giusti-Rodriguez[3], T. Reddy[1], Y. Li[2], R. Gordan[1], M. Love[2], P. Sullivan[2], A. Allen[1], G. Crawford[1], C. Gersbach[1]; [1]Duke Univ., Durham, NC, [2]Univ North Carolina, Chapel Hill, Chapel Hill, NC, [3]Univ. of Florida Coll. of Med., Gainesville, FL

The Major Histocompatibility (MHC) locus is the most SNP-dense region in the human genome and is linked with >100 polygenic disorders. However, it remains unknown which variants affect gene expression. To dissect cis-regulation within the locus, we performed noncoding CRISPR/dCas9-based regulatory element screening with targeted single-cell RNA-sequencing in human induced pluripotent stem cells (iPSCs), iPSC-derived neural progenitor cells (NPCs), and K562 cells. We designed a gRNA library targeting a union set of 537 putative regulatory elements (pREs) based on accessible chromatin regions across all three cell types and 44 previously identified POU5F1 enhancers with TSS-targeting positive controls and nontargeting negative controls for a total of 12,723 gRNAs. We engineered iPSCs, NPCs, and K562 cells to express the repressor dCas9-KRAB, and iPSCs and NPCs to express the activator dCas9-p300 and transduced each cell line with the gRNA library at high multiplicity of infection. We profiled 1.2 million single cell transcriptomes at seven to nine days post-transduction. To assign pRE-gene links, we performed differential expression testing between all cells in which a given gRNA was observed versus all cells in which a different gRNA was observed.

Across all five screens, we identified a pRE for 72.4% of genes detected in the locus and connected genes to 36.1% of all perturbed regions, with 53.7%-75.3% of pRE-gene pairs spanning <100kb and 9.0%-17.0% spanning >1Mb. For all cell types, the overlap of active histone marks increases as the pRE-gene link becomes more significant. Concordantly, 94% of pREs in K562 cells are also in accessible chromatin in that cell type. In contrast, 20%-40% of pREs in iPSCs and NPCs are not in accessible regions in these cell types and in some cases also overlap the repressive mark, H3K27me3.

We next investigated cell-type specificity. 63.8% of pREs displayed function in only one cell type and the iPSCs had the greatest number of unique links. Given the association of the MHC locus with schizophrenia (SCZ), we analyzed the overlap of SCZ GWAS SNPs with the linked pREs. We observed significant enrichment of overlap in the NPCs, but not iPSCs or K562 cells, consistent with the etiology of this disorder. Finally, we delivered individual gRNAs to the iPSCs and confirmed the change in expression of the linked target gene for 11/11 tested links. Collectively, this study clarifies the cis-regulatory landscape within one of the most complex and disease relevant regions in the human genome, characterizes shared and cell-type specific regulatory mechanisms, and reveals cell-type specific relevance to disease-associated genetics.

Session Title: Omics Technologies Poster Session III

PB3366 Distinct and overlapping metabolomic profiles across cardiometabolic diseases reveal converging points of disease etiology.

Authors:

J. Dron[1,2], J. Lee[1], A. Schuermans[1], S. Koyama[3], W. Hornsby[1], G. Peloso[4], P. Natarajan[1,2,5]; [1]Broad Inst. of MIT and Harvard, Cambridge, MA, [2]Ctr. for Genomic Med., Massachusetts Gen. Hosp., Boston, MA, [3]Broad Inst. of MIT and Harvard, Boston, MA, [4]Boston Univ., Boston, MA, [5]Cardiovascular Res. Ctr., Dept. of Med., Massachusetts Gen. Hosp., Boston, MA

Cardiometabolic diseases often co-occur and share common risk factors. To reveal converging and distinct points of disease etiology, we utilized Mendelian randomization (MR) to identify metabolites with distinct and overlapping genetic associations between coronary heart disease (CHD), type 2 diabetes (T2D), chronic kidney disease (CKD), non-alcoholic steatohepatitis (NASH), non-alcoholic fatty liver disease (NAFLD), and hypertension (HTN). Genetic instruments (IVs) for 168 metabolites were selected from GWAS summary statistics of Nightingale NMR metabolomics in ~115K European individuals from the UK Biobank. IVs were selected to be significantly associated with the metabolite of interest (P<5e-8) and clumped into independent loci <10Mb (R^2 <0.001; based on the 1KG European reference panel). GWAS summary statistics for the outcomes of interest were obtained from non-overlapping FinnGen v9 samples, which included ~430K Finnish individuals. Two-sample MR was used to test for causal relationships between each metabolite and outcome. The inverse-variance weighted method was used for the primary analysis, and metabolites with directionally consistent effects across alternative MR methods (MR Egger, weighted median, weighted mode) were retained. MR revealed relationships between 139 (82.7%) metabolites with at least 1 of the 6 outcomes. CHD had a distinct risk profile, with 74.0% (60/81) non-overlapping metabolites mainly defined by low-density lipoprotein (LDL) and very-low-density lipoprotein (VLDL) metrics. T2D shared 54.5% (12/22) metabolites with CHD and was defined almost exclusively by high-density lipoprotein (HDL) metrics. CKD shared 66.7% (6/9) metabolites with CHD, including 4 intermediate-density lipoprotein (IDL) metrics with effects in opposite directions for the two outcomes. NASH also had a distinct risk profile with 76.9% (10/13) non-overlapping metabolites and was primarily defined by larger VLDL particles and triglyceride content. Neither NAFLD nor HTN had exclusive lipid or lipoprotein associations. Simultaneous assessment of multiple diseases and metabolomics data enables the identification of key components of lipoprotein metabolic pathways relevant across diseases.

Session Title: Omics Technologies Poster Session I

PB3367 Diversion in endometrium adhesion pathway in Indian female with recurrent implantation failure (RIF)

Authors:

V. Chettiar[1], A. Patel[2], S. Chettiar[2], D. Jhala[3]; [1]Dept. of Life Sci., Sch. of Sci., Gujarat Univ., Ahmedabad, India, [2]GeneXplore Diagnostics and Res. Ctr. Pvt. Ltd., Ahmedabad, India, [3]Dept. of Zoology, Sch. of Sci., Gujarat Univ., Ahmedabad, India

Understanding implantation of embryo to endometrium is one of the most finely regulated processes required for successful pregnancy. Implantation in the uterus occurs in a limited time frame called the "window of implantation" (WOI) . Knowledge of predicting an accurate WOI in recurrent implantation failure (RIF) patients is still at infancy. There are many methods to predict WOI, recent method of choice being molecular gene expression evaluation of endometrial biopsy. To understand the gene expression profile in Indian females with RIF endometrial samples (n=34) and positive control (n=33) endometrial samples of females with successful IVF implantation in first two attempts, were collected during mid-secretory phase of hormone replacement therapy cycle, P+5 from the females between the age group of 28 to 35 years. Case history of each patient were studied in detail, samples were subjected to RNA isolation and were taken forward for microarray based expression profiling. Results were evaluated using GeneSpring software (Agilent, USA) for adhesion pathway analysis followed by validation experiments. Total of 133 genes were analyzed, of adhesion pathway for their direct and indirect influence on endometrial positioning for implantation. Results revealed major number of differentially regulated genes belonging to different group of cadherin, claudins, Major histocompatibility complex, integrins and chemokine (C-C motif) in females with RIF. Genes involved showed significant role and evaluating their expression profile it can be useful as candidates for biomarker to predict WOI. This study can pave the way for identification of group of genes in Indian female using cost effective techniques like RT-PCR that will make the testing protocol affordable to larger fraction of population.

Session Title: Omics Technologies Poster Session II

PB3368 Efficient computational pipeline for processing and analysis of single-nucleus transcriptome sequencing data

Authors:

P. NM[1], J. Bendl[1], D. Lee[2], G. Hoffman[2], J. Fullard[2], P. Roussos[2]; [1]Icahn Sch. of Med. at Mount Sinai, New York City, NY, [2]Icahn Sch. of Med. at Mount Sinai, New York, NY

Revolutionising our understanding of gene expression at the single-cell level, single-nucleus transcriptome sequencing (snRNA-seq) enables researchers to analyse gene expression profiles from individual nuclei, providing valuable insights into cellular heterogeneity and complex biological processes. As snRNA-seq continues to advance, standardised analysis pipelines are increasingly important to effectively process and interpret the expanding volume of data generated.To address the challenges associated with diverse experimental and computational setup of snRNA-seq experiments, including the requirement for processing multiplexed design and configuration of high-performance computing (HPC) resources, we have developed a comprehensive computational pipeline. Built on Snakemake framework, our pipeline offers an easy setup, scalability, reproducibility for processing snRNA-seq data of any size, from one-off experiments to large-scale sequencing projects. The main pipeline steps include read alignment, quality control and generation of gene expression matrix. The pipeline outputs can be seamlessly used for subsequent analyses such as differential expression, pseudotime trajectories, and various quantitative trait loci (QTLs). The pipeline code and documentation are maintained through version control systems Github and Read the Docs, respectively.

The pipeline output includes convenient quality control plots, summary statistics, and a variety of generally useful data formats to set the groundwork for subsequent project-specific data analysis. Applying our pipeline to over 2,000 samples linked to neuropsychiatric and neurodegenerative disorders, including Alzheimer's disease, Parkinson's disease, schizophrenia, and bipolar disorder, we successfully handled the challenges of large-scale datasets.

In conclusion, our comprehensive pipeline for snRNA-seq enables efficient processing and interpretation of data. The successful application of our pipeline to thousands of samples associated with various neuropsychiatric and neurodegenerative disorders highlights its effectiveness in unravelling the molecular underpinnings of complex diseases.

Session Title: Omics Technologies Poster Session III

PB3369 Efficient quantification of pattern similarity between spatial genomics and cell morphology

Authors:

Y. Yang, J. Hu; Emory, Decatur, GA

Spatial transcriptomics (ST) is a powerful tool for studying gene expression patterns in spatial context, complemented with tissue histology image. Despite the availability of various methods to integrate the gene expression and tissue histology imaging in ST, the lack of a gold-standard metric to accurately quantify the similarity between features across different modalities has been a limitation for data integration. There is a strong demanding to measure the pattern similarity across features from different modalities, which provides insights into the association between molecular profile and cell morphology and enhances our understanding of tissue organization and function. To address this, we introduce VOGC, a spatial pattern similarity measurement using variance of orthogonal gradient curve that effectively captures the pattern similarity of features from different modality in the spatial context. Our method starts from identifying the common regions indicated by different modalities. Next, within each domain, it characterizes the pattern of each features using gradient curves in two orthogonal directions. Then, we calculate the variance of the different curves for each direction in every domain, summing them up to quantify the dissimilarity between two comparing features. The maximum variance observed across all regions serves as an indicator of the highest level of feature disparity. VOGC has been evaluated on multiple datasets, and compared with other popular spatial pattern measurement, i.e., SSIM, spatial correlation and Moran's I. The results demonstrate that our method can better identify image features that exhibit concordance with gene expression as well as those that show dissimilarity than all existing metrics. VOGC can be further used to filter out image features dissimilarity to all gene expression patterns, which captured artificial effect such as blurriness resulting from loss of camera focus. VOGC serves as a bridge between gene expression patterns and tissue histology imaging in ST, providing a robust metric to assess the similarity of features across diverse modalities.

Session Title: Omics Technologies Poster Session I

PB3370 Empowering immunogenetic analysis with Biofilter 3.0 via enhanced annotation and filtering capabilities.

Authors:

V. Truong, X. Li, S. Dudek, J. Wherry, M. Ritchie; Univ. of Pennsylvania, Philadelphia, PA

Autoimmune diseases are a growing global health concern, affecting approximately 4.5% of people worldwide. In addition, 25% of affected individuals suffer from multiple autoimmune disorders. While hundreds of genetic variants have been identified for the most common autoimmune diseases, the small effect sizes and limited explanatory power underscore the need for an integrative knowledge-driven approach. The body of knowledge for genetic and immunological information is vast and oftentimes overwhelming. We discuss how incorporating immune-relevant information can empower the discovery and interpretation of combinations of genetic variations based on prior immunological knowledge. In this work, we present Biofilter 3.0 which contains enhanced annotation and filtering capabilities for empowering immunogenetic analyses. Previous versions of Biofilter provided a convenient single interface for accessing multiple publicly available human biological databases stored within the supporting knowledgebase of the Library of Knowledge Integration (LOKI). We integrated additional public databases of autoimmune and immune-mediated information into LOKI to bolster existing information on genomic locations of SNPs, genes, ontological categories, and interaction pairs. Furthermore, we expanded the functionality of Biofilter to make it possible to leverage this prior knowledge in analyses. This approach yields a collection of functions which enable researchers to efficiently annotate, subset, and filter thousands of SNPs, genes, proteins, and genomic locations based on immunological criteria, including relevant immune terms, pathways, and/or diseases.

Session Title: Omics Technologies Poster Session II

PB3371 Enhanced promoter occupancy of SUMO1 bound proteins in AC16 cells during hypoxia

Authors:

D. Sharma, C. Rai, M. Kumari, A. Kumari, R. Varshney, R. C. Meena; Defence Inst. of Physiology and Allied Sci., Delhi, India

Hypoxia and its intricate regulation is the epicenter of cardiovascular diseases. It triggers reactivation of the fatal gene programme, induces cardiac hypertrophic response, changes the composition of extracellular matrix, affects mitochondrial biogenesis and has an impact on myocardial contractility. Cellular energy demand and supply under hypoxic conditions is also regulated by post-translational modification of various proteins, which complicates studies on individual pathways. SUMOylation is one such transient post-translational modification where small-ubiquitin like modifiers (SUMO) get covalently attached to their target-proteins by a multi-step enzymatic mechanism. In addition to regulating diverse cellular processes and deciding the fate of proteins, SUMOylation is also known to maintain cardiac function and can protect against a hypertrophic response to cardiac pressure overload. However, the detailed transcriptional regulation of cardiac cell responses by SUMOylation during hypoxia, has not yet been fully determined. Here, we uncover a genome-wide relationship between global SUMOylation and chromatin-associated SUMOylation in AC16 cells exposed to hypoxic stress (1% O_2) for 24 h. To address this, we performed immunoassays and chromatin immunoprecipitation sequencing (ChIP seq) with SUMO1 antibody. Chromatin immunoprecipitation and its sequencing was done by our previously optimized protocol. Further to analyse and visualise the immunoprecipitated regions, MACS2 software was used for peak calling, followed by Cis-regulatory element annotation system (CEAS) tool for promoter analysis of regulated genes and gene region annotation. Also, the narrow peak files were further visualised in Integrative Genome Viewer. With the analysis, we found that global SUMOylation of proteins by SUMO1 was significantly decreased while chromatin-associated SUMO1 signals increased at promoter-transcriptional start sites in response to hypoxic stress. The number of peaks detected by MACS2 analysis were 10 fold more in hypoxia exposed AC16 cells, as compared with unexposed control. These peaks are associated with active chromatin regions including promoter, introns, exons and UTR's. The promoter occupancy of SUMO1 associated peaks was higher in hypoxic condition, which represents the enhanced SUMO1 binding to transcription factors. These observations provide evidence that SUMO1 might play a pivotal role in controlling the transcription in AC16 cells during hypoxic stress. These findings can be further analysed and validated to precisely identify SUMO1 bound targets for modulating hypoxic response in cardiovascular diseases.

Session Title: Omics Technologies Poster Session III

PB3372 Enhancing Scalability and Consistency in Translational Multiomics with an Optimized Fixed-Cell ATAC-seq Method.

Authors:

R. Graham, M. Johnson, C. Mao, E. Louie, A. O'Hara, J. Kim, B. Niu, B. Sereni, D. Corney, C. Mozdzierz, H. Latif, G. Zhou; Azenta Life Sci., Burlington, MA

ATAC-seq has an emerging role in decoding mechanisms of gene regulation, offering valuable insights into pathology and treatment response in disease models for cancers, autoimmune conditions, and neurocognitive disorders. However, clinical adoption of conventional ATAC-seq methods has been limited by logistical hurdles, including time-sensitive processing of fresh samples and compromised viability of cryopreserved cells. These constraints, compounded by changes in open chromatin regions (OCRs) following cryopreservation, introduce unintended bias and pose obstacles for the translational impact of ATAC-seq experiments. Here, we introduce an optimized fixed-cell ATAC-seq approach to overcome these limitations and unlock new sample types for ATAC-seq analyses. Our solution improves and simplifies the workflow from sample collection to clinical deliverable. This allows for the investigation of a more diverse range of sample types and facilitates complex experimental designs including time course studies, drug efficacy screens, and cell differentiation analyses. To demonstrate the effectiveness of this method, we compared our optimized fixed-cell method with fresh and cryopreserved samples of GM12878 cells prepared in parallel. Remarkably, we observed consistent patterns of OCR enrichment at key regulatory elements across the three sample preparation methods. We observed consistent OCR enrichment across the promoter region of known highly-expressed B-cell genes including *CD48* and *LCP1*, underscoring this assay's potential for detection of key gene activity in disease models. Our optimized method also enhances compatibility with RNA-Seq data collection and integration, enabling the simultaneous study of chromatin and transcriptional activity from a single sample. In this study, RNA-seq pathway analyses revealed that the top 6 significant pathways are related to B-cell function and known B-cell diseases, demonstrating the opportunity to generate multiomic analyses with this fixed-cell approach. This novel ATAC-seq approach offers enhanced scalability and consistency over conventional ATAC-seq, and presents new opportunities for seamless and efficient multiomic data collection in clinical and research settings.

Session Title: Omics Technologies Poster Session I

PB3373 Estimating heritability of gene expression from single cell measurements by modeling intra-subject variation with linear mixed models

Authors:

Z. Xu, T. Amariuta; Univ. of California, San Diego, San Diego, CA

Heritability estimation has been extensively applied to understanding the genetic component of gene expression in bulk RNA-sequencing data. However, the role that genetic variation plays in regulating gene expression depends heavily on the cell type. Single cell RNA-sequencing techniques enable studies of cell-type-specific genetic regulation of gene expression at unprecedented resolution. Recent heritability estimation using single cell RNA-sequencing relies on aggregating reads across cells of the same individual (pseudo-bulking), leaving the potential value of modeling variation between single cell measurements completely unexplored. In this work, we aim to improve the power and accuracy of gene expression heritability estimation by leveraging per-cell measurements. To this end, we implement a new linear mixed model to explicitly account for the shared environment of cells within an individual. Our model partitions the variance of gene expression into intersubject (genetic) variance, intrasubject (cell-cell) variance, and random noise. We also propose a new restricted maximum likelihood estimation (REML) algorithm using a Poisson distribution to address the common model misspecification issue of applying Gaussian distributions to single cell RNA-sequencing data. Our model is designed to be computationally robust and efficient in finding the best estimation of variance terms via a combination of expectation-maximization (EM) and average information (AI) algorithms. In simulations, we generate heritable gene expression phenotypes across 100 individuals and a variable number of cells per individual to mimic single cell RNA-sequencing data. Cell-cell relatedness is approximated by the 2D UMAP projection of the gene expression data. We found that our model produces unbiased estimates of heritability and that modeling multiple measurements per individual greatly reduced noise in estimating small heritability (i.e., <10%, which is realistic for the heritability of gene expression attributed to cis variants). We observed a 98.9% reduction in heritability standard error when simulating 100 cells per individual compared to conventional pseudo-bulk analysis using GCTA. In real data analysis, we assess the validity of our method applied to the 1K1K single cell RNA-sequencing cohort of 982 individuals and 1.27 million immune cells. We anticipate that our approach will more powerfully quantify the genetic component of gene expression than pseudo-bulk analysis. This work is important for understanding cell-type-specific genetic regulatory mechanisms and co-regulation across fine-grained cellular populations.

Session Title: Omics Technologies Poster Session II

PB3374 Evaluating performance and applications of sample-wise cell deconvolution methods on human brain transcriptomic data

Authors:

R. Dai[1], T. Chu[2], M. Zhang[2], X. Wang[2], A. Jourdon[3], F. Wu[3], J. Mariani[3], F. Vaccarino[3], D. Lee[4], J. Fullard[4], G. Hoffman[4], P. Roussos[4], Y. Wang[5], X. Wang[6], D. Pinto[4], S. Wang[7], C. Zhang[8], C. Chen[2], C. Liu[1]; [1]SUNY Upstate Med. Univ., Syracuse, NY, [2]Central South Univ., Changsha, China, [3]Yale Univ., New Haven, CT, [4]Icahn Sch. of Med. at Mount Sinai, New York, NY, [5]Virginia Polytechnic Inst. and State Univ., Blacksburg, VA, [6]UTHSC, Memphis, TN, [7]Univ. of Texas Hlth.Sci. Ctr. at Houston, Houston, TX, [8]Upstate Med. Univ., Syracuse, NY

Sample-wise deconvolution methods have been developed to estimate cell-type proportions and gene expressions in bulk-tissue samples. However, the performance of these methods and their biological applications has not been evaluated, particularly on human brain transcriptomic data. Here, nine deconvolution methods were evaluated with sample-matched data from bulk-tissue RNAseq, single-cell/nuclei (sc/sn) RNAseq, and immunohistochemistry. A total of 1,130,767 nuclei/cells from 149 adult postmortem brains and 72 organoid samples were used. The results showed the best performance of dtangle for estimating cell proportions and bMIND for estimating sample-wise cell-type gene expression. For eight brain cell types, 25,273 cell-type eQTLs were identified with deconvoluted expressions (decon-eQTLs). The results showed that decon-eQTLs explained more schizophrenia GWAS heritability than bulk-tissue or single-cell eQTLs alone. Differential gene expression associated with multiple phenotypes were also examined using the deconvoluted data. Our findings, which were replicated in bulk-tissue RNAseq and sc/snRNAseq data, provided new insights into the biological applications of deconvoluted data.

Session Title: Omics Technologies Poster Session III

PB3375 Evaluating the genome-wide off-target edits mediated by CRISPR/CAS9 in human iPSCs

Authors:

D. Liu[1], S. Marlow[1], G. Nagappan[1], C-F. Xu[2], A. Cortes[2]; [1]GSK, Cambridge, MA, [2]GSK, Stevenage, United Kingdom

Background: Since its inception, CRISPR-Cas systems have been the core toolkit for targeted genome editing in research. One potential concern of CRISPR-Cas-based therapy is that the guide RNA may induce double-strand cleavages at off-target sites and thus generate unintended genetic alterations. Evaluating the scope and functional impact of such unintended edits is hence crucial for understanding phenotypic changes. In this study, we aim to provide a thorough view of the off-target events in targeted CRISPR-Cas experiments as well as the suggested workstream using whole genome sequencing (WGS). **Methods and results:** The human biological samples were sourced ethically, and their research use was in accord with the terms of the informed consents under an IRB/EC approved protocol. Human induced pluripotent stem cells (iPSCs) were subjected to CRISPR-Cas9 editing using pooled small guide RNAs targeted to two exons, with the intent of creating knockout of two target genes of interest either individually (single KO) or together (double KO). We performed short-read whole genome sequencing (WGS) of the genomic DNA from our derived iPSC clones.We assessed the suitability of three genomic analysis tools and selected GATK Mutect2 somatic variant caller as the preferable analysis pipeline for genome-wide editing evaluation. Mutect2 compared the pre- and post-editing WGS data on 30 iPSC clones (including 5 unedited wild type, 16 single KOs, and 9 double KOs) in genomic regions with in-silico (Cas-OFFinder) prediction of the guide RNA (gRNA) binding. The number of SNPs and Indels identified per clone ranged from 1 to 96 and 0 to 14, respectively. The majority of off-target changes overlap non-coding regions while a small number were missense or frameshift variants. We also identified genes that were frequently affected at a rate on par with the genes that gRNAs were designed for. We are following up on this data by undertaking a series of enrichment analyses and will report on the genomic features and biological networks disproportionally represented among the unexpectedly disrupted regions. **Conclusion:** This study provided benchmarking data and novel insights into the scope and impact of off-target events in CRISPR-Cas systems, which will be highly valuable in understanding phenotypic changes upon editing. By demonstrating the utility of GATK Mutect2 in the context of detecting unwanted editing events, we also expanded the use case of this widely used somatic calling pipeline.

Session Title: Omics Technologies Poster Session I

PB3376 Evaluating the mouse neural precursor line, SN4741, as a suitable proxy for midbrain dopaminergic neurons: A lesson in obtaining genomic rationale for the use of in vitro models of human biology and disease.

Authors:

R. Boyd, S. McClymont, N. Barrientos, P. Hook, W. Law, R. Rose, E. Waite, J. Rathinavelu, D. Avramopoulos, A. McCallion; Johns Hopkins Univ. Sch. of Med., Baltimore, MD

Since the emergence of next generation sequencing, a surge in technological advances has enabled various innovative approaches by which to study human genetic disease and evaluate novel therapeutic strategies. As multiomics data become more commonplace in human genetics research, it seems increasingly imperative to apply these technologies to validate and annotate model systems of human disease that accurately represent the biological context in which disease associated variants mediate their effects. To overcome ethical and technical limitations of in vivo models, model organism-derived cell lines are often employed to investigate disease mechanisms and therapeutic avenues. However, despite their ubiquitous use, many in vitro models lack contemporary genomic analyses supporting their suitability as surrogates of human disease.

One such proxy is the mouse neural precursor cell line, SN4741, which has been used to study genetic variation, cell signaling, transcriptional regulation, neurotoxicity, and therapeutic intervention in Parkinson Disease/PD for >25 years. PD is a neurodegenerative disease characterized by the progressive onset of motor and cognitive impairments due to accumulation of α-synuclein aggregates and loss of midbrain dopaminergic/DA neurons. PD disrupts the quality of life of >8.5 million people worldwide; therefore, it is crucial to determine the extent to which biological surrogates of this disease reflect the processes they are assumed to model.

Using a combination of classic and contemporary genomic techniques - karyotyping, RT-qPCR, scRNA-seq, bulk RNA-seq, and ATAC-seq - we characterized the transcriptional landscape, chromatin landscape, and genomic architecture of the SN4741 cell line to evaluate its suitability as a proxy for midbrain DA neurons in PD research. We found that SN4741 cells are polyploid and exhibit low expression of DA neuron markers. Transcriptional signatures suggest that these cells remain undifferentiated at a permissive temperature (37°C) and differentiate into immature neurons at a non-permissive temperature (39°C) - not DA neuron precursors, as previously suggested. Furthermore, the chromatin landscapes of differentiated and undifferentiated SN4741 cells lack concordance with open chromatin profiles of ex vivo, mouse E15.5 midbrain-derived DA neurons. Our data suggests that SN4741 cells may reflect early aspects of neuronal differentiation but are a less suitable proxy for DA neurons than previously thought. The implications of this study extend broadly, exemplifying the need for robust biological and genomic rationale underpinning the use of in vitro models of human disease.

Session Title: Omics Technologies Poster Session II

PB3377 Exploring mass spec proteomics data using REVEAL: an easy-to-use bioinformatics platform for comprehensive analysis

Authors:

K. Sharma, A. G. van der Velde, J. Zahiri, Z. Pitluk, S. Sarangi; Paradigm4, Inc., Waltham, MA

The crux of precision medicine is to use molecular tools to define biomarkers for diagnosis and disease targets for intervention, sometimes on individual scales. The basic informatics approach is to look for differentially expressed proteins (DEP) between normal and diseased samples. Mass spec proteomics has evolved to the point where recent advances have gained more than 10-fold sensitivity over traditional methods by leveraging numerous techniques like TIMs to enrich less abundant proteins. The fundamental roadblock to further progress is managing and manipulating the large datasets. Key challenges include - cross sample comparisons that are essential for DEP analysis, intrasample queries that easily capture the many to many relationships of protein group to protein to peptide, and research reproducibility to ensure tracking of data processing steps. Considering these issues, this study aims to evaluate alternative statistical methods and tools for detecting DEPs, while analyzing the advantages and disadvantages of employing different algorithms. We utilized REVEAL, a bioinformatics platform that enables easy loading and efficient storage of large multiomics datasets including mass spec proteomics, while providing APIs (REST, R, Python) to perform cross-dataset queries such as searching for protein and precursor signatures or mapping features between ontologies, among others. The primary dataset that we evaluated was a publicly available pan-cancer cell line proteomic map (accession number: PXD030304), which quantified 8,498 proteins across 949 cell lines. In our workflow, the raw data format was converted using ProteoWizard followed by the generation of the intensity matrix using DIA-NN. Subsequently, the batch correction was performed using the Combat package. Finally, various statistical methods implemented in packages: limma, edgeR and deseq2 as well as Wilcoxon rank-sum test which were applied to the batch-corrected intensity. This study demonstrates how REVEAL serves as a solution for managing and analyzing large proteomics datasets, showcasing cross-dataset use cases and comparative algorithm analysis. With data exceeding 1TB, REVEAL effortlessly handles data loading and management in this case study. Additionally, we harnessed REVEAL: Features, utilizing its gene-protein mapping and accession number conversion capabilities for gene ontology (GO) and pathway enrichment analysis.

Session Title: Omics Technologies Poster Session III

PB3378 From Gene Regulation to Metabolites: A Multi-Omics Framework for Investigating Airway Hyperresponsiveness in Pediatric Asthma

Authors:

R. Sharma[1], S. Chu[1], R. Kelly[1], S. Begum[1], K. Mendez[1], J. Celedon[2], C. Clish[3], S. Weiss[1], J. Lasky-Su[1], M. McGeachie[1]; [1]Brigham and Women's Hosp. and Harvard Med. Sch., Boston, MA, [2]Children's Hosp of Pittsburgh of UPMC, Univ. of Pittsburgh, Pittsburgh, PA, [3]Broad Inst. of Massachusetts Inst. of Technology and Harvard, Cambridge, MA

Introduction Airway hyper-responsiveness (AHR) is a prominent feature of asthma, with complex molecular mechanisms involving multiple pathways. This study explores how changes in miRNAs and metabolites during childhood asthma may lead to changes in AHR. **Methods** In the Childhood Asthma Management Program (CAMP), we analyzed 564 children with asthma using longitudinal AHR (PC20) measures and plasma metabolomic profiling, among whom 491 children also provided serum for miRNA sequencing at baseline. Our aim was to uncover time-dependent interactions between metabolites or miRNAs and AHR. Linear Mixed models were used to assess the interactions between metabolite/miRNA and AHR adjusting for age, sex, race, treatment group and recruitment clinic, where interaction $p < 0.05$ was considered significant. Joint pathway enrichment analysis of miRNA and metabolites with significant interactions was performed with MetaboAnalyst V5. **Results** The mean ages at the three studied time-points across 16 years of follow-up in CAMP were 8.8, 12.8, and 17.4 years; the mean PC20 were 2.1 mg/ml, 7.7mg/ml and 11.6 mg/ml, respectively, indicating less AHR as children grew. Decreasing PC20 values was found to be associated with lower levels of lung function (prebronchodilator percent predicted FEV1, $p < 0.001$), asthma symptoms, and a subjective clinical staff assessment of asthma severity ($p < 0.001$). We found that 72 (14.4%) of the 501 identified metabolites and 72 (27%) of the 266 identified miRNAs exhibited a significant interaction with time on AHR. The joint pathway enrichment analysis of significantly interacting metabolites and miRNA target genes revealed the enrichment of several biochemical pathways that have been implicated in AHR development through impact on allergic airway inflammation, mucus hypersecretion, smooth muscle contraction, and oxidative stress. **Conclusion** Our findings provide evidence to support broad changes in metabolites and miRNA regulation of genes accompanied general clinical trends of asthma improvement and decreasing AHR in children with mild to moderate severe asthma. These omic indicators were enriched in pathways associated with key molecular mechanisms involved in AHR. Longitudinal multiomic analysis is likely to be informative in additional populations and conditions.

Session Title: Omics Technologies Poster Session I

PB3379 GenCompass: A cloud-enabled, accelerated, end-to-end workflow for analysis of large whole-genome sequencing (WGS) cohorts.

Authors:

L. E. Egolf[1,2], B. T. Jordan[1,2], B. D. Hicks[1,2], K. Jain[1,2]; [1]Div. of Cancer Epidemiology and Genetics, NCI, NIH, Bethesda, MD, [2]Cancer Genomics Res. Lab., Frederick Natl. Lab. for Cancer Res., Frederick, MD

Ultra-high throughput next generation sequencing (NGS) platforms have dramatically lowered the time and cost of genome sequencing, surpassing Moore's law. This has made population genetics studies feasible, driving a better understanding of the contributions of germline genetics across multiple disease areas. However, computational tools have not advanced as rapidly, and there is a need for accelerated pipelines to analyze data from large cohort studies. Bioinformaticians and data analysts frequently employ many small pipelines to analyze data, encountering difficulties in installation, lack of reproducibility, long runtimes, and unoptimized resource requirements. To address this challenge, we have developed GenCompass (Germline ENsemble Calling Of Mutations with Parabricks Accelerated Software Suite): a modular, resource-optimized, end-to-end comprehensive workflow consisting of NGS data preprocessing, Nvidia Parabricks GPU-accelerated mapping and ensemble variant calling using three popular germline callers, comprehensive quality control and reporting at each step, and variant annotation. This optimization yields a runtime of 3-4 hours and processing cost of $9 per 40x sample on cloud platforms using spot instances. Our workflow is system agnostic and uses standardized Docker images, GitHub, Workflow Development Language, and Cromwell for reproducibility. Here we present results from rigorous benchmarking conducted using guidelines from Genome in a Bottle (GIAB) and National Institute of Standards and Technology (NIST). We tested a total of 9 WGS samples from GIAB and synthetic diploid standard benchmarking datasets, downsampled to 40x coverage to represent a typical cohort study. GenCompass achieved a median F1 score of 98.4% for SNPs (precision 99.6%; recall 97.2%) and a median F1 score of 97.7% for INDELs <50 bases (precision 98.8%; recall 96.5%) genome-wide. Performance remained strong in difficult-to-call regions defined by GIAB for both SNPs (median F1 94.8%; precision 98.0%; recall 91.8%) and INDELs (median F1 97.1%; precision 98.5%; recall 95.8%). In contrast to other commercial and open-source pipelines that are either limited in scope, rely on a single variant caller, or lack thorough benchmarking for population studies, GenCompass offers complete functionality, performs ensemble calling, and is benchmarked with multiple whole genomes using latest community standards. Our open-source pipeline does not depend on any commercial license for software or proprietary hardware. Finally, GenCompass has been tested on WGS datasets up to 80X and can be extended to primary and secondary analysis of non-germline NGS datasets.

Session Title: Omics Technologies Poster Session II

PB3380 Genetic influences of gene co-expression using individual-specific networks without single-cell omics

Authors:

F. Melograna[1], B. Yousefi[1,2,3], Z. Li[1], E. Gervasoni[4], B. Schwikowski[2], I. Konigsberg[5], K. Kechris[6], P. Castaldi[7], M. Cho[8], K. Van Steen[1,4]; [1]KU Leuven, Leuven, Belgium, [2]Inst. Pasteur, Le Perreux-sur-marne, France, [3]Sorbonne Université, Paris, France, [4]Univ. de Liege, Liege, Belgium, [5]Univ. of Colorado, Denver, CO, [6]Colorado Sch. of Publ. Hlth., Aurora, CO, [7]Brigham and Women s Hosp., Cambridge, MA, [8]Brigham and Women s Hosp., Duxbury, MA

Reverse-engineered *individual-specific networks* (ISNs) (Kuijjer et al. 2019) from an aggregate network offer the opportunity to investigate the impact of individual-level network wirings, paths or connectivity on decision-making in the individual's interest. This work investigates the potential of Kuijjer's ISNs to identify *co-expression QTL epistasis* (co2-eqtls). To this end, we identified those zones in a collection of ISNs that exhibited the most significant heterogeneity in network wiring. Our approach uses principles of individual similarity assessment (Euclidian distances between spectral vectors), computation of a measure of ISN wiring heterogeneity (Fiedler values of symmetrically normalized Laplacians), and significance assessment of the subsystems (stepdown *maxT* algorithm). For each significant zone (a "HotZone") in this sense, we analyzed the weighted edges for their relationship to disease phenotypes. We subsequently prioritized the most promising disease-associated edges for co2-eqtl analysis with *MB-MDR* (Van Lishout et al. 2015), a data mining tool for large-scale epistasis detection. To construct individual-specific gene co-expression networks, we used *RNA-seq* data and 4,524 genes with a coefficient of variation of at least 1, adjusted for age, sex and clinical center, available for 2,655 Non-Hispanic White (NHW) and 1,024 African American (AA) smokers, part of COPDgene (www.copdgene.org). We observed 15 (16) significant HotZones in AA (NHW) of size 3-37 (3-33). In AA (NHW) 75% (64%) of Zones that significantly linked to FEV1 via their eigengene representations were Hot, indicating non-negligible stratification of individuals according to their ISNs. Standard and ISN edge analyses were found to be complementary. For NHW HotZones, 3% of edges were significantly associated to *FEV1* and covered 4 Zones. These Zones were enriched for *Reactome* pathways related to the immune system and transport of small molecules. The top FEV1 associated HotZone edges in AA and NHW involved the gene pairs (*OLIG2*-Novel transcript-antisense to *AJAP1*) and (*ABCG2* - *TSPO2*). We detected 40 (AA) and 201 (NHW) co2-eqtls in autosomal chromosomes (adjusted for main effects). Gene interaction networks exhibited varying network properties including hubs *PTPRF* (AA) and *GMDS-DT*, *PRKN*, *CNKSR3*, *LINC02432*, *RBM47* (NHW). Several examples highlighted varying relationships between gene co-expression and polygenic risk score by multilocus genotype profile. In summary, our results showcase that ISNs, even without single-cell omics, can increase our understanding of (genetic modifiers of) co-eQTLs as direct or indirect regulators of gene co-expression.

Session Title: Omics Technologies Poster Session III

PB3381 Genetic prediction performance of molecular traits is driven by their genetic architecture.

Authors:

F. Nyasimi, L. Yanyu, H. Im; Univ. of Chicago, Chicago, IL

Genetic predictors of molecular traits are powerful instruments to probe the biology of complex diseases. Gene expression traits represented as mRNA levels, splicing, and protein levels have been successfully predicted and linked to disease etiology using Transcriptome-wide Association Study (TWAS)-like approaches. However, these only represent a small fraction of the molecular scaffold connecting genotype to phenotype. Expanding genetic prediction downstream is essential to advance our understanding of the mechanisms that drive complex disease etiology.

For gene expression traits, given their sparse genetic architecture in cis, elastic net predictors have been shown to perform well. However, optimal prediction approaches for metabolite levels and other downstream traits remain less understood.

Here we propose a user-friendly software, Arch-Predict, to investigate the genetic architecture of these mediating traits and select the optimal prediction approaches. We use the performance of the elastic net family of prediction models that span the spectrum of sparse to polygenic (infinitesimal model) architecture to identify models that are more likely to perform well. Given a validation set, we provide a snakemake pipeline that will compare a suite of prediction approaches including a range of polygenicity levels in the elastic net family as well as several summary statistics-based approaches, including PRS-CS, lassosum, PRSice, DBSLMM, and LDpred.Application to metabolite levels and brain MRI-derived features showcased the utility of our pipeline. We found that for metabolites, sparse models outperform polygenic ones whereas for brain features, fully polygenic models provide better performance. Our pipeline will be of great value for researchers looking to develop their own prediction models.

Session Title: Omics Technologies Poster Session I

PB3382 Genome In A Bottle Mosaic Benchmark for HG002: Initiating MDIC's Somatic Reference Sample Program.

Authors:

C. Daniels[1], A. Abdulkadir[1], M. Cleveland[2], N. D. Olson[2], M. de Mars[1], J. Zook[2]; [1]Med. Device Innovation Consortium (MDIC), Arlington, VA, [2]Natl. Inst. of Standards and Technology (NIST), Gaithersburg, MD

A landscape analysis conducted by The Medical Device Innovation Consortium (MDIC) identified a lack of well characterized somatic variant reference samples and benchmarks for use in development of cancer biomarker diagnostics. The Somatic Reference Samples (SRS) Initiative was launched by MDIC to address this need by engineering clinically relevant somatic variants into the extensively characterized Genome in a Bottle (GIAB) sample HG002. To establish a baseline for the engineered cell lines, we first identified low frequency mosaic variants in HG002 by leveraging a novel trio-based approach with unedited GIAB genomes (HG002, HG003, and HG004) and the somatic variant caller, Strelka2. To determine that the Strelka2 variant caller was fit for purpose, we used *in-silico* mixtures of real data from HG002 and HG003 to establish the allele frequency limit of detection. Then, high coverage (300X) Illumina data from the Ashkenazi Jewish trio was used for low frequency variant detection in the son. Data from HG002 (son) was passed to Strelka2 as the tumor data and the combined parents' data (HG003-father and HG004-mother) was passed as normal. These HG002 variants identified with Strelka2 were subsequently compared against the Genome In A Bottle v.4.2.1 reference set, and complex variants causing artifacts were removed to create a draft mosaic benchmark. Candidate variants were evaluated by manual curation using the high-coverage (300X Illumina) trio along with four orthogonal datasets: PacBio HiFi, Element, BGI, as well as Illumina duplex sequencing data. We will next conduct an external validation by subject matter experts in mosaic and somatic variant calling using high quality mosaic variant callsets to confirm the draft mosaic benchmark reliably identifies false positives and false negatives. We expect this new benchmark to be particularly useful as a negative control for mosaic and somatic variants. This benchmark will also inform the sequencing and informatic methods we will use to characterize off-target edits and mosaic variants in the forthcoming somatic reference samples from the SRS Initiative.

Session Title: Omics Technologies Poster Session II

PB3383 Genome Integrity assessment by optical genome mapping for cell manufacturing/bioprocessing applications

Authors:

A. Pang[1], A. Hastie[2], A. Chaubey[3]; [1]Bionano, San Diego, CA, [2]Bionano Genomics, San Diego, CA, [3]Bionano Genomics Inc., San Diego, CA

Cells and cell lines are used for multiple applications such as bioprocessing, therapy, and research. To ensure quality of cells and downstream applications, appropriate quality control (QC) methods are critical. Historically, karyotyping has been employed, but it is limited by its very low resolution and tedious laboratory workflow. Other methods such as PCR and targeted sequencing can characterize only small genomic variants at specific loci. Whole genome sequencing can detect small variants genome-wide but has limited sensitivity in detecting structural variants (SVs). Optical genome mapping (OGM) is a novel genome analysis technique that can fill many of the gaps in current capabilities for assessing genome integrity. To find clonal variants, 400 Gbp of data is collected from the parental/control and test samples. For both samples, a de novo assembly is constructed, and homozygous or heterozygous SVs are assessed. Subsequently, the dual variant annotation pipeline identifies unique SVs in the test sample compared to the parental sample. To discover subclonal SVs down to 5% variant allele fraction (VAF), the workflow leverages the generation of 1.5 Tbp of data, requiring a simple adjustment to longer data collection time. Comparison of test sample to parental sample enables easy determination of acquired variants. Finally, generation of ~5 Tbp of data, requiring 2-3 flow cells to be run, enables the detection of SVs at ultra-low VAF down to ~1%. In this study, several dilutions and simulations were performed to examine OGM's limit of detection. Targeting a coverage of 5 Tbp and analysis using the somatic SV-analysis workflow revealed that OGM has the sensitivity to detect deletions >50kbp, insertions >20kbp, duplications >100 kbp, and translocations at ~1% VAF. We have applied the clonal and somatic workflows to verify genomes' integrity after cell immortalization, induced pluripotency, transgene-integration, and gene-editing. The data using the robust and sensitive workflows demonstrate that the OGM platform is a cost-effective solution for cell manufacturing/bioprocessing QC applications.

Session Title: Omics Technologies Poster Session III

PB3384 Genome wide CRISPR screen identifies ceramide-1-phosphate transporter as a regulator of the CLEAR gene network

Authors:

T. Han[1], A. Hirt[2], A. Kim[3], L. Allende[4], M. Kono[4], R. Proia[4], C. Tifft[1]; [1]Natl. Human Genome Res. Inst., Bethesda, MD, [2]NIH, Bethesda, VA, [3]NHGRI, Bethesda, MD, [4]Natl. Inst. of Diabetes and Digestive and Kidney Diseases, Bethesda, MD

The Coordinated Lysosomal Expression and Regulation (CLEAR) gene network contributes to cellular homeostasis by responding to changes in nutritional state, primarily through the activity of transcription factors TFEB and TFE3. Increased TFEB and TFE3 activity are associated with some cancer types, while overexpression of these transcription factors has been shown to alleviate the lysosomal storage phenotype associated with multiple neurodegenerative disease models. Thus, better understanding the regulatory mechanisms underlying the CLEAR gene network will lead to a better understanding of the pathophysiology of these diseases and identify potential therapeutic strategies. We performed a genome wide CRISPR knockout screen to identify genes and pathways that regulate the CLEAR network. Among the top hits from the screen was the ceramide-1-phosphate (C1P) transporter *CPTP*. This gene is of interest because several lysosomal storage disorders are caused by disruption of sphingolipid and ceramide metabolism. We generated a *CPTP* knockdown cell line that demonstrated altered lipid profiles, including elevated C1P and decreased ceramides, hexosyl-ceramides, and sphingomyelin. RNA-seq of our knockdown cells identified 24 differentially expressed CLEAR genes, one of which was *HIF1A*. Similarly, pathway analysis identified enrichment for several gene sets related to HIF1a activity, including hypoxia response, angiogenesis, and hypertrophic cardiomyopathy. Notably, pathway analysis did not identify enrichment of lysosomal or autophagy gene sets, suggesting that the CLEAR activation in this system occurs independent of those canonical mechanisms. Our study indicates that subsets of transcriptional programs exist within the CLEAR network that enable tailored responses to specific environmental stimuli. This further suggests that it may be possible to modulate individual CLEAR gene transcriptional programs based on the therapeutic strategy of interest, e.g. increasing lysosomal biogenesis and autophagy in neurodegenerative diseases or decreasing hypoxia response in solid tumors.

Session Title: Omics Technologies Poster Session I

PB3385 Genome-wide analysis of radiation-induced somatic mutations in mouse long-term hematopoietic stem cells

Authors:

Y. Matsuda[1,2], A. Uchimura[1], Y. Satoh[1], N. Kato[1], M. Toshishige[1], K. Kume[2], J. Kajimura[1], K. Hamasaki[1], K. Yoshida[1], T. Hayashi[1], A. Noda[1], O. Tanabe[1]; [1]Radiation Effects Res. Fndn., Hiroshima, Japan, [2]Hiroshima Univ., Hiroshima, Japan

Molecular mechanisms for late adverse radiation effects such as increased neoplasm incidence after accidental exposure and secondary malignancies after radiotherapy remain unresolved even now. The major cause of the neoplasm development is thought to be somatic mutations resulting from DNA damages by radiation exposure. However, it has not been clarified at the genome-wide level what somatic mutational scars are carried in seemingly normal cells after radiation exposure, and how those scars contribute to neoplasm development. Recently, we focused on long-term hematopoietic stem cells (LT-HSCs), which have a potential to maintain hematopoiesis throughout life, and comprehensively analyzed somatic mutations introduced by whole-body X-irradiation by whole-genome sequencing (Matsuda et al. PNAS 2023). By whole-genome sequencing of clonal cell populations propagated in vitro from single isolated LT-HSCs, we comprehensively identified and characterized somatic mutations in LT-HSCs. In non-irradiated mice, the numbers of spontaneous SNVs and small indels were around 30-60 per genome, and those mutations were increased up to 2 to 3-fold in a dose-dependent manner by 3.8-7.7-Gy X-irradiation. We identified characteristic base substitutions and various mutational signatures specific to radiation exposure, which include small non-repeat deletions occurring outside tandem repeats, multisite mutations (multiple small mutations occurring within short regions), and structural variants. Further analysis of somatic mutations in multiple LT-HSCs from mice exposed to higher doses indicated that large fractions of LT-HSCs originated from single HSCs that survived the irradiation and expanded in vivo to confer marked clonality to the entire hematopoietic system. Further, we report here a detailed analysis of genomic features related to susceptibility to spontaneous and radiation-induced somatic mutations. We believe that these findings in a mouse model will provide clues for elucidating molecular mechanisms of late adverse radiation effects.

Session Title: Omics Technologies Poster Session II

PB3386 † Genomics at scale with the NHGRI AnVIL

Authors:

M. Schatz[1], A. Philippakis[2], AnVIL Team; [1]Johns Hopkins Univ., Baltimore, MD, [2]Broad Inst. of MIT and Harvard, Cambridge, MD

Modern genomics often require very large numbers of samples to detect any subtle patterns that may be present. The scale of these projects opens new opportunities for discovery; however, this scale also introduces major new technical challenges that require overhauling how genomics and genomics data science are performed. Addressing this, the NHGRI Genomic Data Science Analysis, Visualization, and Informatics Lab-Space, or AnVIL (https://anvilproject.org/), provides a secure cloud environment for the storage and analysis of large genomic and related datasets. By providing a unified environment for data management and compute, AnVIL eliminates the need for data movement, allows for active threat detection, and provides elastic, shared computing resources on demand. AnVIL currently provides harmonized access to >600,000 genomes, with many more on the horizon. We also provide access to thousands of software tools, plus several options for interactive and batch analysis.
In this presentation, we describe how AnVIL has been used in several major studies of human genomics. First, we discuss how AnVIL supports the Telomere-to-Telomere (T2T) consortium through a large-scale reanalysis of the 1000 Genomes cohort orchestrated through the Workflow Description Language (WDL) on Terra. This enabled us to identify over one million variants in the newly resolved regions of the human genome, including within the recently completed chromosome Y sequence. Next, we present new results detecting single- and multi-tissue SV-eQTLs by genotyping SVs discovered with long-reads within the GTEx short-read sequencing data. This work was quickly and securely executed through WDLs, Jupyter notebooks, and R/Bioconductor, and lead to the discovery of 5,580 SV-eQTLs where the SV has the highest CAVIAR score (a metric of causality) over other nearby SNVs. Finally, we discuss how to analyze pangenomes and haplotype diversity using Galaxy within Terra. This is critically important to diversity and disease studies, especially to capture and analyze variation not found in any single reference genome.

Session Title: Omics Technologies Poster Session III

PB3387 GenoPhenoCorr: Simplifying the Characterization of Genotype-Phenotype Correlations

Authors:

L. Rekerle, D. Danis, P. Robinson; The Jackson Lab. for Genomic Med., Farmington, CT

Understanding the intricate relationship between genotypes and phenotypes is a fundamental pursuit in genetics. Although many other areas of genetics research have been optimized through new computational technologies, genotype-phenotype correlations have remained tedious and time consuming for researchers. We present GenoPhenoCorr, a Python package that is designed to expedite the identification of genotype-phenotype correlations. GenoPhenoCorr takes Global Alliance for Genomics and Health (GA4GH) phenopacket data as input. A typical use case investigates a cohort of individuals with pathogenic variants in a given gene and compares different classes of variants (e.g., missense vs. nonsense), different locations (e.g., variants in distinct protein domains) or different variants (e.g., if there are variants that occur in multiple individuals). Each phenopacket represents a single patient with a list of recorded phenotypes and at least one variant. The software organizes the data into a table, showcasing all the recorded phenotypes, variants, affected transcripts, and the corresponding proteins along with the frequency of each within the cohort of patients. Users can choose from available comparisons to run based on the information in the table. Genotype phenotype correlations are then calculated by analyzing the counts of individuals with vs. without annotations to HPO terms. Results are presented in an interactive table. We evaluated GenoPhenoCorr on ten cohorts of individuals with diseases including KBG Syndrome and Isolated Sulfite Oxidase Deficiency. We could replicate results of some published genotype phenotype correlation studies and will present new results on other cohorts. GenoPhenoCorr is designed to be run in a Jupyter notebook environment, and requires only that users format existing data as GA4GH phenopackets. It enables interactive exploration of results, provides a library of functions for investigating correlations, and can be easily extended. By integrating novel algorithms, comprehensive data management, and interactive visualization, GenoPhenoCorr simplifies the identification of genotype phenotype correlations.

Session Title: Omics Technologies Poster Session I

PB3388 Genotyping coding-VNTR in the *MUC1* gene using k-mer frequency based alignment-free method allows genetic diagnosis of *MUC1*-related autosomal dominant tubulointerstitial kidney disease

Authors:

H. Saei[1], V. Morinière[2], L. heidet[1], O. Gribouval[1], s. Lebbah[2], F. Tores[1], M. Mautret[1], B. Knebelmann[2], S. Burtey[3], V. Vuiblet[4], C. Antignac[5], P. Nitschke[1], G. Dorval[1]; [1]Imagine Inst. of Genetic Diseases, Paris, France, [2]Univ. Hosp. Necker for Sick Children, Paris, France, [3]Marseille Publ. Univ. Hosp. System, Marseille, France, [4]Univ. Hosp. Ctr. Reims, Reims, France, [5]Imagine Inst.-Inserm UMR1163, Paris, France

The Human genome includes tandem repeats with variable length (VNTR) and a subset of these repeats have been associated with rare human diseases. Specific frameshift variants in the coding-VNTR of the *MUC1* gene cause autosomal dominant tubulointerstitial kidney disease - *MUC1* (ADTKD-*MUC1*). Calling variants from VNTR using short-read sequencing (SRS) is challenging due to poor read mappability, motif complexity, variable repetition, and enormous motif sequence similarity. To address this issue, we developed an efficient computational pipeline named VNtyper, for reliable detection of *MUC1* VNTR pathogenic variants and demonstrated its clinical relevance using two comprehensive patient cohorts. The pipeline utilizes mapping-free genotyping using k-mer frequencies (Kestrel algorithm) to detect small indels within MUC1-VNTR using SRS. To identify and analyze deleterious variants, we employed a combination of our self-developed *MUC1*-VNTR motif dictionary and a mapping-free genotyping algorithm. This approach allowed us to accurately detect and process these variants. The pipeline was applied to SRS data from two distinct cohorts. The first cohort, known as the historical cohort, consisted of 108 families (237 individuals) where at least three family members exhibited ADTKD including 32 families with known *MUC1* related ADTKD. The second cohort, referred to as the replication naive cohort, comprised 2910 patients who had previously undergone testing for monogenic renal diseases using a gene panel including 228 genes. In the historical cohort, our pipeline successfully re-identified all cases previously confirmed to carry pathogenic *MUC1* variants using the fastidious probe-extension assay. Additionally, we identified a previously unknown 25bp frameshift insertion in another family that had been overlooked. Within the replication cohort, we identified and validated 30 new cases with pathogenic variants that were overlooked. By establishing a minimal confidence depth-score, we achieved a sensitivity of 100% and a specificity of 94.40% in VNtyper testing. These findings demonstrate the effectiveness of our well-optimized pipeline in accurately screening the *MUC1*-VNTR region for the detection of deleterious variants. This pipeline could be simply integrated to SRS gene panel testing, and consistently call small indels in the *MUC1* VNTR, hence improving ADTKD-*MUC1* diagnosis, especially as innovative therapies begin to emerge.

Session Title: Omics Technologies Poster Session II

PB3389 Germline short variants detection from UG100 Sequencing Data

Authors:

D. Shem-Tov[1], I. Soifer[1], A. Jaimovich[2], Y. Etzioni[1], D. Lipson[2], O. Barad[1]; [1]Ultima Genomics, Tel-Aviv, Israel, [2]Ultima Genomics, Menlo Park, CA

Accurate detection of short variants is crucial for understanding the genetic basis of complex diseases and guiding precision medicine approaches. Variant calling (VC) algorithms discern true biological mutations from errors, by inspecting multiple reads at each genomic position. Here we present advances in the UG100 sequencing chemistry and analysis methods that result in improved germline VC performance.

HG001-HG005 cell lines were sequenced on the UG100 at 40X coverage. New sequencing chemistry and base-calling algorithm were used to obtain better coverage over AT rich regions and improved coverage uniformity. DeepVariant (DV) algorithm was further optimized. The filtering accuracy of the DV model was improved by enriching the training-set, and by adding an extra channel encoding the size of insertions. Furthermore, the sensitivity of the candidate generator was elevated by separating and calibrating allele-frequency thresholds of homopolymer and non-homopolymer indels. Additionally non-homopolymer insertion sensitivity was improved by considering imperfect-matches between reads to haplotype as potential support. Finally, the robustness of DV to coverage variations was improved by two complimentary methods. Using adaptive down-sampling, the observed mean coverage is automatically adjusted to match the coverage the model was trained on. This mode is suitable for over-sequenced WGS samples, where we want to preserve the natural WGS coverage variance. In coverage-capping mode, we truncate the number of reads shown to the model at a maximal value, which is suitable for cases where the coverage is more variable such as WES. Using coverage-capping enables us to call WES samples with general WGS models, without retraining on WES data.

The modifications lead to a measured average improvement in coverage uniformity F95 from 1.8 to 1.5, and an improvement in base-quality Q30 from 85% to 89%. To evaluate the variant calling performance, we used GIAB v4.2.1 as ground truth, with corresponding High Confidence Regions (HCR) excluding homopolymer regions of length greater than 11 (0.3% of HCR), and low complexity regions (1.5% of HCR). The SNP F1 results improved from 99.6% to 99.8% and indel F1 from 96.4% to 98.6%.

In conclusion, our study demonstrates a set of improvements to the UG sequencer and DV for VC on UG germline data, which offers a robust, accurate, and efficient VC solution, enabling researchers and clinicians to unravel the genetic underpinnings of diseases and advance personalized medicine.

Session Title: Omics Technologies Poster Session III

PB3390 Global metabolomics is a powerful tool for diagnosing rare diseases devoid of clear genetic causalities

Authors:

A. Kennedy[1], K. DeBalsi[1], A. Crawford[2], L. Sommerville[1], A. Evans[1]; [1]Metabolon, Durham, NC, [2]Illumina Inc, San Diego, CA

Inborn errors of metabolism (IEMs) are rare inherited disorders that disrupt metabolic function to result in health complications or death. The symptomologies of many IEMs are non-specific, and a genetic cause is not always clear. In these cases the lack of specific tests can appreciably delay diagnosis, leading to poor outcomes. Addressing this gap in care relies on gaining deeper insight into a patient's phenotype from diagnostic testing. Here, we evaluated untargeted metabolomics as a tool for diagnosing IEMs. Using Metabolon's Global Discovery Platform, we obtained metabolomic profiles of plasma and urine from 29 persons suspected of having an IEM but with no clear underlying genetic susceptibility. Raw peak metabolite intensities were normalized, log-transformed, and Z-scored. Patient metabolite concentrations that were higher than the 95th percentile or lower than the 2.5th percentile of the normal range were flagged as potential diagnostic markers. We found evidence of numerous IEMs in the cohort including dysregulation of glycolysis, fatty acid oxidation, biosynthetic pathways, and mitochondrial function. In nearly every case, global metabolomic profiling identified a pathway disruption that either suggested an IEM or was key to clinical decision making. One patient presented with muscular hypotonia, seizures, and ataxia. Metabolic profiling showed a significant increase in acylcarnitines, hydroxy-acylcarnitines, and beta-hydroxy fatty acids that are elevated in certain mitochondrial disorders. Follow-up targeted molecular testing revealed a diagnosis of mitochondrial aminoacyl-tRNA synthetase deficiency. Another patient presented with non-specific growth delay and intellectual disability. His metabolic profile revealed significantly elevated amino acid metabolites and bilirubin, and significantly reduced carnitine; all suggestive of liver dysfunction. Indeed, this patient was eventually found to have a mutation in *KMT2D*, revealing a diagnosis of Kabuki syndrome. Another patient had significantly decreased metabolites upstream of DHODH, an enzyme involved in pyrimidine synthesis. Follow-up genetic testing showed a pathogenic variant in *DHODH*, the gene that encodes the enzyme. Yet, a pathogenic variant would be expected to decrease DHODH enzymatic activity to thereby elevate upstream metabolites. Finding that this patient showed the opposite effect impacted their treatment plan considerably. Altogether, this study provides compelling support for including global metabolomics in IEM diagnostic testing as a complement to genetic profiling to ultimately reduce the time-to-diagnosis and improve patient outcomes.

Session Title: Omics Technologies Poster Session I

PB3391 † Health and economic outcome analysis of patient journeys in a large seizure cohort: a case for rapid whole genome sequencing.

Authors:

R. Subaran[1], D. Corrigan[1], Z. Liu[1], K. Parsons[2], M. Higashi[1]; [1]GeneDx, Stamford, CT, [2]Gene Dx, Stamford, CT

Seizures are the most frequent neurological emergencies in children, including neonates. Since the risk of adverse neurodevelopmental outcomes increases with the time elapsed between seizure onset and remediation, rapid precision intervention is paramount. However, seizure phenotypes are highly heterogenous, and depending on the underlying molecular and neurophysiological etiology, many seizures can be refractory to, and even contraindications for, common first- and second-line therapies. Rapid whole-genome sequencing (rWGS) has been transformative in seizure intervention, with consensus mounting that differential targeting of epilepsy pathways based on individualized genomic results can significantly improve clinical outcomes. Unfortunately, despite this evidence, the prospect of adopting or expanding rWGS in many clinics, particularly in historically underserved populations, is met by competing economic considerations. Thus, there is an impetus to synthesize and articulate epilepsy patient odyssey data that can allow providers to also make the economic case for rWGS. Here, we explore the treatment odyssey of 20,712 consented and deidentified seizure patients recruited at The Mount Sinai Hospital (New York, NY) with longitudinal electronic health record data covering diagnoses between 2004 and 2020. We stratified epilepsy groups into neonatal-, infant-, early childhood-, late childhood-, and adolescent-onset, then performed patient journey mapping focusing on seizure 1) phenotype and severity, 2) nosocomial procedures, and 3) changes in medication. Using data from third-party payers across five states (MA, CA, MA, OH, PA, TX), we were able to estimate the median cost of imaging-based diagnostic procedures across each journey. We found that failure to resolve seizures in the first visit was associated with an estimated average imaging-based diagnostic cost of $3,929.61 USD (range by state = $2,532.05 - $6,454.50) in the total cohort and most pronounced in adolescent-onset population of the subgroups examined (estimated average = $4,073.52). Patients in this group who required imaging also had the had the highest incidence of status epilepticus (average per journey = 1.25) - a severe seizure phenotype characterized by prolonged and/or rapidly recurring paroxysmal episodes and which is associated with increased risk of brain damage and death - suggesting special urgency for rapid seizure resolution in this group. Therefore, our patient journey mapping analysis in this cohort adds to the clinical case for diagnosis via rWGS and poses an economic case for its adoption.

Session Title: Omics Technologies Poster Session II

PB3392 Heritability, pQTLs, and environmental influence of proteins involved in age and cardiovascular disease using highly multiplex proteomics

Authors:

B. Epstein, T. Lai, **T. Johnson**, M. A. Hinterberg, D. P. Astling; SomaLogic, Inc, Boulder, CO

Background: Highly multiplex proteomic assays have enabled a growth in protein quantitative trait locus ("pQTL") studies. By linking genomic variation to the proteome, pQTLs can reveal the mechanisms underlying the genetic risk of disease and suggest potential therapeutic targets. Here we explore the relationship between genetic effects on variation in protein abundance, as captured by pQTL studies, and associations between phenotypic or physiological variation and protein abundance.
Methods: We compiled a list of pQTLs and protein abundance heritability estimates from 15 pQTL studies based on the SomaScan™ assay. We calculated the association between protein analytes and age and cardiovascular disease risk (CVD), dividing the analytes into those strongly associated (FDR < 10%) and not strongly associated (FDR > 10%) with the endpoints. For analytes in each endpoint association category, we calculated the proportion that were tagged by a cis-pQTL and examined the distribution of heritability.
Results: Across the ~ 5,000 analytes, about half (50.6%) were cis-pQTL-associated, while 30% had significant heritability. By contrast, 70% of analytes strongly associated with age or CVD were also pQTL-associated, and 48-50% (depending on ancestry) had significant heritability. Overall, mean heritability across analytes was 13-14% regardless of endpoint association status.
Conclusions: Proteins associated with aging or CVD are more likely to have evidence of genetic influence than proteins not associated with aging or CVD, supporting the use of proteomic data in Mendelian randomization and proteome-wide association studies. However, there are still many proteins that have a strong association with an endpoint, but no cis-pQTL association or low heritability. These proteins are likely strongly influenced by environmental, rather than genetic, factors. The ability of proteomics to capture the effects of environmental variation complements the genetic effects captured by genomics.

Session Title: Omics Technologies Poster Session III

PB3393 High throughput human sample prep and sequencing on PacBio Revio system

Authors:

D. Moine[1], G. Young[1], J. Burke[1], J. Rocha[2], P. Baybayan[3], I. McLaughlin[4], K. J. Liu[1]; [1]PacBio, Menlo Park, CA, [2]Pacific BioSci.s, Bel Air, MD, [3]Pacific BioSci.s, Menlo Park, CA, [4]Pacific BioSci.s of California, Inc., Menlo Park, CA

Improved throughput and cost of long-read sequencing, driven by recent technological advances of the PacBio Revio system, enables investigation of whole human genomes across larger populations. To support the growing capabilities of long-read sequencing, high throughput (HT) sample and library preparation solutions are necessary. We present a fully automated HT DNA extraction, shearing, and library preparation workflow for different human sample types for PacBio HiFi sequencing. High molecular weight (HMW) DNA extraction is performed utilising Nanobind magnetic disk technology on automated Hamilton NIMBUS Presto or ThermoFisher KingFisher systems. Nanobind disks feature micro-and-nanostructured silica wrinkles to shield bound DNA from damage during extraction. We obtained ~6 µg of DNA per 200 µL blood sample (96 well plate) and ~30 µg of DNA per 1 mL blood sample (24 well plate) in 2.5 hours. HMW DNA is then sheared to 15-20 kb using robotic pipette shearing. Automated PacBio library and loading preparation is then performed on the fully automated Hamilton NGS STAR. The methods presented utilize standard configurations of Hamilton instruments and can easily be incorporated into existing workflows. Data is presented using different sample types for a workflow which can prepare 96 samples from DNA extraction to library ready for loading in ~10 hours. A Revio SMRT Cell typically generates ~30X coverage of high-quality sequence data sufficient for analysis including phasing, 5mC, and variant calling.

Session Title: Omics Technologies Poster Session I

PB3394 High-density and scalable protein arrays are generated for comprehensive single-molecule proteomic studies.

Authors:

J. Kuhar, M. R. Newman, J. K. Robinson, B. Nortman, S. Paul, A. Sohi, M. De Garate, C. Guerrero, K. Cosert, C. Inman, P. Indermühle, P. Hao, T. Aksel, H. Qian, S. Yan, T. Gillies, S. Tan, S. Sankar, S. K. Wilcox, P. Mallick; Nautilus Biotechnology, San Carlos, CA

Introduction: Massively parallel Next-generation DNA sequencing dramatically improved access to genomic data by increasing the scale by which DNA could be read, however to-date proteomic analyses have been severely limited in scale. To solve this challenge, we created a novel, highly-scalable system that allows for single-molecule protein interrogation across a wide dynamic range from minimal sample input. This system comprises a mono-disperse DNA scaffold with a single protein attachment site and a nanoscale-patterned surface with billions of landing pads. Our solution creates massively parallel super-Poisson single-molecule protein arrays where captured proteins are bound and can be interrogated repeatedly by multi-affinity probes or traditional affinity reagents.
Methods: Lysate is processed through a denaturing workflow that chemically disrupts cysteine-mediated complexes and covalently adds a methyltetrazine click group to amines to generate functionalized single polypeptides. The methyltetrazine is subsequently used to conjugate proteins to a DNA scaffold displaying a single trans-cyclooctene (TCO) moiety. In this manner, we efficiently create a protein library consisting of a population of protein-conjugated scaffolds, each displaying only a single individual protein captured from the sample lysate. The scaffold ensures spatial isolation of individual proteins when deposited onto a hyper-dense nanostructured array containing DNA binding sites surrounded by HMDS modified interstitial chemistry. To demonstrate single-molecule occupancy of the array, the scaffolds are dye labeled and loaded for imaging.
Results: When low micrograms of human cell lysate are used with the workflow we observe DNA scaffolds conjugated to a diverse set of proteins. The workflow disrupts model protein complexes and does not impair detection by multi-affinity probes. Minimal co-localization of two or more polypeptides is observed. A nanoarray loaded with protein-scaffold conjugates shows approximately 98% single molecule occupancy.
Conclusions: We demonstrate a workflow for the straightforward preparation of conjugated protein libraries that can be deposited as single molecules on a densely patterned array. When interrogated with multi-affinity probes, these libraries can provide proteomic information at scale.

Session Title: Omics Technologies Poster Session II

PB3395 † High-throughput Allele-Specific Expression Analysis Can Detect Allelic Imbalance in Clinical Patient Samples as a Proxy for Quantitative Measures of Gene Expression

Authors:

H. Mirsafian[1], H. Zimmermann[1], L. Hudish[1], S. Belhadj[1], D. Cheng[2], E. Chao[1], R. Karam[1], M. Schultz[2]; [1]Ambry Genetics, Aliso Viejo, CA, [2]REALM IDx, Aliso Viejo, CA

RNA analysis has recently emerged as a powerful tool for variant classification. However, most RNA analysis performed on clinical RNAseq data involves the detection of qualitatively different transcripts produced as a result of aberrant splicing. Identifying quantitative differences in the amount of transcript produced from an allele at the gene-level is a more difficult task and has historically relied upon quantitative real-time PCR (qRT-PCR) experiments. In this study we used the phASER tool on high-throughput clinical paired DNAseq and RNAseq samples to detect allelic differences in gene expression caused by *cis*-acting variants. phASER employs RNASeq reads to phase heterozygous variants (called from DNASeq data) relative to one another within a given gene. Allele-specific expression (ASE) analysis is then applied to evaluate the relative expression of the two alleles present at heterozygous loci to produce a single expression measurement for each haplotype. To validate this method in our clinical samples, we performed ASE analysis on samples from two patients heterozygous for a known pathogenic variant in the 5`UTR of the *MLH1* gene (*MLH1* c.-27C>A) that has been shown to result in reduced promoter activity. As predicted, this ASE analysis was able to detect significant ASE of *c.-27C>A* and a downstream benign polymorphism (*c.655A>G*) in both probands and not in controls. We then performed this analysis on two novel variants impacting the canonical donor splice sites of the noncoding first exons of *BRCA2* (*BRCA2* c.-40+2T>C) and *TP53* (*TP53 c.-29+1G>C*). We suspected these variants may cause ASE based on literature reports demonstrating ASE for other variants impacting the same canonical donor sites. We identified two patients heterozygous for the *BRCA2* variant and one for the *TP53* variant and were indeed able to show significant ASE for benign downstream heterozygous polymorphisms in the 5`UTR (*c.-52A>G*, *c.-26G>A)*, exon 10 (*c.1114A>G*), exon 11 (*c.3396A>G*), and exon 14 (*c.7242A>G*) of *BRCA2*, and in exon 3 of *TP53* (*c.215C>G*). ASE was not detected in controls. Although preliminary, these data indicate that ASE analysis using phASER is capable of detecting RNA allelic imbalance associated with variants identified through clinical DNA samples. These results raise the tantalizing possibility that high-throughput ASE analysis could be used to identify patients with allele-specific expression of unknown origin. Future directions of this study include validating the method for additional variants and variant types and assessing potential candidate variants in a variety of genes using our in-house database of >300000 paired DNAseq and RNAseq samples.

Session Title: Omics Technologies Poster Session III

PB3396 High-throughput full-length isoform sequencing with long, accurate reads

Authors:

E. Tseng[1], H. Ferrao[1], J. Bruand[1], A. Toepfer[1], M. Mehlferber[2], V. Pavelko[2], G. Sheynkman[2]; [1]Pacific BioSci.s, Menlo Park, CA, [2]Univ. of Virginia, Charlottesville, VA

Long-read RNA sequencing has been demonstrated to eliminate the need for short-read transcript assembly and applied to discovering novel isoforms in disease research. Long-read transcript sequencing using PacBio (the Iso-Seq method) can characterize full-length isoforms and has been used in bulk RNA studies to identify alternative splicing, fusion genes, and allele-specific isoform expressions. However, the high sequencing cost and lower throughput on previous PacBio systems have prohibited the ability to perform scalable isoform classification & quantification over large number of samples.

PacBio is commercializing the MAS-Seq method (Al'Khafaji et al., 2021) - an amplicon concatenation method for throughput increase - for bulk isoform sequencing. Using an 8-fold concatenation approach, the MAS-Seq for bulk Iso-Seq kit achieves an 8-fold throughput increase on PacBio Sequel II/IIe and Revio systems. We demonstrate that we can achieve 40-60 million full-length reads per sample on the Revio system, where the average transcript is ~2.5kb spanning from 100 bp to over 10,000 bp. Further, we show that the concatenation approach maintains the relative abundance of cDNA molecules without concatenation, an important need for facilitating isoform-level quantification and differential analysis.

The companion SMRT Link software release contains a new Iso-Seq workflow that robustly analyzes full-length reads, from isoform-level clustering to transcript characterization and quantification. The workflow integrates with popular tertiary analysis tools such as tappAS and Cerebus for differential expression analysis and joint sample transcript annotation.

Concurrent with the kit and software, we release a high-accuracy Iso-Seq reads HG002 dataset demonstrating saturation of known genes and isoforms that can be used for variant calling and detection of allele-specific isoform expressions. A second dataset consists of 11 WTC-11 samples totaling over 600 million reads - the largest bulk Iso-Seq dataset for a single cell line to date. In contrast to previous PacBio sequencers and Iso-Seq methods, this scope would have required >200 SMRT Cells 8M and over 200 sequencing days to generate the equivalent amount of data. Here, the WTC-11 dataset required 11 Revio SMRT Cells which were sequenced in 3 days. This dramatic throughput increase and reduction in sequencing time will enable high-throughput, population-scale full-length RNA sequencing that has the potential to reveal novel insights into human disease.

Session Title: Omics Technologies Poster Session I

PB3397 HiPhase: Jointly phasing small and structural variants from HiFi sequencing.

Authors:

J. Holt, C. Saunders, W. Rowell, Z. Kronenberg, A. Wenger, M. Eberle; PacBio, Menlo Park, CA

In diploid organisms, phasing is the problem of assigning heterozygous variants to one of two haplotypes. For human genetics, phasing variants within a gene is required for clinical applications including donor matching (HLA typing) and determining drug interactions (pharmacogenomics). Additionally, knowing whether variants occur in cis or trans is an important factor in identifying the causes of autosomal recessive conditions both for known conditions (e.g., Tay Sachs) or when trying to identify the causes of novel rare diseases. While inherited variants can be phased with parental information, read-backed phasing allows all variants, including de novo variants, to be phased using read level evidence. These reads can also be assigned to a haplotype, enabling epigenetic analyses linking genetic and epigenetic changes on the same haplotype to detect allele specific methylation and imprinting disorders (e.g., Prader-Willi syndrome). Reads from PacBio HiFi sequencing provide long, accurate observations that can be used as the basis for identifying and phasing both inherited and de novo variants of all types. However, current phasing tools typically only phase smaller SNV and indel variants, leaving larger structural variants unphased.

We developed HiPhase, a tool that jointly phases small (SNV/indel) to large (structural) variants in a single, easy-to-use program. Compared to the current state-of-the-art approach, HiPhase increased phase block NG50 by 21% (493 kb), decreased errors by 42% (933 switchflips), increased the number of fully phased genes by 3.6% (90.6% of genes), and enabled allele-specific methylation analysis for 2.2% more CpG islands (73.6% of CpG islands). To our knowledge, HiPhase is the first tool to jointly phase small variants (SNV/indel) with structural variants.

Additionally, HiPhase includes novel algorithmic contributions, innate multi-threading, statistics gathering, and concurrent phased alignment output generation. The improvements from HiPhase in fully phased genes and CpG islands can provide more explanations in rare disease research.

HiPhase is freely available on Bioconda with documentation on GitHub:
https://github.com/PacificBiosciences/HiPhase.

Session Title: Omics Technologies Poster Session II

PB3398 † Human Pangenome Reference Consortium Coordinating Center.

Authors:

T. Lindsay[1], L. Antonacci-Fulton[2], S. Cody[1], H. Lawson[1], T. Wang[3]; [1]Washington Univ., Saint Louis, MO, [2]Washington Univ., St. Louis, MO, [3]Washington Univ. in St. Louis, St. Louis, MO

The goal of the Human Pangenome Reference Consortium (HPRC) is to increase the diversity of the human reference by moving from a linear reference to one comprised of many genomes. Including more diversity in the human reference will lead to a better understanding of human variation, which patients and clinicians need for better outcomes in the clinic.
Plans include data generation from 350 genomes over the five-year grant, with many of the genomes selected from the 1,000 Genomes Project. Development efforts focus on inventing the pangenome tools the research and clinical communities need.
An international outreach effort, Human Pangenome Project, is also underway to expand the number of partners leading to additional genomes being made available for the resource. These efforts will include the coordination of events, and discussions in pangenomic resource development with the aim of fostering necessary interaction and debate in the area of genomics and ethical oversight, Identifying projections of diversity in the pangenome, developing strategies to evaluate and measure gaps in diversity, forming international partnerships to establish an accessible resource to improve clinical genomics and biomedical research globally, openly share protocols and methods, and community engagement.
The HPRC is a multi-component consortium, and efficient and effective logistical and scientific coordination is critical to maximizing the value of the project. The HPRC components include the Coordinating Center and Sequencing Center. The HPRC organizational framework comprises a steering committee and working groups with specific work streams. Unique to the HPRC is a funded ELSI component consisting of leaders in this field. Their primary function is to help guide the HPRC as it negotiates sample selection, data use, and other situations.
Within the past year, the HPRC has released our first Human Pangenome assembly graph, comprised of 47 diploid individuals built together into one Pangenome. The complete resource contains each genome's raw sequence data and haploid assemblies. The sequencing data types include Pacific Biosciences HiFi data, Oxford Nanopore Technologies (ONT) long-read sequencing, Bionano optical maps, and Hi-C Illumina short-read sequencing for all HPRC samples. Researchers and clinicians can access the resources provided by the HPRC through GitHub, AnVIL, and NCBI.
We will fully detail the workings of the consortium and available resources that researchers and clinicians can access.

Session Title: Omics Technologies Poster Session III

PB3399 Human plasma metabolite outliers identified in Estonia Biobank knockouts

Authors:

L. Bomba[1], K. Yu[2,1], T. Esko[3], K. Estrada[1], A. Wuster[1]; [1]BioMarin Pharmaceutical Inc., Novato, CA, [2]Univ. of Michigan, Ann Arbor, MI, [3]Estonian Genome Ctr., Inst. of Genomics, Univ. of Tartu, Tartu, Estonia

Metabolomics, in combination with genetic data, is a powerful technology to assess the biochemical consequences of genetic variation. We studied the impact of human gene knockouts (KOs) on the metabolite levels of the Estonia Biobank (EstBB) participants and integrate these findings with electronic health record (EHR) data. We leveraged 723 knockouts selected from 153 loss of function (LoF) variants and 258 controls from the EstBB with 1,505 metabolites profiled using ultra-high-performance liquid chromatography-tandem mass spectrometry. We define a human knockout as an individual carrying at least two predicted homozygous LoF variants in a gene. The LoF variants were validated using Sanger sequencing. We identified 48 associations implicating rare LoF variants in 22 genes and 43 metabolites using Wilcoxon rank sum test. Approximately two-thirds of the associations were found in genes that cause inborn errors of metabolism (IEM). The top associations identified in our analysis included genes and metabolites involved in the degradation pathway of the pyrimidine bases uracil and thymine (*DPYD* and *UPB1*). We found *DPYD* gene knockouts associated with elevated levels of Uracil, confirming that DPD-deficiency is a leading cause of severe 5-Fluorouracil toxicity. Overall, 54% of reported associations are gene targets of approved drugs or bioactive drug-like compounds. Our findings contribute to assessing the impact of human knockouts on metabolite levels offering insights into gene functions, disease mechanism and drug target validation.

Session Title: Omics Technologies Poster Session I

PB3400 Identify cell-type-specific and shared eQTLs in single-cell eQTL analysis by leveraging cell type similarities.

Authors:

C. Lin[1], W. Li[1], C. Su[1], Y. Lin[2], H. Zhao[1]; [1]Yale Univ., New Haven, CT, [2]The Univ. of Sydney, Sydney, Australia

Expression quantitative trait locus (eQTL) studies aim to identify the genetic variants and their effects on gene expression levels, providing insights into the genetic regulations of complex traits. While previous eQTL studies on bulk tissues provide an average effect across cell types and states, recent research has highlighted the significance of cell-type-specific eQTLs. The emergence of population-scale single-cell RNA sequencing (scRNA-seq) datasets with annotated cell types has greatly improved our understanding of cell-type-specific eQTLs. However, existing strategies often neglect the shared genetic effects among similar cell types, leading to potential loss of statistical power. In this study, we propose a statistical framework that leverages cell type similarities with a Bayesian model to identify cell-type-specific and shared eQTLs. Our framework begins by separately analyzing each cell type using MatrixEQTL with a pseudo-bulk aggregation of single-cell data. The estimated eQTL effect sizes are then used to capture the similarities between cell types incorporating the cell type ontology. By fitting priors for cell-type-specific eQTL effect sizes using these similarity matrices, we then apply Bayes' theorem to obtain posterior estimates of eQTL effect sizes that incorporate the information from multiple cell types to derive the false discovery rates. We demonstrate the performance of our framework in cell-type specific eGene identification using the OneK1K scRNA-seq data with peripheral blood samples from 981 donors. We identified a total of 7544 eGenes (genes associated with at least one significant eQTL) across 29 cell types. Notably, 4156 eGenes exhibited cell-type-specific genetic effects, whereas the other 3388 eGenes demonstrated shared genetic effects among similar cell types. Comparing our framework with conducting MatrixEQTL separately in each cell type, we observed an 18% increase in the number of shared eGenes, suggesting increased statistical power to identify shared genetic effects across similar cell types. By leveraging information across cell types, our framework also yielded greater power in identifying eQTLs for rare cell types (cell type proportion < 0.01% in the data), as evidenced by a four-fold increase in eGenes identified in CD4-positive, alpha-beta T cells, and a two-fold increase in eGenes identified in plasmacytoid dendritic cells. These findings underscore the importance of considering cell-type-specific and shared genetic effects when studying the regulatory mechanisms underlying complex traits.

Session Title: Omics Technologies Poster Session II

PB3401 Identifying genetic variation in cancer using targeted long-read sequencing.

Authors:

S. Iyer[1,2], M. Kramer[1], A. N. Habowski[1], D. A. Tuveson[1], S. Goodwin[1], W. McCombie[1]; [1]Cold Spring Harbor Lab., Cold Spring Harbor, NY, [2]Stony Brook Univ., Stony Brook, NY

Structural variations (SV) tend to be recurrent and contribute to cancer. Next-generation sequencing (NGS) technologies are mostly blind to many SVs, with false positive and false negative rates of >50% in SV detection. Long-read sequencing generates read lengths of tens of thousands of bases, helping identify thousands of genomic features previously missed by NGS. However, whole genome, long-read sequencing at the coverage needed to detect variants in heterogeneous samples is impractical due to its high cost. Targeted sequencing on the other hand, improves accuracy and coverage by providing the depth necessary to detect rare alleles in a heterogenous population of cells.

We developed ACME, an Affinity-based Cas9-Mediated Enrichment method, which is an improvement on the widely used nanopore Cas9-targeted sequencing (nCATS). Like nCATS, ACME is an amplification-free approach that can simultaneously assess genomic variations and DNA methylation from the same sequencing run. Importantly, ACME helps reduce background non-target reads, increasing coverage and size of the target regions that are captured. Using ACME to target 10 cancer genes in MCF 10A and SK-BR-3 breast cell lines, we showed a 2- to 25-fold increase in target coverage and 3- to 7-fold increase in number of reads that span the whole target from start to end compared to nCATS. Using ACME, we obtained >60-fold target enrichment, 35-65x coverage, and 3-20 end-to-end reads spanning *BRCA2*, a 95 kb target on our panel. The main advantage ACME offers over other amplification-free long-read targeting approaches is the ability to capture contiguous reads, up to 100 kb in size. This reduces mapping errors and aids in SV detection. As proof of concept, we showed that ACME successfully detected all SVs within our targets previously inferred by whole genome long-read sequencing.

Expanding on this work, we used ACME on four pancreatic ductal adenocarcinoma (PDAC) patient-derived organoids (PDOs). Organoid models represent the full spectrum of disease, making them invaluable to assess personalized treatment options. Using ACME we observed comprehensive SV detection within our 10 targets, of note, a homozygous exon deletion covering an ~500 bp region in the *STK11* gene in one of the organoids, which correlated with decrease in gene expression. We also identified known driver SNPs in the *KRAS* gene in all four organoids.

We are currently working on a multiplexed approach using native barcoding that would not only make ACME even more cost-effective but also take it a step further in helping answer longstanding biological questions.

Session Title: Omics Technologies Poster Session III

PB3402 IL1R+ antimicrobial molecule program and heterogeneity study.

Authors:

C. Hu, P. R. Andrade, M. Pellegrini, R. Modlin; UCLA, Los Angeles, CA

The study of the relationship between Th17 and Treg cells has emerged as a crucial area of research in immunology. T-helper 17 (Th17) cells have been described as “pathogenic” vs. “non-pathogenic” based on their contribution to tissue injury and their secreted cytokine pattern. Our previous study defines “antimicrobial Th17 cells” based on the expression of interleukin 1 receptor type I (IL-1RI), production of IL-26 and the ability to mediate a rapid antimicrobial response against extracellular bacteria. Our transcriptome analysis revealed that the IL-1RI+ Th17 cells express T regulatory (Treg) markers *FOXP3*, *LRRC32*, *CTLA4* and *CD25* (Weiss et al., 2019), leading to the hypothesis that some of these cells represent intermediates in Th17 development from Treg precursors (Valmori et al. 2010).Terminally differentiated cells fulfill opposite functions: Th17 cells cause autoimmunity and inflammation, whereas Treg cells inhibit these phenomena and maintain immune homeostasis. Thus, unraveling the mechanisms that affect the Th17/Treg cell balance is critical if we are to better understand autoimmunity and tolerance (Lee et al., 2018).Here we address this relationship by investigating the effect of TCR activation and IL-1β stimulation on the antimicrobial gene program and heterogeneity of memory Th17 cells populations that express IL1RI, as well as the differentiation trajectory of Th17 and Treg populations in acne and rosacea. We performed scRNA seq data on IL-1RI+ and IL-1RI- memory Th17 cells (*CD4+CD45RO+CCR6+CD161+CCR4+CXCR3-*) from five healthy donors stimulated or not with IL1β or anti-CD3/CD28 for 9 hours. After integration of the scRNA-seq data from the five donors using the R package Seurat, we identified 14 clusters across the IL-1RI+ and IL-1RI- Th17 populations and their conditions (media, IL1β and CD3/CD28). Our analysis identified a continuum between cluster Th17 cells and cluster Treg cells on the UMAP visualization, suggesting the existence of a progression between cellular states within the CD3-activated IL1RI+ Th17 population. In vivo, we performed scRNA seq analysis with the same process on acne and rosacea data. Using the same computational tools as in our in vitro data, we found a similar pseudotime trend going from the Th17 clusters to the Treg clusters, suggesting that the Th17 to Treg phenotypic differentiation route is also present in vivo. Our findings indicate that a portion of the Th17 IL1RI+ population consists of intermediate phenotypes that are differentiating towards T regulatory phenotype, which could be involved in control mechanisms of the inflammation elicited by antimicrobial responses.

Session Title: Omics Technologies Poster Session I

PB3403 Implementation of ONT long-read WGS in routine clinical genetic testing.

Authors:

A. Kaplun, L. Kaplun, G. Krautz-Peterson, N. Neerman, C. Stanley, S. Hussey, M. Folwick, A. McGarry, S. Weiss; Variantyx, Framingham, MA

Whole Genome Sequencing (WGS) has allowed exploration of areas of the genome that might not have been targeted by other approaches, such as intergenic regions. A single technique detecting all genetic variants at once is intended to expedite the diagnostic process while making it more comprehensive and efficient, making WGS a preferable approach as a first line diagnostic test. Nevertheless, there are still several shortcomings that cannot be effectively addressed by short read sequencing, such as determination of the precise size of short tandem repeat (STR) expansions, phasing of potentially compound recessive variants, resolution of some structural variants and exact determination of their boundaries, etc. Fortunately, those weak spots of short read sequencing can be compensated by long read sequencing technology that have comparable or better detection of some types of variants while lacking the mentioned above limitations of short read sequencing. We have developed an integrated clinical genetic testing approach, augmenting short read WGS-based variant detection with Oxford Nanopore Technologies (ONT) long read sequencing, providing improved detection and simultaneous orthogonal confirmation of all types of variants with the additional benefit of improved identification of exact size and position of the detected aberrations. The validation study of this augmented test has demonstrated that ONT sequencing can efficiently verify multiple types of reportable variants, thus ensuring highly reliable detection and a quick turnaround time for WGS-based clinical genetic testing.

Session Title: Omics Technologies Poster Session II

PB3404 Improving methods for high throughput epigenomics studies.

Authors:

B. Marosy[1], H. Ling[1], P. Zhang[1], J. Gearhart[1], J. Paschall[1], C. Oncago[1], M. Mawhinney[1], L. Aker[1], A. Robinson[1], B. Craig[1], Z. Xu[2], K. Doheny[1]; [1]Ctr. for Inherited Disease Res., DGM, Johns Hopkins Univ., Baltimore, MD, [2]Biostatistics and Computational Biology Branch, Natl. Inst. of Environmental Hlth.Sci., NIH, Research Triangle Park, NC

The Center for Inherited Disease Research (CIDR) provides high-throughput genomic services and statistical genetics consultation to investigators working to discover genes that contribute to diseases. In response to the growing requests for epigenetic studies, we recently evaluated an alternative chemistry for methylation detection - an enzymatic library conversion of methylated sites (NEB EMseq) paired with a targeted methylome panel (Twist BioSciences). The performance of this method (Methyl-Seq) was compared with our existing methylation array (EPIC), which uses bisulfite conversion. The Methyl-Seq and Illumina EPIC assays are designed to capture 3.98M and 935K CpG sites, respectively. On average, the Methyl-Seq data (with a mean coverage of 25-55x) detected 6.5M CpG sites per sample, approximately 7 times as many as the EPIC array. The peaks of raw beta value distribution in the sequencing data are tighter than the EPIC array for both methylation controls and experimental samples. The Pearson correlation coefficient, derived from HapMap duplicate pairs (9 for EPIC and 3 for Methyl-Seq), exhibits comparable values for both sequencing and array data, exceeding 0.99. For both, the correlation of beta values between non-duplicate pairs (n=12 for Methyl-Seq and n=27 EPIC) falls within the range of 0.88 to 0.91, roughly 10% lower than the duplicate pairs. The between-technology correlation is 0.96 to 0.97 (18 cross-platform duplicate pairs). Our current EPIC array studies typically involve hundreds to a few thousand samples. However, the size of recently funded studies is expanding to the tens of thousands, necessitating the implementation of automated bisulfite conversion methods. We have successfully validated the Zymo EZ-96 DNA Methylation Lightning assay on our in-house automation, in preparation for population-scale studies. 32 samples were run with a mean of 932,871 (99.6%) %CpG detected at p=.01. Lastly, we have automated EPIC array sample level QC to date using *ewastools* and *minfi* packages. We are exploring the *ENmix* package to allow for a more in-depth QC including normalization on both sample and probe level, with the aim of providing our investigators with a dataset closer to “analysis ready.” From our initial analyses, Methyl-Seq is a promising new platform for epigenomic studies.

Session Title: Omics Technologies Poster Session III

PB3405 Inducible CRISPR interference in iPSC-derived neurons

Authors:

M. Nelson, D. Ramos, M. E. Ward; NIH, Bethesda, MD

CRISPR interference (CRISPRi) is a method for targeted gene knockdown. A single guide RNA (sgRNA) targets a catalytically dead Cas9 enzyme (dCas9) to a specific transcriptional start site, whereupon a transcriptional repression domain fused to dCas9 silences the targeted gene. CRISPRi facilitates multiplexed genetic screens in induced pluripotent stem cell (iPSC)-derived neurons: a library of sgRNAs can repress a catalogue of genes in a population of neurons, whose phenotypes can then be studied. Lentivirus is typically used to deliver sgRNA libraries, and common lentivirus backbones have poor tropism for iPSC-derived neurons. Thus, sgRNA libraries are commonly delivered to cells as iPSCs, which are then differentiated into neurons. In this paradigm, the effects of gene knockdown in iPSCs, differentiating cells, and post-mitotic neuron-like cells are conflated, making it difficult to study neuron-specific effects of gene knockdown. Here, we developed a method for inducible CRISPR interference using a Cre recombinase system in which dCas9 is expressed upon Cre delivery to post-mitotic iPSC-derived neurons. This limits screens solely to neuronal phenotypes. Our versatile platform for inducible CRISPRi is compatible with numerous sgRNA libraries, robustly induces dCas9 without detectable baseline expression, and results in gene knockdown in iPSC-derived neurons.

Session Title: Omics Technologies Poster Session I

PB3406 Inferring super-resolution tissue architecture by integrating spatial transcriptomics with histology.

Authors:

D. Zhang[1], A. Schroeder[1], H. Yan[1], H. Yang[1], J. Hu[2], M. Y. Lee[1], K. S. Cho[3], K. Susztak[1], G. X. Xu[1], M. D. Feldman[4], E. B. Lee[1], E. E. Furth[1], L. Wang[3], M. Li[5]; [1]Univ. of Pennsylvania, Philadelphia, PA, [2]Emory Univ., Atlanta, GA, [3]MD Anderson Cancer Ctr., Houston, TX, [4]Indiana Univ., Indianapolis, IN, [5]Univ. of Pennsylvania, Philadelphia, PA

Spatial transcriptomics (ST) has demonstrated enormous potential for generating intricate molecular maps of cells within tissues. Here we present iStar, an end-to-end workflow that integrates ST data and high-resolution histology images to predict spatial gene expression and characterize tissue architecture with super-resolution. Our method not only enhances gene expression resolution to near-single-cell levels in ST, but also enables gene expression prediction in tissue sections where only histology images are available.

Session Title: Omics Technologies Poster Session II

PB3407 Integrated ambient modeling and genetic demultiplexing of single cell RNA+ATAC multiome experiments with Ambimux

Authors:

M. Alvarez[1], T. Li[1], S. Lee[1], U. Arasu[2], I. Selvarajan[2], T. Örd[2], E. Rahmani[3], Z. Chen[4], O. Avram[3,4,5], A. Kar[1], D. Kaminska[6,7], V. Männisto[8], E. Halperin[1,3,4,5,9], J. Pihlajamäki[7,10], C. Luo[1,11], M. Kaikkonen[2], N. Zaitlen[3,12], P. Pajukanta[1,9]; [1]Dept. of Human Genetics, David Geffen Sch. of Med. at UCLA, Los Angeles, CA, [2]A. I. Virtanen Inst. for Molecular Sci., Univ. of Eastern Finland, Kuopio, Finland, [3]Dept. of Computational Med., David Geffen Sch. of Med. at UCLA, Los Angeles, CA, [4]Dept. of Computer Sci., Univ. of California Los Angeles, Los Angeles, CA, [5]Dept. of Anesthesiology and Perioperative Med., David Geffen Sch. of Med. at UCLA, Los Angeles, CA, [6]Dept. of Med., Div. of Cardiology, UCLA, Los Angeles, CA, [7]Inst. of Publ. Hlth.and Clinical Nutrition, Univ. of Eastern Finland, Kuopio, Finland, [8]Inst. of Clinical Med., Univ. of Eastern Finland and Kuopio Univ. Hosp., Kuopio, Finland, [9]Inst. of Precision Hlth., Univ. of California Los Angeles, Los Angeles, CA, [10]Dept. of Med., Endocrinology and Clinical Nutrition, Kuopio Univ. Hosp., Kuopio, Finland, [11]Ctr. for Neurobehavioral Genetics, Semel Inst. for NeuroSci. and Human Behavior, David Geffen Sch. of Med. at UCLA, Los Angeles, CA, [12]Dept. of Neurology, Univ. of California, Los Angeles, Los Angeles, CA

Single cell technologies have advanced at a rapid pace, providing assays for various molecular phenotypes. Droplet-based single cell technologies, particularly those based on nuclei isolation, such as simultaneous RNA+ATAC single cell multiome, are susceptible to exogenous ambient molecule contamination, which can increase noise in cell type-level associations. We reasoned that genotype-based sample multiplexing can provide an opportunity to infer this ambient contamination by leveraging DNA variation in sequenced reads. Thus, we developed Ambimux, a likelihood-based method to estimate ambient fractions and demultiplex single cell multiome experiments using genotype-level data. Ambimux models the ambient or nuclear probability at the read level, and thus can classify empty droplets and estimate droplet-specific ambient molecule fractions in each modality. We first evaluated our method using simulated data sets across a range of parameters. As an example, we simulated an 8-sample multiome data set across a uniform range (0-100%) of ambient contamination. We found that Ambimux closely estimated the ground truth droplet contamination fractions in the RNA (MSE=0.0088) and ATAC (MSE=0.0055) modalities. As a result, Ambimux maintained high specificity (>95%) and was able to correctly assign singlets at considerably high ambient fractions (up to 80%) for both RNA and ATAC modalities. In comparison with models that do consider ambient contamination, e.g. Demuxlet, these only maintained similar sensitivity levels at considerably lower ambient fractions (up to 26%). We then generated a real data set of seven visceral adipose tissue biopsies run on a single 10X Multiome channel. We ran Ambimux and detected 4,595 singlets, capturing 96% of those identified by Demuxlet in either modality and taking ~9 hours less than Demuxlet. Then, we sought to evaluate the fidelity of the ambient fraction estimates from Ambimux. We split the 4,595 singlets into size-matched ambient-enriched (>=25% contamination in either modality) or nuclear-enriched (<25% in both) droplets and performed gene-peak linkage analysis in 310 adipocyte marker genes. While the clean set of droplets resulted in 522 significant gene-peak links across 217 marker genes, the ambient-enriched droplets exhibited less links (n=107) across fewer marker genes (n=79), suggesting that the ambient droplets identified by Ambimux hamper the identification of biologically meaningful signals. In summary, we developed a joint single cell multiome demultiplexing method, Ambimux, that accurately models and estimates ambient molecule contamination in each modality.

Session Title: Omics Technologies Poster Session III

PB3408 Integrated multimodal analysis of simultaneous single-cell profiling opens a new frontier for characterizing cell states.

Authors:

H. Lee[1,2]; [1]Brigham Women's Hosp. and Harvard Med. Sch., Boston, MA, [2]Univ. of Massachusetts Amherst, Newton, MA

Recent advances in single-cell sequencing methods uncover cellular heterogeneity and dynamics of individual cells that conventional bulk-population level measurements cannot reveal. Additionally, simultaneous multimodal single-cell profiling promises a new frontier for single-cell genomics, demanding rigorous computational methods to define cell states based on multimodal data. In this study, we present our experimental and analytic framework based on peripheral blood mononuclear cells collected at the baseline visit of four participants in the autoimmunity-blocking antibody for tolerance (AbATE) clinical trial and four participants in the high-dose immunosuppression and autologous transplantation for multiple sclerosis (HALT-MS) clinical trial at Immune Tolerance Networks (ITN). Equal number of sorted T cells (CD3+CD19-CD14-CD56-) and B cells (CD3-CD19+CD14-CD56-) from each subject are pooled together for droplet-based single-cell assay and downstream analyses. We simultaneously sequenced the multiplexed samples through two sequencing platforms, single-cell multiome ATAC+Gene Expression and CITE-seq (RNA and 163 antibody proteins) and TCR/BRC repertoire. We performed demultiplexing and quality-control on the pooled samples based on dense genotyping information of the subjects and obtained 7102 nuclei for the multiome platform and 21850 cells for the latter platform for analysis. The T cells and B cells are identified and characterized by simultaneous profiling of gene expression, surface protein, BCR/TCR repertoire and chromatic accessibility through the two sequencing platforms in parallel. For downstream analyses, we first merge the RNA data from both platforms and link the merged RNA data with surface protein data by using canonical correlation analysis. we adopt machine learning techniques such as dimension reduction and clustering algorithms and annotate each cluster of cells based on marker genes and reference mapping. We examine clonal expansion using BCR/TCR data and identify open chromatin regions in our peak-calling analysis using ATAC data in the annotated clusters. This experimental and analytic framework highlights the potential application of simultaneous multimodal single-cell profiling in tracking changes of profile over time during immune-modulating therapies.

Session Title: Omics Technologies Poster Session I

PB3409 Integrative approach to reveal adverse impacts of butylparaben in homologous human genes from *Daphnia magna* via BLAST and network analysis.

Authors:

J. Lee[1], Y. Jang[1], S. Dong Yeop[1], H. Jun Pyo[1], S-W. Lee[2], D. Choi[3], E-M. Cho[4], J. Lee[5], **Y. Seo**[1]; [1]Dongguk Univ., Goyang, Korea, Republic of, [2]Korea Conformity Lab., Incheon, Korea, Republic of, [3]Korea Conformity Lab., Incheon, Korea, Republic of, [4]Seokyeong Univ., Seoul, Korea, Republic of, [5]Gachon Univ., Seongnam, Korea, Republic of

Butylparaben, an estrogen-like compound, is a common pollutant in aquatic ecosystems and a major threat to aquatic life and other organisms, including humans. Some previous reports have indicated that this pollutant has been toxic to human despite this pollutant having been conventionally regarded as 'safe' and widely used commercially such as an antimicrobial preservative in cosmetics. However, most research on butylparaben toxicity in humans is limited to acute toxicity, while only a few studies have elucidated the underlying molecular mechanisms of butylparaben exposure in humans. Considering its property and usage, we conducted a chronic reproductive test with *Daphnia magna* (*D. magna*), which has a unique parthenogenic and short reproductive cycle. As a result, we discovered that butylparaben at a dose of 0.32 mg/L for 21 days significantly reduced *D. magna* fecundity. Furthermore, we identified 843 differentially expressed genes (DEGs) through RNA sequencing, using the groups that demonstrated a decrease in fecundity. To predict the adverse effects of chronic butylparaben exposure to humans, we performed BLAST analysis on the daphnia DEGs and we identified 87 homologous human genes. Through network analysis, it was predicted that various diseases such as neurogenetic, male urogenital, and endocrine disorders and *SRS* and *POU*, would be selected to as potential biomarkers for human exposure to butylparaben. Our integrative approach allowed for the identification of adverse effects of following long-term exposure to butylparaben in humans and provided a perspective of identifying genes that should be considered as potential biomarkers for environmental disease following long-term exposure to a chemical.

Session Title: Omics Technologies Poster Session II

PB3410 Integrative network analysis identifies miRNA signatures for neuropathology and cognition in Alzheimer's disease

Authors:

Y. Park[1], S-W. Han[2], J-M. Pyun[3], K. Nho[4]; [1]Seoul Natl. Univ. Bundang Hosp., Seongnam-si, Korea, Republic of, [2]Chuncheon Sacred Heart Hosp., Chuncheon-si, Korea, Republic of, [3]Soonchunhyang Univ. Seoul Hosp., Seoul-si, Korea, Republic of, [4]Indiana Univ Sch. of Med., Indianapolis, IN

Background The exact mechanisms underlying the pathogenesis of Alzheimer’s disease (AD) remain unclear due to the involvement of complex neurochemical and genetic factors. MicroRNA (miRNA) dysregulation is a potential mechanism contributing to gene expression changes in AD. Integrative network analysis based on the correlation of miRNA profiles can provide insights into the pathogenesis of AD.Methods We performed miRNA co-expression network analysis to identify network modules associated with AD, neuropathology markers, and cognition and their hub miRNAs using brain tissue miRNA profiles from a large longitudinal study of aging cohort (N=702) as a discovery dataset. We also performed association analysis of the hub miRNAs with AD, neuropathology markers, and cognition. After selecting target genes of the hub miRNAs, we performed association analysis of the hub miRNAs with their target genes and then performed pathway-based enrichment analysis. Furthermore, we performed differential expression analysis of target genes using brain tissue RNA-Seq data. For replication analysis, we performed a consensus miRNA co-expression network analysis using the ROSMAP dataset and an independent miRNA profile dataset (N=16) from the Gene Expression Omnibus (GEO). Finally, we performed a machine learning approach using penalized logistic regression to assess the performance of hub miRNAs for the classification of AD.ResultsNetwork analysis identified a module (M3) as significantly associated with AD, amyloidopathy, and cognition. Five hub miRNAs (miR-129-5p, miR-433, miR-1260, miR-200a, and miR-221) of M3 exhibited significant associations with AD, neuropathology markers, and/or cognition. Gene-set enrichment analysis of target genes associated with their corresponding hub miRNAs identified significantly enriched biological pathways including ErbB, AMPK, MAPK, and mTOR signaling pathways. Consensus network analysis identified two consensus network modules as significantly associated with AD in ROS/MAP and marginally associated with AD in GEO. Notably, consensus network analysis also identified miR-129-5p and miR-221 as hub miRNAs. Machine learning analysis showed that the AD classification performance (area under the curve (AUC)=0.847) of age, sex, and *APOE* ε4 carrier status was improved by 4.2% with inclusion of five AD-associated hub miRNAs.Conclusion An integrative network approach identified miRNA signatures as associated with AD, neuropathology markers, and cognition, enhancing our understanding of AD pathogenesis and leading to better performance of AD classification as potential diagnostic biomarkers.

Session Title: Omics Technologies Poster Session III

PB3411 Interpretable Multi-View Integrative Approaches Imputing Serum Short-Chain Fatty Acids from Gut Microbiome

Authors:

A. Liu[1], B. Tian[2], K-J. Su[1], L. Jiang[1], C. Zhao[3], M. Song[4], Y. Liu[5], G. Qu[1], Z. Zhou[1], X. Zhang[1], Y. Gong[1], C. Qiu[1], Z. Luo[1], Q. Tian[1], H. Shen[1], Z. Ding[1], H-W. Deng[1]; [1]Tulane Univ., New Orleans, LA, [2]Central South Univ., Changsha, China, [3]Michigan Technological Univ., Houghton, MI, [4]Univ. of Southern Mississippi, Hattiesburg, MS, [5]Sch. of Basic Med. Sci., Central South Univ., Changsha, China

Short-chain fatty acids (SCFAs) are the main metabolites produced by bacterial fermentation of dietary fiber within the gastrointestinal tract. The SCFAs produced by the gut microbiota (GM) are absorbed by the host, reach the bloodstream, and are distributed to different organs, thus influencing host physiology. Recent deep learning approaches have paved the way for innovative predictive models being capable of analyzing complex gut microbiome data to provide insights into disease risk, metabolic functionality, and other microbiome-associated outcomes, offering new opportunities for personalized medicine. However, there are insufficient explorations on predictive models that can accurately link the GM composition and the actual production of serum SCFAs in the host. Most existing predictive models (e.g., disease prediction model) do not fully account for the complex interplay among host characteristics (race, age, exercise, etc.), dietary habits, and GM composition. Here, we developed an innovative multi-task multi-view integrative approach (M^2AE, Multi-task Multi-View Attentive Encoders) to predict the abundances of serum SCFAs by capturing the intricate interplay among three views, i.e., gut microbiomes, dietary habits, and host characteristics. This method jointly explores view-specific representation and cross-view correlation for effective prediction. We demonstrated that this method outperforms traditional regression-based approaches (Isobutyric acid: Spearman correlation coefficient (SCC) = 0.327, p value < 0.05; Hexanoic acid: SCC = 0.308, p value< 0.05). The host characteristics, dietary habits, GM composition, and serum SCFAs data are from 964 males in our own Trans-omics Integration of Multi-omics Studies for Male Osteoporosis Study. Furthermore, this method identified that the bacterial species Bacteroides (stercoris, oleiciplenus), Prevotella (copri), and Flavonifractor (plautii) showed a more significant contribution of gut microbiome in predicting serum SCFA profiles. To further detect the complex associations between GM composition and SCFAs, we also further constructed the GM interplay networks for high SCFAs and low SCFAs groups, respectively. To summarize, the microbial association network analysis revealed the complex interplay of GMs for high SCFAs and low SCFAs groups. Our method and findings can enhance our comprehension of the complex biological processes involved in SCFA production, offering fresh perspectives on how dietary habits and GM composition contribute to this process.

Session Title: Omics Technologies Poster Session I

PB3412 Investigating deep learning nuclei segmentation of H&E images for bulk RNA sequencing expression deconvolution.

Authors:

J. Balan, S. K. McDonnell, Z. Fogarty, N. B. Larson; Mayo Clinic, Rochester, MN

Recent developments in single-cell RNA sequencing (scRNA-seq) technologies facilitate characterization of transcriptomic patterns for individual cell types. However, scRNA-seq remains costly, requires high-quality tissue samples, and presents numerous analytical challenges. With large bulk RNA-seq datasets available for various tissues and conditions, new algorithmic approaches have emerged for expression deconvolution of bulk RNA-seq using well-defined scRNA-seq reference panels. However, these panels may not be available on relevant cell types and/or fully capture expression profile variability. With the common availability of pathological H&E-stained slides for tissue-based bulk RNA-seq and the ease of slide digitization, an alternative approach is to characterize and quantify cell types using deep learning (DL). In this study, we explored the integration of bulk RNA-seq expression data with DL-based cell proportion estimates from digitized pathological H&E slides in 120 normal prostate tissue samples.

Bulk RNA-seq gene expression was quantified using an HTSeq and normalized via CQN (24,075 total genes). Corresponding H&E slides were manually annotated by pathologist to provide crude estimates of percent epithelium (deciles; range = 40-80%) and presence of infiltrating lymphocytes. The slides were then digitized using the Aperio Leica GT-450 digital pathology scanner platform. The digitized slides in SVS format were analyzed using a DL model (Hovernet) to perform nuclei segmentation and quantify epithelial, neoplastic, inflammatory, connective, dead, non-neoplastic epithelial, lymphocyte, macrophage, and neutrophils. We then performed gene expression variance decomposition using linear mixed models to quantify explained variation by estimated cell-type proportions. Hovernet segmentation demonstrated overall agreement with manual epithelium estimates (Spearman corr. = 0.392, P<1e-05) while additionally providing single-cell resolution estimates. Expression variance decomposition analyses yielded comparable findings between the two cell-type proportion methods, with the manual and DL-based methods explaining >10% of total variation in 4079 (8.4%) and 5051 (10.5%) of total genes, respectively. Furthermore, DL-based methods tended to explain more expression variability on a per gene basis. The subset of 163 genes with high explained variation by epithelial proportion (>30%) were significantly enriched for relevant biological pathways. These results further highlight the potential of augmenting bulk RNA-Seq with H&E slide digital pathology for innovative multi-omics analyses.

Session Title: Omics Technologies Poster Session II

PB3413 Isolate and neaten VNTRs (InVNTR): A command-line tool for VNTR analysis and visualization.

Authors:

E. Kaufman, P. N. Valdmanis; Univ of Washington, Seattle, WA

Variable number tandem repeat (VNTR) length and internal repeat structure are associated with several disorders, including neurodegenerative disease. As a greater number of long-read genomes with accurate resolution of VNTRs become available, a suite of tools to investigate VNTR variation and structure at a high throughput are needed. We developed Isolate and neaten VNTRs or InVNTR as a command-line tool for rapidly parsing and analyzing VNTRs in whole genomes. It can be used to identify repeat expansions or contractions, as well as variation in repeat structure. InVNTR simplifies the process of creating data visualizations like Seattle plots, allowing for simple repeat visualization, either delimiting by sub-motif or by length of repeat motif. It requires unique sequence directly before and/or after the VNTR, and repeat length or a sub-motif by which to split it by. InVNTR has shown its utility for analysis of VNTR repeat allele length, accurately identifying individuals and populations with repeat expansions in both exons and introns in publicly available repositories of long read genomes, such as from the Human Pangenome Reference Consortium and the Human Genome Structural Variation Consortium.

As proof of principle, we utilized InVNTR to identify repeat sequence and structure in a series of the largest coding VNTRs in the human genome. We identified 12 VNTRs with variability in allele length in mucin genes, of which the pattern of repeat expansion could be identified in 11. A 34bp VNTR in *MUC2* has a 50-fold dynamic range in repeat length, while a VNTR in *MUC12* with a repeat unit of 84bp is more than 77 times larger in some individuals than others, corresponding to a nearly 35kb expansion event within an exon. We leveraged short read WGS data from the 1000 Genomes Project and found high concordance between short and long-read repeat length estimates. From this analysis we found a significant ~1.5-fold increase in average repeat length for *MUC2* specifically in samples of East Asian ancestry relative to the four other superpopulations ($p<0.0001$ by one-way ANOVA followed by Tukey's multiple comparison test), perhaps reflecting selection for increased repeat length. Collectively InVNTR is a powerful tool to identify and visualize human variation in VNTRs.

Session Title: Omics Technologies Poster Session III

PB3414 Large-scale Circulating Proteome Association Study (CPAS) and its use for improved hip fracture prediction

Authors:

E. Coward[1], T. Austin[2], H. A. Fink[3], D. I. Jalal[4,5], A. E. Törnqvist[6], P. Buzkova[2], J. I. Barzilay[7], T. Lu[8,9,10], L. D. Carbone[11,12], M. Gabrielsen[13], L. Grahnemo[6], K. Hveem[13], C. Jonasson[14], J. Kizer[15,16], A. Langhammer[1,17], K. J. Mukamal[18], M. Nethander[6], B. M. Psaty[2,5], J. A. Robbins[19], Y. Sun[20], A. Skogholt[13], B. ÅSvold[1], R. J. Valderrabano[21], J. Zheng[22], B. Richards[9,8,10,23], C. Ohlsson[6,24]; [1]Norwegian Univ. of Sci. and Technology, Trondheim, Norway, [2]Univ. of Washington, Seattle, WA, [3]Geriatric Res. Ed. and Clinical Ctr., Minneapolis, MN, [4]Carver Coll. of Med., Iowa City, IA, [5]Iowa City VA Med. Ctr., Iowa City, IA, [6]Univ. of Gothenburg, Gothenburg, Sweden, [7]Kaiser Permanente of Georgia, Atlanta, GA, [8]Jewish Gen. Hosp., Montreal, QC, Canada, [9]McGill Univ., Montreal, QC, Canada, [10]5 Prime Sci. Inc, Montreal, QC, Canada, [11]Charlie Norwood VAMC, Augusta, GA, [12]Augusta Univ., Augusta, GA, [13]NTNU, Trondheim, Norway, [14]Norwegian Univ. Sci. and Technology, Hoevik, Norway, [15]Univ. of California San Fransisco, San Fransisco, CA, [16]San Fransisco VA Hlth.Care System, San Fransisco, CA, [17]Levanger Hosp., Levanger, Norway, [18]Beth Israel Deaconess Med. Ctr., Brookline, MA, [19]Univ. of California, Davis, CA, [20]Emory Univ., Atlanta, GA, [21]Harvard Med. Sch., Boston, MA, [22]Univ. of Bristol, Bristol, United Kingdom, [23]King's Coll., London, United Kingdom, [24]Sahlgrenska Univ. Hosp., Gothenburg, Sweden

Hip fractures are associated with significant disability, high cost, and mortality. However, the exact biological mechanisms underlying hip fractures remain incompletely understood. In an exploratory search of the underlying biology as reflected through the proteome, we performed a comprehensive Circulating Proteome Association Study (CPAS) meta-analysis on incident hip fractures.

Analyses included 6,430 subjects from two prospective cohort studies (Cardiovascular Health Study [CHS] and HUNT Study) with available circulating proteomics data (aptamer-based 5K SomaScan version 4.0 assay; 4,979 aptamers). Associations between circulating protein levels and incident hip fractures were calculated for each cohort using age and sex-adjusted Cox regression models.

Participants experienced 643 incident hip fractures. Inverse-variance weighted meta-analyses yielded more significant associations than either individual study and identified 23 aptamers that were significantly associated with incident hip fractures (conservative Bonferroni correction 0.05/4,979, $P<1.0\text{x}10^{-5}$). The most strongly associated aptamers with hip fracture risk corresponded to two proteins of the growth hormone/insulin growth factor system (GHR and IGFBP2) as well as GDF15 and EGFR. High levels of several inflammation-related proteins (CD14, CXCL12, MMP12, ITIH3) were significantly associated with increased hip fracture risk.

These analyses identified several circulating proteins and pathways consistently associated with incident hip fractures. These findings underscore the usefulness of the meta-analytic approach for comprehensive CPAS in a similar manner as has previously been observed for large scale human genetic studies.

Based on the CPAS results in CHS only, we developed a protein-based risk score for hip fracture prediction including 18 proteins. The performance in predicting incident hip fractures was determined in the independent HUNT validation cohort, as well as in a separate HUNT cohort with 1,988 subjects with proteomics data from the 7K SomaScan assay. The performance was better than existing genetic risk scores for bone mineral density and fractures, and it improved fracture discrimination and reclassification significantly in the independent validation data sets. Thus, proteomics may add clinically useful information for hip fracture prediction.

Session Title: Omics Technologies Poster Session I

PB3415 Learning transcriptional signatures from high-dimensional data: Creating a set of spatially-resolved patterns of gene coexpression in the mouse cortex.

Authors:

K. Woyshner, Y. Wang, K. Hansen, G. Stein-O'Brien, L. Goff; Johns Hopkins Univ., Baltimore, MD

With the development of spatial transcriptomic technologies, researchers are able to profile gene expression across tissues of interest, retaining information about the location of individual genes, cell types, and features. This opens up the potential for more comprehensive profiling of complex and intricate structures such as the mammalian brain. However, meaningful interpretation of these high-dimensional data remains difficult. Non-negative matrix factorization is a machine learning algorithm that can reduce high-dimensional data into a meaningful set of latent space factors. When applied to transcriptomic data, these factors capture shared and distinct biological and cellular processes. We have adopted non-negative matrix factorization to learn a set of spatially-resolved patterns of gene coexpression in the mouse brain to better understand the diversity of transcriptional signatures across different brain regions and to establish a baseline representation of biological features and cellular processes engaged across the mammalian brain. Applying this technique to the Allen Brain Atlas in situ hybridization dataset of adult mouse brain, we have identified spatially resolved patterns that delineate distinct features of the mouse cortex. Many factors correlate strongly with known anatomical regions, while others capture shared spatial features that do not correlate with a single structure.

By projecting external gene expression datasets into this learned latent space catalog, we have the ability to predict the spatial extent of additional features such as disease-relevant signatures, and areas of shared developmental- and degeneration-associated regulatory networks to elucidate the spatial effect based on transcriptional changes. Defining a catalog of spatially informed transcriptional signatures in a healthy adult mouse brain would provide a more comprehensive understanding of the composition and organization of the mammalian brain and serve as a foundational resource for exploring the spatial extent of transcriptional changes involved in disease.

Session Title: Omics Technologies Poster Session II

PB3416 Leveraging Clinical Intuition to Improve Accuracy of Phenotype-Driven Prioritization

Authors:

M. Beckwith[1], D. Danis[2], J. Jacobsen[3], D. Smedley[3], C. Boerkoel[4], P. Robinson[5]; [1]The Jackson Lab. for Genomic Med., Famington, CT, [2]The Jackson Lab. for Genomic Med., Farmington, CT, [3]Queen Mary Univ. of London, Cambridge, United Kingdom, [4]Univ. of British Columbia, Vancouver, BC, Canada, [5]The Jackson Lab., Farmington, CT

Next-generation sequencing technologies (NGS), including exome sequencing and genome sequencing, have had dramatic effects on the cost, accuracy, and utility of genetic testing. A major focus of translational genomics research is to improve the performance of NGS-based diagnostics in rare disease. Efforts have focused on improving the performance of phenotype-driven prioritization, which leverages a comparison of the phenotypic abnormalities of the individual being investigated with those previously associated with human diseases or genetically modified model organisms to rank candidate variants. To our knowledge, no available algorithm explicitly models clinical intuition. However, the clinical assessment often involves the identification of the general group of diseases, such as non-syndromic intellectual disability and epileptic encephalopathy. We recently introduced LIRICAL, a likelihood-ratio based approach for phenotype-driven prioritization, in which the pretest probability of diseases was modeled uniformly. Here, we introduce L4CI, which extends the LIRICAL algorithm to model clinical intuition, exploiting the ontological representation of diseases of the Mondo ontology. For example, increasing the pretest probability adjustment value for metaphyseal chondrodysplasia-retinitis pigmentosa syndrome from 0, which corresponds to the uniform model used in LIRICAL, to 20 increases the disease rank from 52 to 1. Therefore, adjusting the pretest probability of a group of diseases that reflect clinical intuition can improve the ranking of candidates with the LIRICAL algorithm. Additionally, L4CI can be used with a group of genes associated with diseases that are commonly seen in specialty clinics for disease groups such as skeletal dysplasia. We provide a desktop application with graphical user interface that can be used to run the L4CI algorithm and produce an HTML output file.

Session Title: Omics Technologies Poster Session III

PB3417 Leveraging Existing DNA Sequence-to-Expression Models for Enhanced Single-Cell Gene Expression Prediction

Authors:

Y. Zhou[1], T. Adeluwa[1], S. Gona[1], L. Zhu[2], M. Chen[1], H. Im[3]; [1]The Univ. of Chicago, Chicago, IL, [2]The Univ. of Chicago, chicag, IL, [3]Univ. of Chicago, Chicago, IL

Advancements in single-cell RNA sequencing (scRNA-seq) technologies have enabled the characterization of cellular heterogeneity and the identification of cell-type specific gene expression patterns. However, accurately predicting gene expression at the cell-type level remains a challenge. To address this problem, we developed Single-cell PrediXcan, a multitask model specifically designed for multi-cell-type gene expression prediction. Our model adopts a transfer learning approach, making use of the cutting-edge 'enformer' DNA-to-epigenetics deep learning model (Avsec, Ž., et al. 2021) and incorporating epigenetic features inferred from DNA sequences to improve prediction accuracy. Most existing gene expression prediction methods operate at the tissue level, lacking the resolution to capture cell-type specific variations. Existing methods designed for cell-type level prediction do not utilize epigenetic features during training and consequently demonstrate suboptimal performance. In contrast, Single-cell PrediXcan focuses on training a neural network model on scRNA-seq datasets with predicted epigenetic features, enabling accurate gene expression prediction in different cell types.

Here we show that our Single-cell PrediXcan model trained on Enformer predicted epigenetic features with single-cell RNAseq data from the OneK1K dataset (Yazar, S., et al. 2022. Science) achieves a correlation of 0.85 across all the cell types in the dataset between measured and predicted expressions. These results highlight the potential to generate single-cell prediction models for diverse contexts, advancing the development of our proposed method, single-cell prediXcan.

Moreover, Single-cell PrediXcan paves a way for transcriptome-wide association studies (TWAS) at the cell-type or cell-state level, which could provide new insights into the molecular mechanisms underlying complex traits and diseases.

Session Title: Omics Technologies Poster Session I

PB3418 Leveraging multiomics data to improve diagnostic yield: A prospective study of 29 familial cases.

Authors:

A. Crawford[1], J. Albay[1], D. Hernandez[1], R. Wang[1], Q. Zhang[1], E. Chekalin[1], Y. Zhuang[1], C. Glidewell-Kenney[1], J. Taylor[1], A. Clause[1], A. Malhotra[1], A. Coffey[1], A. Chandrasekhar[1], K. Bluske[1], K. Robinson[1], R. Bombardi[1], I. Khrebtukova[1], E. Puffenberger[2], K. Brigatti[2], D. Basel[3], J. McCarrier[3], B. Sadikovic[4], J. Kerkhof[4], A. Kennedy[5], K. DeBalsi[5], R. Taft[1]; [1]Illumina Inc., San Diego, CA, [2]Clinic for Special Children, Strasburg, PA, [3]Med. Coll. of Wisconsin, Milwaukee, WI, [4]Verspeeten Clinical Genome Ctr., London Hlth.Sci., London, ON, Canada, [5]Metabolon, Durham, NC

Background

Undiagnosed rare diseases affect at least 200 million individuals worldwide. Although whole exome sequencing (WES) and whole genome sequencing (WGS) have improved the ability to rapidly diagnose many of these patients, approximately 50% of cases remain without a definitive molecular diagnosis. This study aimed to investigate the potential of multiomic assays to narrow the diagnostic gap across a broad spectrum of clinical presentations suggestive of a genetic disease.

Methods

We conducted a prospective, observational study encompassing 29 consenting families including probands, biological parents, and available unaffected siblings. Study participants were recruited based on a suspected genetic origin of their clinical presentation and participation of at least two unaffected family members. Five families had previously received a molecular diagnosis and served as positive controls, whilst the remaining 24 families had not received a molecular diagnosis from previous whole genome sequencing. Probands presented with a diverse range of phenotypes across multiple systems and spanned in age from 2 to 43 years. We performed a set of multiomic assays on blood samples including both long and short read whole genome sequencing, deep whole genome sequencing, methylation profiling, plasma proteome analysis, bulk T cell and B cell immune repertoire sequencing, whole transcriptome sequencing, and metabolomic profiling. Due to the limited sample size and lack of publicly available datasets, most of the analyses were conducted in a one-versus-all approach, comparing the proband results to the cohort.

Results

Systematic analysis resolved 9 of 24 previously undiagnosed cases. This included the identification of three novel gene-disease relationships, three cases solved using short read WGS data, two cases with whole transcriptome analysis, and two cases through detection of disease-associated whole epigenome signatures using EpiSign™. The genetic findings in at least two cases did not explain the entirety of the participant's phenotypes and were thus considered only partially resolved. Supporting evidence for the molecular findings was observed in multiple multiomic assays, including plasma proteomics and metabolomics, providing additional confirmation of the reported results in both positive controls and the newly solved cases.

Conclusion

The integration of multiomic analyses demonstrates great potential to increase the diagnostic rate for patients with undiagnosed disease. Sequence data from these pilot cases will be made available for further research in partnership with the NHGRI GREGoR Consortium.

Session Title: Omics Technologies Poster Session II

PB3419 Localized spatial dimensionality reduction on spatial transcriptomics

Authors:

J. Chen, Y. Li, D. Li; Univ. of North Carolina at Chapel Hill, Chapel Hill, NC

Dimension reduction has become a crucial technique in the analysis of omics data, particularly due to its high dimensionality, such as gene expression measurement, which often involves more than 10,000 genes. However, the selection of genes for analysis remains an open question, and no or minimal selection can be computationally prohibitive. Therefore, incorporating dimension reduction is essential in such analyses. In the field of transcriptional studies, spatial transcriptomics has gained popularity due to its ability to capture gene expression while preserving spatial coordinates. It has been observed that spots or cells in close spatial proximity tend to exhibit similar gene expression patterns. Consequently, when performing dimension reduction, considering the spatial location may enhance reduction efficiency. In this study, we propose a novel method, Localized Spatial Dimensionality Reduction (LSDR), for dimension reduction while incorporating spatial coordinates. For each spot or cell, we leverage neighboring spots or cells to perform PCA analysis. Additionally, we introduce a penalty term to ensure that the dimension reduction matrix reflects the spatial distance between spots or cells. Through simulation studies and real data analysis, we demonstrate the advantages of our LSDR approach. By obtaining a dimension reduction matrix corresponding to each spatial coordinate, we can leverage transfer learning to apply this matrix to other slides in a common coordinate space without the need for retraining the model. Furthermore, utilizing the local dimension reduction matrix demonstrates better performance on the application of clustering methods to detect biologically meaningful spatial domains.

Session Title: Omics Technologies Poster Session III

PB3420 Long read RNA-sequencing demarcates cis and trans-directed alternative RNA splicing.

Authors:

G. Quinones-Valdez[1], K. Amoah[1], X. Xiao[2]; [1]UCLA, Los Angeles, CA, [2]Univ. of California, Los Angeles, Los Angeles, CA

RNA splicing is largely determined by the contribution of both *cis*-acting genomic elements and *trans*-acting splicing factors. Genetic variants may disrupt *cis*-regulatory motifs and alter splicing, constituting an important link between genotypes and phenotypes. Determining splicing events that are directed primarily by the *cis*- or *trans*-acting mechanisms will greatly inform our understanding of the genetic basis of disease. Here, we leverage long-read RNA-seq to demarcate splicing events that are primarily *cis*- or *trans*-directed, using a new method, isoLASER, that performs *de novo* variant calling, gene-level phasing and allele-specific splicing analysis for long reads. Using ENCODE data from human and mouse samples, we identified 2,807 and 5,322 unique exonic parts, respectively, exhibiting *cis*-directed splicing. Our findings demonstrate that the inclusion of most alternatively spliced exons is predominantly regulated by tissue-specific factors, while the splicing of *cis*-regulated exonic parts remains largely consistent across tissues with a shared genetic background. Additionally, we observed that genes under *cis*-directed splicing are frequently associated with immune function, and the novel exons resulting from *cis*-regulation significantly overlap with SINEs, LINEs, and simple repeat elements known to shape immune response. Leveraging isoLASER's joint analysis function, we analyzed pooled samples from the same tissue type but with different genetic backgrounds to identify likely functional variants. Remarkably, the putative functional variants exhibit significant alterations in RNA-binding protein (RBP) binding and represent a subset of splicing quantitative trait loci (sQTLs) enriched with highly significant p-values, validating the functionality of these candidate variants. Furthermore, we analyzed *cis*- and *trans*-directed splicing of brain samples from patients with Alzheimer's disease (AD) and healthy controls. Our analysis uncovered many cis-directed events, some of which reside in genes of close AD relevance. We highlight a notable difference in the genetic modulation of *HLA-A* gene splicing between AD and controls, which was further confirmed through sQTL analysis using short-read data from the Mount Sinai Brain Bank Consortium. In summary, this study establishes a clear demarcation between genetically and non-genetically driven splicing. We also provide a framework for identifying such events and splicing-associated variants in cohorts with a limited sample size.

Session Title: Omics Technologies Poster Session I

PB3421 Long read single cell whole genome sequencing from human brains

Authors:

C. Proukakis[1], E. Kalef-Ezra[1], D. Horner[1], M. izydorczyk[2], M. Toffoli[1], N. Holmes[3], S. Scholz[4], F. Sedlazeck[2]; [1]Univ. Coll. London, London, United Kingdom, [2]Baylor Coll. of Med., Houston, TX, [3]Univ. of Nottingham, Nottingham, United Kingdom, [4]NIH, Bethesda, MD

Somatic mutations occur in human brains, and may have a role in neurodegenerative disorders. Their study is challenging because of the low allelic fraction, which may necessitate sequencing of single cell DNA. This is preceded by whole genome amplification, which is possible by a number of methods. Multiple displacement amplification (MDA), performed isothermally by the phi-29 enzyme, yields long amplicons (>10 kb), making the detailed study of all classes of somatic structural variants (SV) by long reads theoretically possible. It results, however, in incomplete and uneven amplification, which could be improved by dividing the genome into droplets on the X-Drop (Samplix), as well as complex chimeras.Multiple system atrophy (MSA) is a rare synucleinopathy, a sporadic neurodegenerative disorder similar to Parkinson's disease. We previously demonstrated somatic CNVs (gains) of the disease-relevant alpha-synuclein gene using FISH, and large CNVs in almost a third of cells from two brains using low coverage Illumina single cell whole genome sequencing (scWGS). We have selected these two MSA brains for further study by long read scWGS. We prepared nuclear preparations from the cingulate cortex, and one control brain cortex, and isolated single nuclei using the CellRaft device (Cell Microsystems), followed by droplet MDA. We performed Nanopore sequencing of 6 cells from each brain, using two library prep protocols: one digesting MDA side-chains with T7 endonuclease, and a PCR-based one. Up to 6 cells were multiplexed per flow cell, with maximum yields of 25 Gb for MinION R9.4 (PCR protocol, N50 ~2.8 kb) and 115 Gb for Promethion R10.4 (T7 protocol, N50 ~3 kb). Sniffles2 analysis allows detection of Alu insertions and other SV in areas adequately covered, after filtering chimeras which manifest as inversions. As a comparison, we studied 5 cells from one MSA brain amplified with Primary Template Amplification (BioSkryB), which also uses phi-29, but the addition of "stop" nucleotides leads to shorter amplicons and more complete and even genome coverage. MinION sequencing (3.3 Gb yield) led to much lower N50 as expected (0.77 kb), but no inversions. Further filtering of single cell dMDA SV calls and comparison to bulk Nanopore (55x) and Illumina (85x) data is underway.

Session Title: Omics Technologies Poster Session II

PB3422 Long-read RNA sequencing identified the different gene isoforms in different brain regions and their relationship with DNA methylation.

Authors:

M. Shimada[1], Y. Omae[1], A. Kakita[2], R. Gabdulkhaev[2], T. Miyagawa[3], M. Honda[3], A. Fujimoto[4], K. Tokunaga[1]; [1]Natl. Ctr. for Global Hlth.and Med., Tokyo, Japan, [2]Brain Res. Inst., Niigata Univ., Niigata, Japan, [3]Tokyo Metropolitan Inst. of Med. Sci., Tokyo, Japan, [4]Univ. of Tokyo, Tokyo, Japan

Background: In neuropsychiatric disorders, site-specificity in the brain has been suggested and differences in gene expression could potentially explain this mechanism. However, long-read transcriptome analysis of the human brain is very limited, and none has focused on brain region-specific gene expression. In this study, we performed long-read RNA sequencing on 12 samples that were derived from three different brain regions of the same individuals: the cerebellum, hypothalamus, and temporal cortex. We then compared the differences in expression profiles between regions. **Results:** We identified numerous novel isoforms; approximately 40% of the detected isoforms were novel, some of which were confirmed by publicly available LC-MS/MS data of the brain cortex. We found that, compared to other regions, many genes with higher expression levels in the cerebellum and temporal cortex were associated with neuronal pathways, whereas those with higher expression levels in the hypothalamus were primarily linked to immune pathways. In the analysis of genes expressed in high numbers of isoforms in each brain region, immune-related genes were detected only in the hypothalamus as well, although the correlation between the number of isoforms and expression levels per gene was not strong. In addition, we investigated genes with different major isoforms in each brain region, even with similar overall expression levels among regions. As a result, *GAS7* was found to show the most significant difference in the major expressed isoforms among the three brain regions and the region-specific *GAS7* isoforms were confirmed by western blotting. Many of the genes with different major isoforms among regions were found to be involved in "actin filament-based process" and "cell projection organization" pathways, suggesting that region-dependent isoforms may have distinct roles in dendritic spine and neuronal formation in each region. Furthermore, we investigated the involvement of DNA methylation in these isoforms differentially expressed in each region. We found that there are significantly more CpG sites differentially methylated around genes that express different major isoforms among regions, especially in the region within 1 kb downstream from the transcription start sites between isoforms that have different TSSs between regions. **Conclusions:** Our results provide potentially valuable findings for future research on brain disorders and shed light on the mechanisms underlying isoform diversity in the human brain.

Session Title: Omics Technologies Poster Session III

PB3423 Long-read RNA sequencing identifies consistency amongst tissues and laboratories for RNA-isoforms of known genes and *de novo* gene bodies.

Authors:

L. Gordon[1], M. Page[2], B. Aguzzoli Heberle[1], J. Brandon[1], K. Nations[1], M. Wadsworth[3], S. Goodwin[4], E. Ghiban[5], R. Wappel[5], S. Mavruk Eskipehlivan[5], P. Nelson[1], J. Miller[1], M. Ebbert[1], J. Fryer[6]; [1]Univ. of Kentucky, Lexington, KY, [2]Univeristy of Kentucky, Lexington, KY, [3]Univ. of Kentucky, lexington, KY, [4]Cold Spring Harbor Lab., Cold Spring Harbor, NY, [5]Cold Spring Harbor Lab, Cold Spring Harbor, NY, [6]Mayo Clinic, Jacksonville, FL

Intro: Traditionally, short-read sequencing is used for RNASeq studies that, due to technical limitations, forces researchers to collapse all RNA isoforms for a given gene into a single value—ultimately hiding the diversity and variation of individual RNA isoforms. Long-read sequencing, however, can sequence an entire RNA molecule in a single read, enabling researchers to identify and quantify all RNA isoforms from known genes and new gene bodies. Since long-read sequencing is relatively new, there have been few studies characterizing and quantifying individual RNA isoforms across body tissues. Here, we compare our deep long-read human brain RNASeq to the recently released GTEx data by Glinos et al. and demonstrate the reliability of quantifying isoform expression, and the expression patterns for newly discovered genes and isoforms across tissue types. **Methods:** We sequenced 12 frontal cortex post-mortem aged samples (Brodmann area 9/46; 50% female) using 1 Oxford Nanopore PromethION flow cell/sample. We quantified expression of new isoforms in 9 tissue types from GTEx data. GTEx data were sequenced using the same protocol, except using 1 MinION flow cell/sample. Analysis included pychopper, mimimap2, Bambu, and Ensembl annotations for GRCh38 v107 (July 2022). **Results:** In our deep brain sequencing, we identified 1534 new isoforms from 1255 known genes and 1861 isoforms from 1677 new gene bodies (relative to GRCh38 v107), where 431 (28%) and 269 (14%) of these isoforms are high-confidence and expressed at median CPM (counts per million) > 1, respectively. Limiting to only medically relevant genes, we identified 54 (45%) new isoforms expressed at median CPM > 1. In the 9 GTEx tissues, 10% to 32% of new isoforms from known genes were present with median CPM > 1, notably 32% in the same brain area as our data. The percentage of new isoforms from new gene bodies with a CPM > 1 fell between 0.9% and 6% (6% in same brain region). New isoforms from medically relevant genes increased to between 20% and 43% (42% in same brain region). Additional statistics will be included during presentation. **Conclusions:** Properly characterizing and quantifying individual RNA isoforms across all genes is essential to truly understanding human health and disease. Here, we identified hundreds of new high-confidence RNA isoforms since GRCh38 v107 (July 2022), and quantify their expression across human tissues—allowing us to begin discerning individual RNA isoform function. We also demonstrate that long-read sequencing gives consistent results for the same tissue across different samples and laboratories.

Session Title: Omics Technologies Poster Session I

PB3424 Long-read sequencing reveals transcriptome variations in a right ventricle pressure overload Porcine model administered autologous umbilical cord blood-derived mononuclear stem cell therapy.

Authors:

N. Odogwu, C. Hagen, S. Cantero, B. Rasmussen, S. Oommen, T. Nelson, Wanek Pre-clinical Pipeline; Mayo Clinic, Rochester, MN

Background The genetic complexity, variable phenotypic manifestations, familial inheritance patterns, and a wide variety of genes that characterize congenital heart disorders (CHDs) including hypoplastic left heart syndrome make them difficult to understand. Stem cell therapy and their paracrine activity has been implicated in cardiac repair and restoration of heart function however, their mechanism is still under intense investigation. **Objective**To provide a mechanistic and functional understanding of how umbilical cord blood (UCB) derived mononuclear stem cells potentiate cardiac regeneration and improve cardiac function. We evaluated the effects of administering UCB-derived mononuclear stem cell therapy on failing heart transcriptome by synergistically combining filtering, and multiple multi-omics sequencing approaches to pinpoint candidate genes and gene variants that may affect cardiogenic pathways. **Method** A right ventricle pressure overload porcine model of heart failure was created by pulmonary arterial binding (PAB). Piglets were administered single and multidose umbilical cord mononuclear cells (UCB-MNCs) (3 million/kg) therapy on the right ventricle heart chamber and followed for 12 weeks. Chamber-specific transcriptome expression was assessed using Illumina RNA sequencing, and long-read RNA sequencing. Differentially expressed genes were validated by quantitative real-time PCR. **Results and Conclusion** Next-generation RNA sequencing, refined by pathway analysis, pinpointed distinct transcript genes across the right and left ventricles that are recognized for their role in cardiac pathophysiologies and clinical benefits. Importantly, *WT1, ACTN2*, were significantly upregulated in the left ventricle in pig-administered single-dose UCB-MNCs (3 million/kg) **(P_{WT1}, P_{ACTN2} < *0.05; Kruskal-Wallis*)**. *TBX 20* gene was highly upregulated in the multidose group (P_{TBX20} < 0.05; ***Kruskal-Wallis,*** multidose vs single dose). Long-read RNA-seq unmasked allelic changes and *cis* effects of genetic variants on transcripts. Given the role of genetic variants affecting transcript structure in disease risk, we resolve that a high-resolution characterization of the transcriptome with long-read data is crucial for the discovery of regulatory mechanisms of disease-associated genetic variants.

Session Title: Omics Technologies Poster Session II

PB3425 Low-pass whole genome sequencing with imputation provides an accurate and cost-effective alternative to array genotyping for GWAS and PRS approaches.

Authors:

A. Mulford[1], A. R. Sanders[1,2]; [1]NorthShore Univeristy Hlth.System, Evanston, IL, [2]Univ. of Chicago, Chicago, IL

Introduction. Reduced sequencing costs have prompted interest in low-coverage whole genome sequencing (lcWGS), with imputation, as an alternative to other genome-wide genotyping approaches. Previously reported comparisons have been done with simulation or downsampling approaches generally in European ancestry (EA) samples, or for direct empirical comparisons in much smaller N datasets.
Methods. DNAs from the Genomic Health Initiative (GHI), an institutional biobank, underwent lcWGS at an average 0.5× coverage. Generated paired-end FASTQ files underwent quality control (QC), and passing samples were imputed using the 1,000 Genomes Phase 3 haplotype reference panel. Variants from imputation of lcWGS data with a posterior probability of ≥0.90 were compared to previously generated array genotypes (UK Biobank Axiom Array, following manufacturer's recommended QC filters) for 5,783 individuals across an average of 640,000 variants using SnpSift and BCFtools.
Results. The average concordance for common (>5%), uncommon (1-5%), and rare (<1%) minor allele frequency (MAF) bins was high overall (0.989, 0.993, 0.995) with consistent concordance rates across autosomes and chromosome X, with very similar results for subsets self-identifying as non-Hispanic EA (N=4,305) or African ancestry (AA; N=326). Restricting comparisons to exclude homozygous reference matches showed decreased average concordance overall (0.976, 0.902, 0.934), again with similar results in EA and AA subsets. Genotype grids showed heterozygote mismatches as the most common type of discordance, as expected.
Conclusions. We found high genotype concordance rates overall and in various subsets (MAF bins, self-reported ancestry, non-reference genotypes) for two methods: lcWGS with imputation vs. array genotypes. Here, we demonstrate the applicability (i.e., high genotype concordance) in an institutional biobank setting with a much larger and more diverse sample population that previous reports. We will further present evidence of the utility of this genotyping method by describing replication of known genome-wide association study (GWAS) associations and polygenic risk score (PRS) findings for exemplar disorders and variants.

Session Title: Omics Technologies Poster Session III

PB3426 Machine learning and misalignment: Towards a comprehensive approach.

Authors:

D. D'Onofrio, J. Fu, B. N. Pusey, D. R. Adams; NIH, Bethesda, MD

INTRO: The NIH Undiagnosed Diseases Program (UDP) enrolls participants with significant illnesses that remain undiagnosed despite extensive standard of care clinical evaluation. Increasingly, individuals present with a history of a negative clinical exome or genome study. For this reason, the UDP utilizes genomic analysis pipelines optimized for detecting variants that may be missed by other analyses. This high-sensitivity approach generates many false-positive results. Methodologies for prioritizing results include the identification of short read misalignments. We hypothesize that currently available mapping and genotype scores do not capture all the available information for quantifying alignment quality for individual variants. To address this question, we are developing a machine learning-based tool to assist with ranking variants based on the quality of the alignment associated with short read variants. **METHODS:** We are generating a list of alignment characteristics (i.e., mapping quality, coverage, % of soft clipping, density of apparent variants, # of haplotypes), building on those incorporated into existing tools such as the GATK pipelines. These characteristics are being built into a random forest classifier from the python library scikit-learn. This model will be trained and tested with a combination of highly characterized genomes (Genome in a Bottle and Platinum Genomes), synthetic genomes, and a set of 7748 hand-curated variants from prior UDP evaluations. **RESULTS:** Initial assessment of potential model classifiers has demonstrated marked operator bias in hand-curated datasets. Evaluation of a heuristic alignment filtration system from a prior project suggests that some specific alignment patterns, such as >2 haplotypes covering the called variant, provide information that is not captured by traditional mapping or genotype score filters. **CONCLUSIONS:** We present preliminary data for in-progress work on a machine learning classifier designed to assist with the prioritization of results in noisy short read variant datasets. Our hope is that the work will prompt discussion and feedback that will be useful as tool development proceeds.

Session Title: Omics Technologies Poster Session I

PB3427 MAJIQ-CLIN: A novel tool for the identification of Mendelian disease-causing variants from RNA-seq data

Authors:

D. Issakova[1], J. K. Aicher[1], S. Jewell[1], G. R. Grant[1], N. F. Lahens[1], E. Bhoj[2], Y. Barash[1]; [1]Univ. of Pennsylvania, Philadelphia, PA, [2]Children`s Hosp. of Philadelphia, Philadelphia, PA

Exome sequencing (ES) is the current standard of care for patients with suspected Mendelian genetic disorders; however, the diagnostic rate is only 31%. Traditional ES analysis methods have difficulty incorporating differences in RNA splicing. Alternative RNA splicing is a highly regulated process in which different segments of pre-mRNA are joined, or spliced, together resulting in different mature mRNA transcripts. Alternative splicing occurs in over 90% of human genes, and changes in gene splicing can drastically alter the mRNA transcript and severely affect function of the mature protein. It is estimated that 15-50% of human pathogenic variants alter splicing, with several synonymous splice-altering variants being causal for known Mendelian disorders. It is therefore extremely crucial to develop reliable diagnostic tools to detect aberrations in RNA splicing from patient RNA-seq.
Previous studies have approached this problem, contrasting RNA-seq splicing variations found in patient samples to those found in healthy controls. However, tools developed for this task, such as FRASER and LeafCutterMD, have several limitations. Specifically, LeafCutterMD does not quantify intron retention, nor correct for the effect of confounding factors. In addition, LeafcutterMD and FRASER also require re-running many or all modeling steps from scratch upon the introduction of new data, which results in long run times that are impractical for clinical use.
We present MAJIQ-CLIN, a clinical pipeline developed using an updated version of our previously published MAJIQ pipeline for quantifying alternative splicing. Unlike FRASER and LeafCutterMD, MAJIQ-CLIN detects both unusually enriched outlier variants and entirely novel splice sites in patients. Like FRASER, MAJIQ-CLIN quantifies intron retention and corrects for the effect of known and unknown confounding factors. However, MAJIQ-CLIN can reuse quantifications of controls and learned models of unobserved confounders, requiring only fast updates of intermediate files when comparing a new sample against the controls. We assess MAJIQ-CLIN's accuracy using both synthetic data with spiked-in splice variations as well as several datasets of solved test cases, demonstrating it compares favorably to both LeafCutterMD and FRASER, and demonstrate its use in clinical cases.
In summary, we introduce a new method to accurately and effectively detect disease-causing splicing outliers in patients with suspected Mendelian disorders, with many advantages over currently available tools. By employing MAJIQ-CLIN we hope clinical geneticists will be able to improve molecular diagnosis rates and clinical care.

Session Title: Omics Technologies Poster Session II

PB3428 Massively parallel reporter assays and 3D chromatin structure in innate immune cells identify novel putative Alzheimer's Disease risk genes.

Authors:

M. Bond, I. Y. Quiroga, S. D'Costa, J. L. Bell, I. R. Sahasrabudhe, J. C. McAfee, K. S. Reed, S. Lee, M. Patrucco, H. Won, D. H. Phanstiel; Univ. of North Carolina, Chapel Hill, NC

Alzheimer's Disease (AD) is a progressive neurodegenerative disease that is the leading cause of dementia world-wide, affecting over 55 million people. Genome-wide association studies have identified 75 genomic loci associated with AD; however, interpreting these loci is difficult for multiple reasons. Multiple variants fall within linkage disequilibrium (LD) and are inherited together, making causal variants difficult to identify. Further, a majority of variants are located within the noncoding genome and it is unclear which genes they are affecting. Massively Parallel Reporter Assays (MPRA) allow us to identify putative active enhancers, yet it is also critical to perform MPRAs in relevant cell types and biological contexts. Therefore, we performed MPRAs to quantify the transcriptional activity of AD-associated variants in resting and activated human macrophages revealing hundreds of expression modifying variants. To understand these epigenetic changes and connect them to the genes that they regulate, we mapped H3K27ac (CUT&RUN), chromatin accessibility (ATAC-seq), gene expression (RNA-seq), and chromatin loops (Hi-C) in resting and activated iPSC-derived microglia. In total, we identified 20,992 chromatin loops across resting and activated iPSC-derived microglia. Among these, 9,544 loops linked 14,576 enhancers to 11,487 genes. 58 of these loops connected 414 AD-risk variants that were queried in our MPRA assays to 85 genes. 212 loops, 19,256 enhancers, and 2,790 genes changed in response to immune activation, highlighting the need to study these regulatory events in the correct biological context of disease. Continued analyses and data integration will reveal further mechanistic insight and identify novel AD risk genes that can be potentially targeted for therapeutic development. This intersection of MPRA results and dynamic regulatory networks with AD GWAS in a cell-type and context specific fashion has added to our understanding of the etiology underlying AD.

Session Title: Omics Technologies Poster Session III

PB3429 Metabolomic profile in Pre-eclampsia: a case control study in third trimester.

Authors:

C. Colmenares[1], P. Bautista-Niño[2], D. Quintero-Lesmes[2], C. Riaño[2], N. Serrano Díaz[2]; [1]Univ. del Valle. Fundación Cardiovascular de Colombia, Floridablanca, Colombia, [2]Fundación Cardiovascular de Colombia, Floridablanca, Colombia

Aim: Establish metabolomic profile associated with Pre-eclampsia and its severity (HELLP and early Pre-eclampsia) through metabolome wide association analysis and weighted correlation network analysis. **Methodology**: Case-control study derived from Colombian GenPE biobank that recruited pregnant women at the moment of delivery. Metabolomic analysis was performed on serum samples by Proton Nuclear Magnetic Resonance (Nightingale Health Ltd). 250 were assessed (169 direct measures and 81 ratios from direct measures): Fatty acids, lipoproteins, inflammation, fluid balance, apolipoproteins, glycerides and phospholipids, glycolysis related metabolites, lipoprotein particle concentration and size, amino acids, ketone bodies, cholesterol. Metabolomic quality control was good (Success rate for quantification was above 95% for 246 metabolites) and only 4 serum samples did not report values for any metabolite. Imputation (half minimal) and standardization (inverse normal transformation) was carried out prior to the analysis. For metabolome wide association analysis logistic (Pre-eclampsia) and multinomial (HELLP syndrome and early Pre-eclampsia) regression was conducted adjusting for potential confounders and p value was corrected through FDR (Liu method). A weighted metabolite co-expression network analysis (WGCNA) was constructed using a soft thresholding power of 20, R2 of 0.7, minModuleSize of 5, mergeCutHeight of 0.25 and then, correlation between modules and clinical outcomes were assessed to identify significant modules. **Results:** 596 women were included (controls n=357 and cases n=239). Age was 18.6 years (DE 2.8) for controls and 20.7 years (DE 5.2) in cases; metabolites associated with less risk for Pre-eclampsia were those related to HDL (XL-HDL-FC, XL-HDL-C, XL-HDL-PL, XL-HDL-L, XL-HDL-CE and HDL size), XL-HDL-FC was also associated with less risk of HELLP syndrome and early Pre-eclampsia. WCGNA identify seven modules, and one was inversely associated with all outcomes (corr -0.21 pval: 2,7 x10-7 for Pre-eclampsia, -0.12 pval: 0.004 for HELLP, -0.2 pval: 1.2 x 10-6 for early Pre-eclmapsia). Metabolites in that module were XL-HDL-P, XL-HDL-L, XL-HDL-PL, XL-HDL-C, XL-HDL-CE and XL-HDL-FC. **Discussion:** As in cardiovascular diseases, HDL-related metabolites of different sizes had a protective effect for Preeclampsia and its severity. A recent mendelian randomization analysis reported a causal and protective effect for HDL in this disease. These findings provide evidence of molecules of interest for establishing risk or developing potential therapeutic targets.

Session Title: Omics Technologies Poster Session I

PB3430 † Microbiome Preterm Birth DREAM Challenge: Crowdsourcing Machine Learning Approaches to Advance Preterm Birth Research

Authors:

J. Golob[1], T. T. Oskotsky[2], A. S. Tang[2], A. Roldan[2], V. Chung[3], C. Ha[2], R. J. Wong[4], K. Flynn[3], A. Parraga-Leo[2], C. Wibrand[2], S. S. Minot[5], G. Andreoletti[6,2], I. Kosti[2], J. Bletz[3], A. Nelson[3], J. Gao[7], Z. Wei[7], G. Chen[7], Z. Tang[7], P. Novielli[8], D. Romano[8], E. Pantaleo[8], N. Amoroso[8], A. Monaco[8], M. Vacca[8], M. De Angelis[8], R. Bellotti[8], S. Tangaro[8], A. Kuntzleman[9], I. Bigcraft[9], S. Techtmann[9], D. Bae[10], E. Kim[10], J. Jeon[11], S. Joe[11], The Preterm Birth DREAM Community, K. Theis[12], S. Q. Ng[13], Y. S. Lee[13], P. Diaz-Gimeno[14], D. MacIntyre[13], G. Stolovitzky[15], S. V. Lynch[2], J. Albrecht[3], R. Romero[16], D. K. Stevenson[4], N. Aghaeepour[4], A. Tarca[12], J. Costello[17], M. Sirota[2]; [1]Univ. of Michigan, Ann Arbor, MI, [2]Univ. of California San Francisco, San Francisco, CA, [3]Sage Bionetworks, Seattle, WA, [4]Stanford Univ. Sch. of Med., Stanford, CA, [5]Fred Hutchinson Cancer Ctr., Seattle, WA, [6]Astellas Gene Theraphy, San Francisco, CA, [7]Univ. of Wisconsin Madison, Madison, WI, [8]Univ. degli Studi di Bari Aldo Moro, Bari, Italy, [9]Michigan Technological Univ., Houghton, MI, [10]Gwangju Inst. of Sci. and Technology (GIST), Gwangju, Korea, Democratic People's Republic of, [11]Korea Res. Inst. of BioSci. and Biotechnology (KRIBB), Daejeon, Korea, Democratic People's Republic of, [12]Wayne State Univ., Detroit, MI, [13]Imperial Coll. London, London, United Kingdom, [14]Inst. de Investigacion Sanitaria La Fe (IIS La Fe), IVI Fndn., Valencia, Spain, [15]Columbia Univ., New York, NY, [16]NIH, US Dept. of Hlth.and Human Services, Detroit, MI, [17]Univ. of Colorado Anschutz Med. Campus, Aurora, CO

Introduction: Every year 11% of infants are born preterm with significant health consequences. The vaginal microbiome has been extensively correlated in single center studies with the risk for preterm birth (PTB), opening the promise of using microbiome data to build rigorous, generalizable, and robust predictive models to identify at-risk pregnancies. However, combining data from different microbiome studies into a single dataset is needed for greater statistical power but a non-trivial feat. We aggregated multiple vaginal microbiome studies and leveraged this data for a crowdsourced challenge to predict the risk of PTB. **Methods:** Using a novel technique (MaLiAmPi) we aggregated public raw data from 9 vaginal microbiome studies (total: 3578 samples from 1268 pregnant individuals) and two not-public datasets for validation (total: 331 samples from 148 pregnant individuals). We organized a challenge that took place July-September 2022 to predict: (a) preterm (PTB; <37 weeks) or (b) early preterm birth (ePTB; <32 weeks). Challenge teams worldwide received the training data, then built and submitted their models through our online platform. Model performance metrics on the validation data were returned to the teams. Teams were limited to 5 total submissions with the top performing model selected for ranking. We carried out several techniques to ensure robustness of the resulting rankings including test set label inversion, bootstrapping, over- and under-sampling. **Results:** Top performing models (among 148 and 121 submissions from 318 teams) achieved AUROC scores of 0.69 and 0.87 predicting PTB and ePTB respectively, exceeding organizers' baseline models performance. Most top performing models used ensemble approaches including tree-based methods, e.g., random forest and relatives. A few used regression approaches with feature pruning and clustering, or neural networks. Alpha diversity, VALENCIA community state types (CSTs), and taxonomy-independent composition were important features in top performing models. Through feature permutation combined with independently-developed highly predictive models, we identified multiple organisms that associate with PTB risk. Specifically, there seems to be physiologically-relevant species- and strain-level variability within the Lactobacillus and Gardnerella genera across trimesters that indicates a potential role for intraniche competition in vaginal microbiome during pregnancy and the risk for early PTB. **Discussion:** This work is a model for implementing a crowdsourcing strategy to power the translation of microbiome data into clinically relevant predictive models and better understand PTB.

Session Title: Omics Technologies Poster Session II

PB3431 † Mind the reference gap

Authors:

J. Eisfeldt[1], K. Bilgrav Saether[1], E. Ten Berk De Boer[1], I. Bunikis[2], C. Carvalho[3], A. Ameur[4], A. Lindstrand[5]; [1]Karolinska Inst, Stockholm, Stockholm, Sweden, [2]Uppsala Univ., Uppsala, Sweden, [3]Pacific Northwest Res. Inst., Seattle, WA, [4]Uppsala Univ, Uppsala, Sweden, [5]Karolinska Inst.t, Stockholm, Sweden

Since its release in the early 2000s, the human reference genome has been an invaluable resource in the field of human genomics. Yet, despite significant improvements, considerable sections of GRCh38 remain absent or computer simulated. The recently published Telomere to Telomere (T2T) reference genome offers a continuous sequence, promising heightened structural variant detection and improved mapping. We initiated this study when an inv6(p12.3q16.1) was undetected in the GRCh37 genome, but clearly visible in T2T, GRCh38, and the genomes of both chimpanzees and bonobos. The inversion was first detected by karyotyping in a male patient with hearing impairment, intellectual disability, strabismus, diplopia, anosmia, autistic features and hypogonadism. We next aligned short read genome data from 100 healthy individuals to various reference genomes, including T2T, GRCh37, GRCh38, and those of chimpanzee and bonobo. This allowed us to detect regions present in primate genomes but missing in at least one human reference, Differential Reference Regions (DRRs), thereby uncovering an intriguing similarity between GRCh37 and the chimpanzee genome. Our analysis reveals that most analyzed DRRs demonstrate low coverage, and it underscores the impact of such regions on structural variant detection and variability. Despite the promising prospects of the T2T reference, we found GRCh38 to be the most dependable for optimal structural variant detection and mapping. Our results highlights the need for high-quality diploid assemblies or graph genomes to enhance the efficiency of structural variant workflows. Specifically, a pangenome or de novo assembly may be necessary for comprehensive population representation, particularly in the diagnosis of rare diseases. Our findings collectively call attention to the criticality of ancestral sequences missing from current human reference genomes.

Session Title: Omics Technologies Poster Session III

PB3432 Miqa: A no-code, automated bioinformatic QA platform to continuously improve the accuracy and reliability of computational tools used in clinical genomics.

Authors:

V. Lau, G. Berry, K. McKeag; Magna Labs, Falls Church, VA

Software quality impacts research output and decision making in clinical genomics. While next-generation sequencing (NGS) data continues to scale in size and complexity, quality assurance (QA) processes of bioinformatic pipelines struggle to keep up with the speed of their development. Reliable bioinformatic tools for genetic variant detection are crucial to providing accurate disease diagnosis and patient treatment. Continuous and systematic evaluation in clinical bioinformatics is often hindered by inadequate resources (time and funding) or software engineering expertise. Researchers often rely on periodic benchmarking studies for performance data that may not be suitable to their particular genomic use case or datasets. Alternatively, bioinformatic engineers reportedly spend up to 40-60% of software development and maintenance time on manual troubleshooting. The “shift left” model of regularly detecting software defects early in the development process has proven to drastically improve software quality and delivery. This common practice in conventional software engineering is not well implemented in computational biology. To address the technological and knowledge gap in bioinformatic software testing, we have developed a no-code, automated QA platform, Miqa, to foster continuous quality surveillance and benchmarking of clinical NGS analysis pipelines. Miqa can fully automate the end-to-end QA process from test data preparation, NGS pipeline test execution to real-time QA result reporting. Its user-friendly graphical interface designed for scientific users regardless of programming skill level. To demonstrate Miqa’s real-time software bug detection, we performed a case study on the GitHub issue #1446 of the open-source variant caller, bcftools, which was generating inconsistent variant calls. Using Miqa, we benchmarked bcftools version 1.4 to 1.12 using an admixed small variant truthset from the Platinum Genome reference samples NA12877 and NA12878, using common quality metrics such as concordance, specificity and sensitivity. Miqa identified over 10% drop in variant call concordance in v1.4.1 which was released in 2017, but this bug was not reported until after the release of v1.12 in 2021. Miqa’s interactive visualization tools quickly pinpointed the cause of such discordance to an inconsistent read depth in indel regions in a specific version update. This case study highlighted the utility of QA automation tools like Miqa to proactively detect errors at each software release or code change. Ensuring software quality and reliability can further drive the safe adoption of clinical genomic tests to improve patient care.

Session Title: Omics Technologies Poster Session I

PB3433 MntJULiP and Jutils: Tools for differential splicing detection and visualization from complex RNA-seq data sets

Authors:

W. Lui[1], G. Yang[1], Z. He[1], S. Sabunciyan[2], L. Florea[3]; [1]Dept. of Computer Sci., Johns Hopkins Univ., Baltimore, MD, [2]Dept. of Pediatrics, Johns Hopkins Sch. of Med., Baltimore, MD, [3]Dept. of Genetic Med., Johns Hopkins Sch. of Med., Baltimore, MD

Differences in alternative splicing (AS) patterns can reveal important markers of phenotypic differentiation. Increasingly large and complex data sets that include multiple confounders such as sex, age, ethnicity and clinical attributes, are emerging from disease study cohorts and population-level projects, demanding highly specialized data analysis tools. We developed two tools, MntJULiP and Jutils, for differential splicing detection and visualization that can efficiently handle large-scale and complex RNA-seq data collections.

The first tool, MntJULiP, detects intron-level AS differences from RNA-seq data using a Bayesian mixture model. It identifies changes in intron splicing abundance and intron splicing ratio, and can capture AS variation more comprehensively than other tools (PMID:36104797). Notably, it allows for multi-way comparisons, which are better suited for complex and time-series experiments. To account for external relationships in the data, we recently introduced covariates as linear components in the model. The second tool, Jutils, visualizes AS variation with heatmaps, PCA and sashimi plots, and Venn diagrams. Our tools are scalable and can process thousands of samples within hours.

To illustrate, we applied our methods to 1,398 GTEx RNA-seq samples from 13 brain regions. The first comparison, among tissues, revealed distinct groupings between the cerebellar, cortex, and basal ganglia regions, which did not change when accounting for covariates including ‘sex’ and ‘age’ at death. Secondly, we compared 120 frontal cortex RNA-seq samples by age groups (‘20s’, ..., ‘70s’). Changes in the frontal cortex in aging are thought to contribute to sex-specific differences in the prevalence of neurological disorders (PMID:37284017). More differences were observed with more distant age groups. When ‘sex’ was used as a covariate, a significant increase in the number of differences was observed between the ‘20s’ and ‘40s’ groups, indicating a possible mark of sex-specific differentiation. Lastly, we compared the 83 male (M) and 37 female (F) frontal cortex samples to identify sex-specific differences in splicing. Jutils heatmaps revealed a subgroup of 22 samples (12F and 10M) with a distinct AS pattern, which became evident when regressing for ‘age’. This subgroup was higher aged (50-79), pointing to a distinct splicing program in a subset of individuals with aging.

MntJULiP and Jutils are highly effective and efficient analysis tools for large-scale complex RNA-seq datasets with confounding factors, and can reveal new insights into disease and population substructure. Funding: R01-GM129085.

Session Title: Omics Technologies Poster Session II

PB3434 Molecular and cellular characterization of the Parkinson's disease olfactory bulb.

Authors:

A. Garton, K. Lehnert, M. Curtis, J. Jacobsen; Ctr. for Brain Res., Univ. of Auckland, Auckland, New Zealand

Parkinson's disease (PD) is a neurodegenerative disorder associated with motor symptoms (bradykinesia, rigidity, and resting tremor) and non-motor symptoms (such as sensory disturbances). Pathologically, PD is characterized by the presence of neuronal inclusions primarily composed of alpha-synuclein in the midbrain, and by a progressive degeneration of dopaminergic neurons in the substantia nigra pars compacta. It is a heterogenous and complex disorder, with more than 20 genes causative of or associated with PD; however, for the vast majority of cases the underlying cause remains unknown. Interestingly, up to 95% of PD cases have olfactory dysfunction prior to motor symptom onset, and alpha-synuclein pathology has been identified in the early stages of PD in the olfactory bulb (OB). However, the human OB remains largely uncharacterized, particularly in the context of disease processes. The aim of this project is to characterize the cellular structure of the human OB, and subsequently compare PD and non-PD OB tissue to identify vulnerable cell populations and transcriptional cascades in the OB that may lead to PD. This will be achieved by integrating transcriptomic (bulk, single-nuclei, and spatial RNA sequencing) and RNAscope in situ hybridization evidence from PD and control OBs from the New Zealand Neurological Foundation of New Zealand Human Brain Bank.To characterize and compare cell types in the PD OB, a pilot experiment (one PD OB and one non-PD OB) is being conducted, integrating spatial and single-nuclei RNA sequencing (RNA-seq) to assess transcriptional profiles of cells. For spatial RNA-seq, one medial and one lateral OB section was mounted on 10X spatial gene expression slides. Sections were stained with NeuN, UEA1-lectin, alpha-synuclein, and DAPI prior to library generation. Initial results show distinct spatial gene expression profiles that mirror known patterns of OB tissue architecture and organization (based on previous IHC work). For example, sections with anterior olfactory nucleus staining show significant differential expression of known marker *UCHL1* (PGP9.5). The remaining OB tissue was sectioned for single-nuclei RNA-seq (to obtain cell-type specific gene expression profiles), bulk RNA-seq, and will be used for RNAscope for the confirmation of pathological leads. Utilizing spatial transcriptomics in conjunction with single-nuclei RNA-seq will allow us to generate a spatially resolved map of the OB at single cell resolution for the first time in humans. This could lead to the identification of vulnerable cell types early in the disease cascade for PD, and ultimately identify novel intervention strategies prior to symptom onset.

Session Title: Omics Technologies Poster Session III

PB3435 Morphology profiling driven by deep learning characterizes functional changes in CRISPR knockout cell lines

Authors:

V. Lu[1], A. Jovic[1], T. Vollbrecht[1], J. Mei[1], T. Pham[1], C. Corona[1], R. Carelli[1], K. Saini[1], J. Kim[1], Z. Lian[1], N. Banovich[1], D. Schraivogel[2], L. Steinmetz[2], K. B. Jacobs[1], S. C. Boutet[1], M. Masaeli[3]; [1]Deepcell, Inc, Menlo Park, CA, [2]EMBL Heidelberg, Heidelberg, Germany, [3]Deepcell, Menlo Park, CA

CRISPR/Cas9-based screens have revolutionized functional genomics by enabling the knockout (KO) or expression modulation of hundreds or thousands of genes in parallel. The link between perturbations and corresponding cellular phenotypes can be performed using phenotypic assays, sequencing of gRNA repertoires, or single cell RNA sequencing. However, these assays can be very expensive and inherently include many perturbations with weak or no effects on overall phenotype.

Accordingly, there is a need for orthogonal and high-resolution methods for characterizing phenotypes and isolating cells which receive meaningful and phenotypically relevant perturbations. To this end, we performed high-dimensional morphology profiling on multiple CRISPR KO cell lines to determine phenotypic differences arising from single gene perturbations. We used the REM-I platform, which combines deep learning, microfluidics, and high resolution imaging to characterize and isolate cells of interest.

We trained a self-supervised model using 115 deep learning and morphometric features extracted from images of single cells. To evaluate the application of this model in genetic screens, a library of 11 single gene KOs were generated in three cell lines: HEK 293 (Human Embryonic Kidney cells), Jurkat (T Lymphocyte), and K562 (Myelogenous Leukemia), and compared to a Cas9 only negative control cell line. Individual genes were selected based on previously reported effects on cell morphology, such as vesicle formation, cell size, and nuclear size. We found single gene KOs resulted in significant morphological changes relative to control cells. Morphology profiling results using the REM-I platform captured high-dimensional functional resolution of single gene KOs, including distinct responses from cell line to cell line, indicating the same gene perturbation translates to different phenotypic effects in accordance to cell context. We also found consistent phenotypic patterns in KOs from similar gene expression pathways within the same cell lines, suggesting that morphology as a functional readout is robust and orthogonal. Lastly, we observed several rare morphological phenotypes (~1% of the population), which give further insight into the relative importance of a single gene in each cell context. Future directions include sorting cells from a pooled CRISPR screen and expanding the scope of these results to large scale genetic perturbation screens. In conclusion, we show that deep learning-driven morphology analysis is a promising approach for capturing a wide range of phenotypes induced by genetic perturbations, such as CRISPR KO.

Session Title: Omics Technologies Poster Session I

PB3436 Moving beyond sequence: leveraging DNA structure in genomic deep learning models

Authors:

R. Keivanfar[1], F. Yang[1], K. S. Pollard[2], N. M. Ioannidis[1]; [1]UC Berkeley, Berkeley, CA, [2]UCSF, San Francisco, CA

Genomic deep learning models that learn complex motifs and regulatory grammars from DNA sequence input have had success in predicting molecular phenotypes directly from sequence, such as transcription factor binding, histone modifications, chromatin accessibility, and gene expression. Once trained, these models have the potential to be powerful tools for interpreting personal genomes by predicting the molecular effects of individual genetic variants and haplotypes. There is great need for variant prioritization methods for non-coding variants, which contribute substantially to trait heritability. However, recent studies suggest that current models perform poorly at understanding variation in personal genomes and predicting the direction of effect of many genetic variants, despite being accurate from locus to locus using the reference genome in which they were trained. We hypothesized that these models, in their current form, do not adequately capture the role of DNA's structural attributes, known as “shape features”, in dictating molecular functions such as protein-DNA interactions. Here we investigate the extent to which current genomic deep learning models implicitly account for DNA shape information when processing sequence data alone. Initial findings suggest a potential gap in these models' ability to capture the structural features of DNA and to explain the binding of transcription factors that rely on DNA shape. We explore the impact of integrating explicit DNA shape features into sequence-based deep learning models to enrich their understanding of DNA-protein interactions, paving the way for the use of DNA shape features in predicting a wider array of molecular phenotypes.

Session Title: Omics Technologies Poster Session II

PB3437 Multi Origins Cell type Alignment (MOCA): Deep Learning-Based Integration of Fibroblast Populations across Human Organs Reveals Universal Developmental Trajectories in Fibrotic Tissues

Authors:

Y. Lyu, B. Fang, E. Ayaub, K. Ask, J. Loureiro; Novartis Inst.s for BioMed. Res., Boston, MA

Fibroblasts are critical to maintain organ integrity, supporting other cell types, and driving fibrosis development. The presence and persistence of activated fibroblasts contribute to progressive fibrotic scarring observed in various organs including the lung, liver, heart, and skin. To identify and characterize fibroblast states in healthy and diseased organs, previous studies have integrated fibroblast populations from different tissues at a single-cell resolution. However, such directly comparing transcriptional profiles from different origins lacks the power to reveal universal gene regulatory dynamics and disease fibroblast developmental trajectories, as it fails to differentiate common fibroblast transcriptional programs from origin-specific ones.To address this limitation, we present "Multi Origins Cell type Alignment" (MOCA), a deep learning-based framework designed to align scRNAseq data across multiple origins, including distinct biological and technical sources. MOCA consists of two key components: a stacked autoencoder that captures distinct transcriptional states within cell populations, and an origin discriminator that removes origin-specific biological and technical artifacts. Such framework allows effectively capturing universal cell population dynamics from control to disease conditions, and uncover conserved transcriptional features shared by same cell type across different origins.Applying MOCA to single-cell transcriptional profiles of 28,228 fibroblasts/myofibroblasts from 173 human fibrosis/control tissues across 6 independent (public) studies, we construct a map illustrating universal fibroblast dynamics from control to fibrotic conditions. Our analysis uncovers a common disease-specific fibroblast shared by fibrotic liver and lung, along with its conserved transcriptional features. Moreover, we projected fibroblast population from in vitro mouse model onto the human fibroblast trajectory to evaluate its translation potential.Furthermore, MOCA's versatility extends beyond fibroblasts, as it can be easily adapted to analyze other cell types such as macrophages. We also demonstrate the robustness and scalability of MOCA as a batch removal method.In conclusion, our study introduces MOCA as a deep learning-based framework for aligning scRNAseq data across multiple origins. By enabling the construction of a map depicting universal fibroblast dynamics from control to fibrotic disease conditions, MOCA uncovers key fibroblast activation states present in multiple fibrotic tissues, offering novel insights into therapeutic strategies and targetable cell populations.

Session Title: Omics Technologies Poster Session III

PB3438 Multi-cohort cerebrospinal fluid proteomics identifies robust molecular signatures for asymptomatic and symptomatic Alzheimer's disease

Authors:

M. Ali[1], J. Timsina[1], L. Wang[1], A. Do[1], D. Western[1], Knight Alzheimer Disease Research Center (Knight ADRC), Alzheimer disease neuroimaging initiative (ADNI), Dominantly Inherited Alzheimer Network (DIAN), Fundació ACE Alzheimer Center Barcelona (FACE), Barcelona-1 (Pau), Y. Sung[2], C. Cruchaga[1]; [1]Washington Univ. in St. Louis, St. Louis, MO, [2]Washington Univ. in St. Louis, St Louis, MO

Alzheimer's disease (AD), the most common form of dementia, is a complex polygenic disease characterized by the accumulation of Amyloid-β and hyperphosphorylated Tau proteins in the brain. An increased level of these proteins, among others, in the brain and cerebrospinal fluid (CSF) can be detected years before the symptoms of AD appear. Therefore, studying AD proteome in CSF can reflect its diverse underlying pathophysiology and pave the way for reliable diagnostic and therapeutic advancements.

Here, we present one of the largest AD proteomic profiles, based on 7,029 protein analytes measured in CSF of a total 3,065 individuals in a three-stage study. Firstly, discovery was performed in 1,170 samples from the Knight ADRC and FACE discovery cohorts using the ATN framework (AT- = 680 and AT+ = 490). Secondly, the proteins that passed multiple test corrections in these datasets were further tested in 593 individuals (AT- = 235 and AT+ = 358) from the ADNI and Barcelona-1 replication cohorts. Lastly, the proteins that passed multiple Bonferroni corrections on the meta-analysis were further utilized for creating AD prediction models and conducting pathway and network enrichment analyses to gain mechanistic insights into AD pathophysiology.

The differential abundance analysis in discovery identified 3,565 proteins significantly (FDR < 0.05) associated with AT(n) status. Of these, 2,543 also passed FDR in the replication with the same direction of effect, of which 2,173 passed the stringent Bonferroni correction and were used for downstream analyses. Some of the highly significant proteins included YWHAG ($P = 4\times10^{-219}$), SMOC1 ($P = 6\times10^{-208}$), and NRGN ($P = 3\times10^{-119}$). Using lasso regression, we identified a distinctive signature of 25 proteins with robust and high predictive capability for AD in three independent cohorts. They were unique to AD and showed no predictive power for other dementia such as FTD, DLB, or PD. Disease and pathway enrichment analyses highlighted several neurological disorders (e.g. AD, tauopathy, and synucleinopathy) and neuronal functions (neuron projection morphogenesis, synapse assembly, and axonogenesis) to be significantly enriched (FDR < 0.05) in the altered AD CSF proteome.

Our findings show the promising potential of AD CSF proteomics in the development of a robust AD prediction model. The employed systematic analysis revealed differentially abundant proteins and various biological pathways that are compromised in AD, thereby, increasing our understanding of AD biology. Our findings may accelerate the development of effective intervention therapies that target the earliest molecular triggers of AD.

Session Title: Omics Technologies Poster Session I

PB3439 Multiomic single cell analysis of primary pancreatic ductal adenocarcinoma enhanced by fixation.

Authors:

J. Dahl[1], K. A. Kennedy[1], J. Remington[1], N. Lester[2,3], J. Su[2,3], C. Shiau[2,3], J. A. A. West[1], W. L. Hwang[2,3], **D. H. Gong**[3]; [1]BioSkryb Genomics, Durham, NC, [2]Massachusetts Gen. Hosp. and Harvard Med. Sch., Boston, MA, [3]Broad Inst. of MIT and Harvard, Cambridge, MA

Introduction: Single-cell analysis of certain tumor types has been challenging based on the high levels of DNAses and RNAses found within the tumor cells. While traditional formalin fixation can inactivate these enzymes, it can also limit downstream single-cell analysis. The purpose of our study was to develop a non-formaldehyde fixation protocol that preserves single cells for whole genome and transcriptome amplification and evaluate its performance in clinical samples.
Methods: We fixed NA12878/HG001 cells with different formaldehyde-free fixatives and used live (unfixed) cells as a control. We then compared the performance characteristics of single-cell genomes and transcriptomes amplified using ResolveOME (BioSkryb Genomics) in both fixed and unfixed cells. Based on these data, we selected a fixative whose performance in fixed cells was undistinguishable from live cells. We next applied this workflow to human pancreatic ductal adenocarcinoma (PDAC) primary cells, which are notorious for high levels of DNAse and RNAse activity. Manually dissociated PDAC cells were sorted for viability and into pan-cytokeratin positive (CK+) tumor cells and tumor microenvironment (CK-) cells. Half of each cell population was fixed, while the other half was left unfixed. Whole genomes and transcriptomes were amplified using ResolveOME. Resulting libraries were sequenced by low pass whole genome sequencing. Bioinformatics analysis was performed using BaseJumper bioinformatics platform (BioSkryb Genomics).
Results: Fixation enhanced the probability of amplification from both CK+ and CK- PDAC single cells. Transcriptional profiling of these two populations by principal component analysis segregated single tumor cells from non-tumor cells across both components of the analysis. As expected, the CK- cells, identified as tumor infiltrating immune cells by transcriptional analysis, exhibited normal ploidy across the genome. Analysis of individual CK+ cells revealed known copy number alterations and transcription profiles consistent with known PDAC tumor biology. Key upregulated genes included *ERBB3* and *MET* in the MAPK pathway, and *CIB1* and *TENT4A* in the double strand break repair (DSBR) pathway. Mitochondrial pathways were also up-regulated in CK+ relative to CK- cells, likely reflecting a tumor-specific response to increased energy needs within the cells.
Conclusions: We have developed a method for improved retrieval of multiomic details from challenging cell types. This approach leverages a new fixation approach coupled with whole genome and transcriptome amplification, enabling single-cell analysis in human PDAC samples.

Session Title: Omics Technologies Poster Session II

PB3440 Multi-omics Integration Analysis for Osteoporosis Biomarker Discovery

Authors:

C. Qiu, K. Su, X. Zhang, A. Liu, Y. Gong, L. Jiang, W. Lin, Z. Luo, Q. Tian, H. Shen, H-W. Deng; Tulane Univ. Sch. of Med., New Orleans, LA

Osteoporosis is an increasingly serious public health problem, which is characterized by low bone mineral density (BMD). To date, the functional roles of the most reported biomarker that are associated with BMD remain unclear. To increase the understanding of the biological mechanisms of BMD variation, it is necessary to leverage multiple-dimensional datasets to capture the complexity of biological networks. The recent advent of high throughput technologies has created an opportunity for integrating multi-omics data to build a novel model of the molecular changes in osteoporosis. In this study, we performed a most comprehensive trans-omics study with various omics data, such as transcriptomics, methylomics, metabolomics, and lipidomics, from 546 subjects, to identify interactions/networks and particularly causal regulatory relationships within and especially those between omic molecules with the purpose to discover molecular genetic mechanisms underlying osteoporosis etiology in vivo in humans. By focusing on potential interesting biomarkers (e.g., differentially expressed genes, differentially methylated CpG sites, differential metabolites, and differential lipids) generated from single-layer analysis, we applied a data integration analysis for biomarker discovery using latent components (DIABLO) to identify the key omics biomarkers for BMD variation. We identified 19 mRNA biomarkers, 46 methylation biomarkers, 12 metabolomics biomarkers, and 8 lipid biomarkers, respectively. Pathway analysis showed that most of the identified biomarkers are enriched in bone-related signaling pathways, such as MAPK/TGF-β pathway. In addition, our preliminary results of network analysis also showed an extensive correlation among the identified biomarkers. Further causality analysis will be performed to test their causal regulatory effects within and between omic molecules. In conclusion, the integration of trans-omics and comprehensive bioinformatics analyses provided us a holistic understanding of the underlying functional mechanisms, molecular regulatory information flow, and the interactive molecular systems among different omic molecules for osteoporosis risk.

Session Title: Omics Technologies Poster Session III

PB3441 Multiomics reveals oral microbiome mediated biological mechanisms underlying obesity in Emiratis.

Authors:

T. Denekew, A. A. Shibl, A. Ahmad, C. E. Leonor, L. Utenova, E. Zhang, M. Abd AlBaqi, T. ., R. Ali, Y. Idaghdour, A. R. Jha; New York Univ. Abu Dhabi, Abu Dhabi, United Arab Emirates

The gut microbiome has been causally linked to human diseases but the contributions of the oral microbiome - the second largest microbial ecosystem in the human body - in human health and diseases is markedly underexplored. Here, we characterized the oral microbiome structure using 16S rDNA sequencing in 629 Emiratis enrolled in the UAE Healthy Future Study (UAEHFS). Our results revealed robust associations between the oral microbiome composition and diversity with BMI in this cohort. We further validated our results after matching for potential confounders and demonstrated that obese individuals (BMI≥ 30) had significantly less diverse microbiota and were compositionally different from those of healthy weight ($18.5 \leq BMI \leq 24.9$). To elucidate the oral microbiome driven biological mechanisms underlying obesity, we performed a multi-omics analysis integrating whole-metagenomics shotgun sequencing, untargeted metabolomics, blood markers, and health data. Strain-level microbial community structure obtained from metagenomic sequencing corroborated results from 16S analysis. Notably, we identified several bacterial species differentially abundant between the obese and non-obese groups. Functional profiling revealed a large repertoire of differentially enriched metabolic pathways in obesity-associated microbiomes. Of particular interest were pathways involved in biosynthesis of essential vitamins, which exhibited notable depletion in obese individuals. These pathways were further validated using clinical biomarkers obtained from blood and urine samples. By incorporating untargeted metabolomics analysis into our study, we aim to uncover metabolic signatures and further elucidate metabolic alterations associated with obesity and the oral microbiome. Overall, our study illuminates the unexplored connection between the composition and function of the oral microbiome and obesity. These findings lay the foundation for utilizing oral microbiome markers as diagnostic tools for cardiometabolic conditions, presenting a promising avenue for personalized healthcare and disease prevention.

Session Title: Omics Technologies Poster Session I

PB3442 Multi-transcriptomic analyses reveal altered expression profiles of immune response genes in Pick's disease

Authors:

N. Tamvaka, A. I. Soto-Beasley, Y. Ren, M. G. Heckman, Z. S. Quicksall, S. F. Roemer, E. Udine, M. van Blitterswijk, D. Dickson, O. A. Ross; Mayo Clinic, Jacksonville, FL

Background: Pick's Disease (PiD) is a rare neurodegenerative disorder characterized by dementia, with frontotemporal degeneration and pathognomonic tau cytoplasmic inclusions known as Pick bodies observed at autopsy. PiD is a type of primary tauopathy along with Progressive Supranuclear Palsy (PSP) and Corticobasal Degeneration (CBD). Tauopathies can be classified as 3-Repeat (3R; PiD), 4-Repeat (4R; PSP, CBD), and 3R+4R (Alzheimer's disease, AD) based on the predominance of the tau isoform present in the cytoplasmic inclusions characterizing each disease. PiD is the only known 3R tauopathy, however due to disease rarity it remains significantly understudied and the causes that are driving this preferential 3R tau accumulation in PiD remain unclear. Furthermore, no study has characterized the transcriptomic profile of PiD, which is imperative for gaining insight into disease specific mechanisms and pathways contributing to disease pathogenesis. Methods: We have performed the first multi-transcriptomics experiments on PiD cases (n=28) using short-read (bulk), long-read (PacBio, IsoSeq), and single-nuclei (snRNA) RNA sequencing approaches. We used the significant differentially expressed genes (DEGs) to inform into differential transcript expression, cell type specific expression, network analysis, and gene expression association studies with neuropathology measures and we compared our findings to control samples (n=15) and PSP cases (n=25). Results: Differential gene expression analysis of the bulk data identified 14 significant DEGs between PiD cases and controls and IsoSeq data was used to quantify specific transcript expression of those genes. Network analysis revealed their enrichment in immune system associated modules significantly upregulated in PiD cases. A subset of those (*CCL2* and *SERPINA3*) were also significantly overexpressed when comparing PiD and PSP cases. Interestingly, *CCL2* was significantly positively associated with increased Pick body burden in PiD. *CCL2* and *SERPINA3* are two known modulators of the immune system and our snRNA Seq data confirmed their expression in cell populations involved in immune processes. Conclusions: Our data highlights the use of multi-transcriptomics to capture the unique transcriptional signature of 3R tau pathology in PiD and suggests the involvement of inflammatory and immune processes in the pathophysiology of the disease. We are further investigating the relationship between the DEGs and integrating this data with spatial transcriptomics and whole genome sequencing to better define 3R tau-specific disease relevant pathways and nominate molecular therapeutic targets.

Session Title: Omics Technologies Poster Session II

PB3443 Nanopore Adaptive Sampling combines both targeted and low-pass whole genome sequencing in a single assay.

Authors:

E. Harrington[1], S. Aganezov[2], C. Tyer[3], H. Greenfield[4], S. Hickey[3], P. James[4], **S. Juul**[5]; [1]Oxford Nanopore, New York, NY, [2]Oxford Nanopore Technologies, Inc., Reading, MA, [3]Oxford Nanopore Technologies, Alameda, CA, [4]Oxford Nanopore, Oxford, United Kingdom, [5]Oxford Nanopore Technologies, New York, NY

In the 15 years since the first large-scale genome-wide association study (GWAS), there has been an explosion in the number of diseases and traits with significant associations. These discoveries not only shed light on the biology underlying complex traits but also enable the use of polygenic scores (PGS) to predict phenotypes. Genotyping arrays have played a crucial role in this by providing a cost-effective method to assay common SNVs in large cohorts. More recently low-pass whole-genome sequencing (lpWGS) in combination with genotype imputation has emerged as an alternative method for GWAS and PGS calculation. However unlike higher-coverage WGS and targeted sequencing, these technologies are not well-suited to the detection of rare or novel variation, limiting their ability to explore the effects of rare variants and make phenotype predictions based on the full allele-frequency spectrum. Previous work has shown that genotypes can be accurately imputed using Oxford Nanopore lpWGS data. We expand on this work to explore the effect of DNA fragment length on imputation accuracy. We show that read count is a more important metric than coverage and that <1kb reads can be used for accurate imputation. We also show that PGS for common diseases can be accurately calculated at depths as low as 0.5X. We think that nanopore lpWGS will be particularly useful for populations and applications not well served by current arrays. The absence of a design-cycle combined with the low up-front cost of nanopore sequencing removes a barrier to entry for large genotyping projects. Adaptive Sampling (AS) is a method of targeted sequencing unique to the Oxford Nanopore platform. It works by sequencing part of a strand of DNA, and, based on the initial sequence of the strand, a decision is made to either eject it from the pore or continue sequencing the entire strand. Most often this decision is based on whether the initial sequence maps to a target region in the genome. The result is that these regions are covered to a higher depth with long reads, which can be used to call and phase SVs, SNVs, repeat expansions, CNVs, and epigenetic modifications. The rest is covered to a lower depth with short reads, essentially lpWGS data. We show that AS data can be used for accurate imputation, providing a method to simultaneously assay common genetic variation genome-wide as well as genotype and discover variants in target regions. Several other features of AS make it particularly useful: i) it can target regions refractory to other methods (e.g. HLA); ii) the informatic nature of the targeting method means that there are no up-front costs in designing a panel, enabling bespoke panels to answer specific hypotheses.

Session Title: Omics Technologies Poster Session III

PB3444 Novel Optical Genome Mapping Algorithm for Detection of Structural Variants in Hematological Cancers

Authors:

M. Muggli[1], T. Anantharaman[1], M. Molnar[1], S. Al-Saffar[1], N. Miller[2]; [1]Bionano, San Diego, CA, [2]Bionano Genomics, San Diego, CA

Optical genome mapping (OGM) enables detection of genomic structural and copy number variants that cannot be detected reliably by next-generation sequencing (NGS) technologies and are often missed by conventional cytogenetic techniques. Bionano has developed bioinformatics analysis pipelines for calling structural variants in constitutional and cancer samples. The Bionano Solve de novo assembly pipeline performs a whole genome de novo assembly using a consensus overlap graph method and is intended for constitutional analyses. The Rare Variant Analysis (RVA) pipeline, aimed at low allele-fraction cancer applications, uses an approach where candidate structural variant sites are identified through a molecule-to-reference alignment, followed by localized assembly around the candidate sites.

Bionano has developed a new analysis pipeline that aims to combine the low-allele fraction detection capability of the RVA pipeline with the whole genome coverage and ability to detect smaller structural variants enabled by the de novo assembly pipeline. The Guided Assembly pipeline uses the reference genome as an initial seed followed by extension, refinement and structural variant calling steps. This new analysis method has been evaluated through comparison to previous results from both existing pipelines as well as standard benchmarking datasets used to estimate structural variant calling performance.

Guided assembly was performed for 10 acute myeloid leukemia (AML) samples and 10 myelodysplastic syndrome (MDS) samples which had previously been analyzed with the RVA pipeline. Structural variant calls from guided assemblies were compared to previous results and evaluated for concordance. Simulated samples representing structural variants at 5%, 10%, 20%, 30% and 50% variant allele fraction were created and detection sensitivity and precision for different structural variant types and size ranges were calculated.

Concordance between the Guided Assembly results and previous RVA pipeline results was established. Increased sensitivity for detecting insertion variants smaller than 5 kb and larger than 200 kb was observed, while finding equivalent performance for other variant types with the updated methods.

Session Title: Omics Technologies Poster Session I

PB3445 Optimization of long-read platforms for clinical utility.

Authors:

D. Muzny[1], Q. Meng[1], Y. Han[1], H. Shen[1], G. Weissenberger[1], V. Vee[1], J. Hwang[1], S. Dugan[1], G. Metcalf[1], S. N. Jhangiani[1], J. Hu[1], H. Doddapaneni[1], Y. Zhou[1], Z. Khan[1], M. Mahmoud[1], F. J. Sedlazeck[1,2,3], R. A. Gibbs[1], All of Us Research Program; [1]Human Genome Sequencing Ctr., Baylor Coll. Med., Houston, TX, [2]Dept. of Molecular and Human Genetics, Baylor Coll. of Med., Houston, TX, [3]Dept. of Computer Sci., Rice Univ., Houston, TX

Long-read sequencing platforms, including Pacific Biosciences (PacBio) and Oxford Nanopore (ONT), have the potential to revolutionize personalized medicine through the ability to accurately assess all clinically relevant structural variants (SVs), repeats and rearrangements, often undetectable with short read technologies. While much progress has been achieved in the application of long read technologies for research, true production-scale processing for clinical reporting requires consistent high-quality data, low DNA sample input requirements (1-3ug), platform and software stability, data standards as well as competitive cost models. The Baylor College of Medicine - Human Genome Sequencing Center (BCM-HGSC) has assessed long-read data from both the PacBio Sequel IIe/Revio and the ONT PromethION long-read platforms by the generation of >25x whole genome sequence data. Samples were selected from a Hispanic population that was part of the NIH All of Us Research Program (AoURP) Freeze one (98K) short read data and to date, we have completed 403 ONT genomes averaging 30x coverage/PromethION flow cell and 173 PacBio genomes averaging 38x coverage/three PacBio Sequel IIe SMRT cells each of at ~20kbp N50 read length. Optimized protocols for library, sequencing, and analysis pipelines were established, along with performance metrics that will provide assessment of each platform's readiness for clinical applications. Specifically, library and sequencing kits (PB Express 3.0/Sequel V3 kits and ONT LSK114/ R10.4.1 kits) were optimized to consistently deliver >24Gb/Sequel IIe SMRT cell and >80Gb/PromethION flow cell with an insert size >15Kb. Laboratory optimizations included precision size cuts using Pippin HT, controlled shearing parameters, optimization of all library steps for maximum yield and titration of instrument loading amounts to maximize yield. Key production metrics were established for both platforms, including mean coverage, % genome coverage at 10x, HIFI yields, mean read length, contamination, and Q30 or Q10 mapped bases. Initial accuracy for SNVs was assessed using the standard NIST control (HG002) at 99.8% for PacBio and 97.1% for ONT. For SVs we observe similar high accuracies of 97.0% and 92.6%. These platforms are now ready for integration into the clinical laboratory for application in routine clinical testing. This work was part of the operational development of platforms and data for the All of Us Research program.

Session Title: Omics Technologies Poster Session II

PB3446 Optimizing generation of human isogenic iPSC lines via CRISPR prime editing

Authors:

E. Mansell[1], A. Swift[1], A. Lee[1], E. Winnicki[1], E. Li[1], T. Huynh[1], N. Narisu[1], C. Krilow[1], C. Robertson[2,1], M. Erdos[1], F. Collins[1], L. Bonnycastle[1]; [1]NIH - Natl. Human Genome Res. Inst., Bethesda, MD, [2]Univ. of Michigan, Ann Arbor, MI

CRISPR prime editing (PE) is a transformative gene editing tool that enables precise and versatile genomic modification without inducing double-stranded breaks. PE requires multiple components, including a prime editor enzyme complex, pegRNA (finds the target and provides the new sequence), and sgRNA (directs prime editor complex to nick non-edited strand). Recent studies have identified additional PE components that can improve editing, including PEmax (optimized prime editor complex), epegRNA (structural addition to stabilize pegRNA), and MLH1dn (dominant negative version of mismatch repair protein). Effective design/combinations of these elements are essential for optimal editing efficiency. Here, we use PE to introduce type 2 diabetes (T2D) candidate risk variants (*ABCC8* rs137852671
, *MTNR1B* rs10830963, *ADCY5* rs11708067, *TCF7L2* rs7903146, *HNF4A* rs1800961, *CAMK1D* rs11257655, *CAM K2B* rs878521) into human induced pluripotent stem cells (iPSCs), a cell type which can be differentiated into numerous disease-relevant cell types. We optimized PE systems (components and guide RNA design) across seven T2D risk loci and developed a highly efficient and effective pipeline that generated isogenic lines carrying heterozygous or homozygous risk or non-risk alleles for six of the seven T2D loci. We found that PE can support editing in iPSCs, but optimization of all components is critical to achieve high efficiency. PE systems utilizing PEmax, epegRNA modifications, and MLH1dn provide significant benefit, with maximal editing efficiency ranging from 36-73% depending on the target. We screened for edited cells at two different stages, a primary next-generation sequencing screen of pooled cells to identify transfections with the highest number of edited alleles, and a secondary screen to sequence individual clones. This approach reduced downstream single clone isolation efforts by up to 90%. Editing success and degree of optimization required for each variant differs considerably depending on the sequence at the target site. pegRNA design also plays a critical role, as slight variations in sequence guide parameters can have significant effects. In general, previously established guidelines for pegRNA design should be followed during the initial optimization for each target. After that, making small changes to pegRNA lengths and/or shifting the position of the pegRNA may be required to enhance editing rate to a level that is feasible to screen for edited clones. Although considerable effort is required to optimize PE, it is a promising approach to generate isogenic iPSC lines, enabling the study of specific genetic changes in a common genetic background.

Session Title: Omics Technologies Poster Session III

PB3447 Optimizing Multimodal Single Cell Sequencing to Fine-Map an Alloimmunization Susceptibility Locus in Sickle Cell Disease

Authors:

A. Motter, A. Haldipur, Y. Han, E. Banfield, N. Hanchard; Natl. Human Genome Res. Inst., NIH, Bethesda, MD

Regular transfusion of red blood cells (RBCs) is a mainstay of treatment in sickle cell disease (SCD). Despite this, persistent transfusions increase the risk of developing alloantibodies against donor RBCs, a phenomenon disproportionately exhibited by individuals with SCD. The reason some individuals develop many alloantibodies while most do not is unclear, although host genetic factors are thought to play a prominent role. We have previously identified a novel genetic association between the development of alloantibodies and a genetic variant of African ancestry at chromosome 5q33, further localized to an ENCODE project enhancer annotated to the proximal immunoregulatory genes *ADRB1* and *IL12B*. We proposed to fine-map the risk locus by co-assessing transcription and epigenetic regulation of the candidate genes in risk variant carriers and non-carriers. To facilitate this, we undertook an optimization study to establish an experimental and computational pipeline for single cell RNA and ATAC-sequencing that maximizes cell viability, sequencing depth, and cell type integrity. Peripheral blood mononuclear cells (PBMCs) were isolated from freshly drawn whole blood and subjected to different sample processing- (cell viability enrichment using magnetic levitation vs. no processing) and storage- conditions (freshly isolated vs. cryopreserved), followed by the 10x Genomics Single Cell Multiome ATAC + Gene Expression. We found that cryopreserved, viability-enriched PBMCs provided 97% cell viability post-thaw (3.7 million live cells/ml, 99k dead cells/ml) and retained the highest integrity of cell types compared to the non-frozen control (7% B, 44% CD4+ T, 13% CD8+ T, 1.3% DC, 17% monocytes, 6% NK, 8% other). These results validate pre-sequencing cell viability sorting as an optimal step in real-world single cell application that will allow for efficient processing and analysis of patient samples, whilst providing biologically representative results that can reliably inform multimodal single cell studies.

Session Title: Omics Technologies Poster Session I

PB3448 Overcoming Limitations of RNA-Seq Library Construction from FFPE Samples Using a Novel Workflow

Authors:

C. Jiang, A. Olcott, L. Lingelbach, E. Jordan Dreskin; Bio-Rad Lab., Hercules, CA

Formalin-fixed, paraffin-embedded (FFPE) tumor tissue is a common biospecimen for oncology research and often proposed as a key resource for RNA sequencing (RNA-Seq) to evaluate tumor gene expression signatures. However, RNA purified from FFPE samples typically yields highly degraded material whose poor quality is suboptimal for RNA-Seq library construction. The RNA-Seq workflow can further impair data quality, as traditional library preparation workflows may result in the loss of valuable information due to small fragments being washed away during ribodepletion. Moreover, small, degraded fragments are a challenging template for ligases, and the resulting adapter addition may be inadequate. Here we evaluated a novel workflow that utilizes a proprietary enzyme that bypasses the bias and limitations of reverse transcriptases (RT) and ligases, coupled with a post-library ribodepletion strategy that retains more of the transcriptome. This study also evaluated FFPE RNA extraction methods from breast cancer samples using 2 commercially available kits (RecoverAll and PureLink) to evaluate quality metrics such as RNA integrity number (RIN), DV200, and the recovery of small RNAs during NGS library construction as well as transcriptome coverage. A matching fresh frozen RNA sample was extracted using a phenol-based extraction reagent. Results show that our novel workflow generates a richer data set that represents more of the transcriptome, including both small and long RNAs from FFPE samples. By performing ribodepletion after library preparation, smaller RNAs were detected than when pre-library ribodepletion was used. The excellent analytical performance of this novel library preparation workflow supports its use in RNA-Seq for common archival FFPE specimens in oncology research.

Session Title: Omics Technologies Poster Session II

PB3449 Partitioning the plasma proteome into genetically mediated causal signals and environmental sensors of complex disease risk

Authors:

K. Tsuo[1,2], D. Gadd[3,4], M. Kurki[5,2,6], Z. Zheng[5,2,6], Z. Kuncheva[3,7], D. Baird[8], A. Lehisto[6], R. Marioni[3,4], C. Foley[3,7], H. Runz[8], H. Huang[1,2], B. Sun[8,9], C-Y. Chen[8], M. Daly[1,2,6], A. Martin[1,2]; [1]Broad Inst. of MIT and Harvard, Cambridge, MA, [2]Analytic and Translational Genetics Unit, Massachusetts Gen. Hosp., Boston, MA, [3]Optima Partners, Edinburgh, United Kingdom, [4]Inst. of Genetics and Cancer, Univ. of Edinburgh, Edinburgh, United Kingdom, [5]Broad Inst. of MIT and Harvard, Boston, MA, [6]Inst. for Molecular Med. Finland, Univ. of Helsinki, Helsinki, Finland, [7]Bayes Ctr., The Univ. of Edinburgh, Edinburgh, United Kingdom, [8]Biogen, Cambridge, MA, [9]Dept. of Publ. Hlth.and Primary Care, Univ. of Cambridge, Cambridge, United Kingdom

The plasma proteome uniquely bridges the gap between inherited and modifiable risk factors for disease. Previous disease-specific studies have identified handfuls of proteins that play potential causal roles in mediating genetic effects on disease onset. However, few studies have systematically categorized disease-associated proteins into causal mediators of genetic risk vs. sensors of environmental risk. Here, we used one of the largest proteomic studies to date analyzing ~3K plasma proteins, measured in 54,306 participants in the UK Biobank (UKB), that have been tested for association with time-to-onset for several diseases in previous work. Here, to characterize the proteins significantly associated with incident disease (N=956), we first employed Mendelian randomization (MR) using pQTL instruments from UKB for these proteins and the largest publicly available GWAS of 19 disease outcomes from other studies to assess potential causal roles of these proteins in disease onset. As an example, we observed MR evidence for levels of EFEMP1, a proposed therapeutic target for malignant pleural mesothelioma, on chronic obstructive pulmonary disease (COPD) (β = -0.17, P = $4.3x10^{-5}$). We validated our findings by conducting MR using pQTL instruments from FinnGen for the same proteins and the largest available disease GWAS from UKB and other biobanks. We next evaluated the utility of the proteins without MR signals as predictive biomarkers of disease by applying a decomposition approach not previously applied proteome-wide (termed de-Mendelization), which adjusts for the genetically-mediated variance of the protein levels to improve the observational protein-disease associations. Identifying proteins that could be useful predictors of disease through this approach, we hypothesized that these proteins may serve as quantitative sensors of commonly-measured environmental risk factors and could be more precise predictors of disease than the factors themselves. For example, we found that more than half of the proteins associated with incident disease (576/956) were significantly associated with self-reported smoking status in UKB, and for COPD, after adjusting for smoking, the remaining disease-associated proteins were still significantly associated with smoking. We are comparing prediction models with survey variables of smoking and these proteins to investigate their utility as sensors of cumulative smoking risk. Our study presents a novel strategy that integrates large-scale MR and genetic adjustment to partition protein-disease associations, illustrating another avenue for translating GWAS findings into drug target and biomarker discovery.

Session Title: Omics Technologies Poster Session III

PB3450 Performance of an artificial intelligence-enabled genome sequencing system for universal newborn screening, diagnosis, and precision medicine for 410 severe childhood genetic diseases in over 460,000 individuals.

Authors:

J. Schleit[1,2], M. Wright[1,2], W. Mowrey[3], L. D. Smith[1], C. Kunard[4], M. Bainbridge[1,2], S. Batalov[1,2], E. Blincow[1,2], S. Caylor[1,2], C. Chambers[5], G. del Angel[3], K. Ellsworth[1,2], A. Feigenbaum[1,2,5], E. Frise[6], L. Guidugli[1,2], K. Hall[4], C. Hansen[1,2], C. Hobbs[1,2], M. Kiel[7], L. Van Der Kraan[1,2], C. Krilow[8], Y. H. Kwon[1,2], L. Madhavrao[1,2], J. Le[1,2], J. Leipzig[8], S. Lefebvre[3], Y. Liang[3], R. Mardach[1,2,5], D. Oh[1], L. Olsen[1,2], G. Powley[8], L. Protopsaltis[1,2], M. Saad[1], G. Scharer[1], S. Shelnutt[8], K. Sund[4], M. Tokita[1,2], S. S. Mehtalia[4], S. Papadopoulos[8], E. Sanford[1], S. Schwartz[7], D. Tran[4], M. Reese[6], K. Wigby[5,1,2], M. J. Willis[1], A. Wolen[8], M. Yandell[9], T. Defay[3], S. Kingsmore[1,2,10], M. J. Owen[1,2]; [1]Rady Children's Inst. for Genomic Med., San Diego, CA, [2]Rady Children's Hosp., San Diego, CA, [3]Alexion, Astra Zeneca Rare Disease, Boston, MA, [4]Illumina, Inc., San Diego, CA, [5]Dept. of Pediatrics, Univ. of California San Diego, San Diego, CA, [6]Fabric Genomics, Inc., Oakland, CA, [7]Genomenon Inc., Ann Arbor, MI, [8]TileDB Inc., Cambridge, MA, [9]Univ. of Utah, Salt Lake City, UT, [10]Keck Graduate Inst., Claremont, CA

Newborn screening (NBS) dramatically improves outcomes in selected, severe, childhood disorders by identification and treatment initiation at/before symptom onset. Expansion of the NBS Recommended Uniform Screening Panel (RUSP) has lagged identification and approval of effective therapeutic interventions leading to unnecessarily delayed diagnosis and suboptimal outcomes in hundreds of severe, early childhood onset single locus (mendelian) genetic diseases. BeginNGS supplements RUSP-NBS for these diseases by combining NBS-by genome sequencing (GS) with automated interpretation and virtual guidance regarding confirmatory testing, specialist referral, and urgent management. While clinical development is incomplete, here we report results of the second retrospective study of BeginNGS. Of 518 gene-disorder dyads evaluated for BeginNGS using modified Wilson and Jungner principles and Delphi methods, 410 disorders associated with 341 genes and 1,654 effective therapeutic interventions were retained, with a total incidence in newborns of ~5%. The analytic performance of BeginNGS is trained with genome sequences of large retrospective true positive (TP) and true negative (TN) cohorts. Previously we reported training of 29,865 pathogenic (P) or likely P (LP) variants associated with 388 disorders in 459,083 TP and FN individuals. Following training, BeginNGS had 3 FP/1,000 subjects and TP rate (sensitivity) of 88.8% (119 of 134 diagnoses). Here we will report results of BeginNGS version 2 with 410 disorders, ~50,000 variants from both ClinVar and Genomenon's Mastermind, as well as new artificial intelligence software for detection of novel loss of function variants (Transformer) in a larger cohort including 7,575 critically ill children with suspected genetic disorders and their parents. We will also report results of an actual/counterfactual comparison of the clinical utility, average cost per diagnosis, and incremental cost cost effectiveness ratio of rapid diagnostic GS ($7,000 cost per newborn with return of results during an intensive care unit admission) with BeginNGS ($500 with results on day of life (DOL) 10) and RUSP-NBS ($211 with results on DOL 10).

Session Title: Omics Technologies Poster Session I

PB3451 Pipeline for improved genotype imputation using blended genome-exome sequencing and a diverse reference panel for large-scale population studies

Authors:

M. Gatzen, C. Kachulis, C. Patterson, M. Coole, E. Malolepsza, M. Shand, J. C. Meldrim, M. DeFelice, N. J. Lennon; Broad Inst. of MIT and Harvard, Cambridge, MA

Recent advancements in sequencing technologies have provided the research community unprecedented opportunities for conducting large-scale population studies to understand the genetic underpinnings of disease. Whole genome sequencing (WGS) remains the gold standard for genetic studies, but even though it has become more affordable the relatively high cost remains a barrier to the feasibility of many population studies. Whole exome sequencing (WES) is a more affordable option, however, the shortcomings of being blind to significant portions of the genome may be prohibitive for certain research questions. Imputation from genotyping arrays provides a bridge between affordability and information about large regions of the genome, however, the limitation of only being able to capture predefined alleles results in reduced applicability to diverse populations and disease characteristics. We recently launched blended genome-exome (BGE) sequencing which provides full-depth coverage over the exome and low-pass coverage over the entire genome. The genome portion is a basis for imputation that is superior to arrays while the exome portion provides valuable high-confidence exome calls. We evaluate the imputation accuracy by comparing genotypes imputed from BGE data sequenced on the Illumina NovaSeqX to matched WGS data for a cohort of diverse ancestries. We find that using BGE as an input to imputation results in superior accuracy to existing genotyping arrays while allowing for leveraging arbitrary reference panels due to BGE sequencing not being limited to certain genotyping sites. Furthermore, we updated the reference panel to use the 1000 Genomes Project + Human Genome Diversity Project panel provided by gnomAD, which adds additional 828 samples from 54 populations to the 1000G cohort. After removing singleton sites this panel increases the number of covered sites by 91% as compared to the commonly used 1000 Genomes Project panel. As a result, the coverage of sites used for polygenic risk scores of 10 chronic medical conditions implemented by the NHGRI eMERGE network in a recent clinical study increases from 99.3% to 99.8%. Leveraging the GLIMPSE imputation tool we created a cloud-native imputation pipeline that is optimized for cost-effective processing of large-scale cohorts. The combination of BGE data as an input for imputation with an improved, more diverse reference panel significantly improves the accuracy of results as compared with current approaches. Combined with the high-confidence exome calling this method provides a cost-effective and accurate solution for population genetics studies without the need for multiple analysis modalities.

Session Title: Omics Technologies Poster Session II

PB3452 Poverty shapes the transcriptome of immune cells.

Authors:

N. Arnold[1,2], J. Resztak[2], A. Alazizi[2], D. Witonsky[2], N. Noren Hooten[3], M. Evans[3], V. Odera-Marah[1], D. Dluzen[4], R. Pique-Regi[2], F. Luca[2]; [1]Morgan State Univ., Baltimore, MD, [2]Wayne State Univ., Detroit, MI, [3]Natl. Inst. on Aging, NIH, Baltimore, MD, [4]Johns Hopkins Univ., Baltimore, MD

Psychosocial factors exert a powerful influence on health and contribute to disparities among marginalized populations. For example, overall life expectancy at birth throughout the United States tracks with poverty level, educational attainment, economic security and other upstream social determinants of health. Socioeconomic status (SES) and psychosocial factors may affect gene expression in peripheral blood mononuclear cells, suggesting a molecular mechanism for some health disparities. Here we investigated the effects of poverty among participants in the Healthy Aging in Neighborhoods of Diversity across the Life Span (HANDLS) study. We examined 204 participants whose reported household income was either above or below poverty. The population sample was balanced by self-reported race and sex. We performed RNA sequencing in peripheral blood mononuclear cells to assess differential gene expression patterns associated with poverty. We found 138 genes differentially expressed between individuals living in poverty compared to those living above poverty. We found 104 differentially expressed genes in women, however, we did not observe any genes differential expressed in men living in poverty. Differentially expressed genes in individuals living in poverty are enriched in processes related to wound healing and coagulation. Of the genes differentially expressed in individuals living in poverty, *VIL1* and *EEF1DP7* are also associated with hypertension in transcriptome-wide association studies. Our results suggest that the impact of poverty on gene expression varies between men and women. We show that poverty may affect wound healing and coagulation processes by causing changes in gene expression in immune cells.

Session Title: Omics Technologies Poster Session III

PB3453 Predicting chromatin interactions and gene expression by artificial intelligence methods

Authors:

M. Chilinski[1,2], J. Lipiński[3], A. Agarwal[2], Y. Ruan[4,5], D. Plewczynski[1,2]; [1]Warsaw Univ. of Technology, Warsaw, Poland, [2]Univ. of Warsaw, Warsaw, Poland, [3]Cellular Genomics, Warsaw, Poland, [4]The Jackson Lab. for Genomic Med., Farmington, CT, [5]Life Sci. Inst., Zhejiang Univ., Hangzhou, China

In our work, we have developed 3DNBERT - a Artificial Intelligence (AI) model for the prediction of the HiChIP interactions. We have used deep learning model, that is composed of the BERT architecture as well as transformer encoder for the ChIP-Seq and ATAC-Seq signals. Our pipeline first trains ChIP-BERT for detection of the ChIP-Seq and ATAC-Seq signals, then 3DN-BERT in two cases - one for prediction of the anchors, and one for predicting which anchors will be connected together.
The final model is trained on all chromosomes and experimental epigenetic profiles, except for the chromosome 9 that is used for validation purposes. Our learning architecture firstly predicts the epigenomic signal in the testing chromosome, then predicting the potential anchors, and finally identifying which anchors will form chromatin loops with each other. We are predicting from DNA sequence all the downstream genomic information - as we are using SNPs, indels and structural variants (SVs) for creating of the personalised and phased genomes, that are used as input to the network.
We have conducted additional research, which the extend the spatial structure in silico identification with the gene expression prediction. This time we used convolutional neural network as in ExPecto - improved by including the spatial information. This allowed us to increase the expression prediction accurancy using various ChIA-PET experimental datasets. We have obtained an improvement of the spearman correlation score of up to 0.042 (in case of RNAPOL2 ChIA-PET GM12878).
Concluding, our results show that the modern Artificial Intelligence algorithms are capable of predicting the chromatin spatial interactions from the DNA sequence, and those interactions can be further used to improve the gene expression prediction. Our research is one of first steps in connecting DNA sequence, spatial organisation of the chromatin, and the gene expression in quantitative form.

Session Title: Omics Technologies Poster Session I

PB3454 Profiling of stranded RNA biotypes during early mouse gonad development

Authors:

F. Ou[1], H. SUN[2], Z. Wang[1], K. Miu[1], P. Shi[2], W-Y. Chan[1]; [1]The Chinese Univ. of Hong Kong, Hong Kong, China, [2]The City Univ. of Hong Kong, Hong Kong, China

Gonad development involves complex spatial and temporal genetic and epigenetic regulations. While several scRNA-Seq and scATAC-Seq investigations have provided refined gonadal transcriptome and chromatin accessibility profiles of different species, the functional networks of non-coding RNAs in gonad development remain largely unknown. This study performed the first complete stranded RNA-Seq to profile all the stranded RNA biotypes from mouse male and female gonads at four different time points. 2419 differentially expressed genes (DEGs) were identified from E12.5-vs-E11.5 with cut-offs as $|LogFC| \geq 1$ and $FDR \leq 0.05$, and 333 and 770 DEGs were identified from E13.5-vs-E12.5 and E14.5-vs-E13.5, respectively. microRNA, lincRNA, and antisense were the major differentially expressed RNA biotypes in addition to protein-coding mRNA from the datasets. RT-qPCR and RNA spatial assay further validated the spatial and temporal expression of selected DE miRNAs. miR202 showed dramatic expression fold change with a low expression level, which is consistent with the previous report by Wainwright EN et al. in 2013. miR6236 was the most highly expressed miRNA and showed significant downregulation from E11.5 to E12.5 and then significant upregulation from E13.5 to E14.5, suggesting that mmu-miR-6236 plays important roles in early sex determination and germ cell development initiation with different pathways. Target identification suggests that miR6236 interacts with lincRNA Miat and promotes Foxd3 expression, which may further contribute to cell-type specific transcription factors chromatin reconstruction.

Session Title: Omics Technologies Poster Session II

PB3455 Protein Detection Correlations: Examining the Correlation of Protein Detection in High-throughput Proteomics Platforms in Million Veterans Program Proteomics Study.

Authors:

C. Yandava[1], P. Schreiner[2,1], D. Panyard[3,4], P. Devineni[1], B. Gorman[2,1], P. Tsao[5,6], S. Pyarajan[1,7]; [1]VA Boston Hlth.care System, Boston, MA, [2]Booz Allen Hamilton, Lexington, MA, [3]Stanford Univ., Sunnyvale, CA, [4]VA Palo Alto Hlth.Care System, Palo Alto, CA, [5]VAPAHCS/Stanford Univ., Palo Alto, CA, [6]Stanford Univ. Sch. of Med., Stanford, CA, [7]Brigham Women's Hosp., Boston, MA

Proteins carry out the major functions of an organism and are involved in structural, enzymatic, inter and intracellular communications, immunity and many other functions. Protein levels are indicators of the health of an organism and used for diagnostic, prognostic, and therapeutic purpose. The Million Veteran Program (MVP) at VA is a national research program to understand the genetic basis for the illnesses of veterans and is a great resource for clinical data in diverse and minority ethnic groups. We measured the plasma protein levels of about 900 individuals, using Olink Explore 1536 and Somascan v4, comprising both a control group and those with diseases, using both Somascan and Olink platforms. The Olink Explore 1536 platform profiles 1472 proteins of which 1463 are unique and correlations among proteins of this platform are studied. Three proteins of these proteins - *IL6*, *CXCL8* and *TNF* - are replicated in 4 distinct detection panels. These replicate proteins can be used to perform QC based on their correlations of log2 transformed Normalized Protein eXpression (NPX). There are very high correlations of the same protein profiled across all the panels. *CXCL8* has the highest correlation coefficients (between 0.94 and 0.99) between panels, followed by *IL6* (0.9 to 0.95) and *TNF* had the lowest correlation between 0.8 and 0.87. Apart from the replicates, there are 3337 pairs (381 proteins) with r values over 0.8. We evaluated the relationship of these proteins that could possibly explain coordinated expression for functional roles or indicate technical issues such as cross reactivity due to sequence similarity. One pair of proteins, *AMY2a* and *AMY2b* perform the enzymatic reaction of hydrolyze 1,4-alpha-glucoside bonds in oligosaccharides and polysaccharides. They are 98.83% identical and 99.02% similar in their sequence. Their detection is very highly correlated (0.988). Another pair of proteins with high correlation (0.95) is with proteins *PILRA* (Paired immunoglobulin-like type 2 receptor alpha) and *PILRB* (Paired immunoglobulin-like type 2 receptor beta). These proteins are not very identical (61.76%) or similar (66.01%) in their sequence. *PILRA* is an inhibitory receptor and *PILRB* is an activating receptor and cell signaling is balanced by the activity of these proteins. It is reasonable to suggest that these 2 proteins necessarily have similar protein levels to perform their function. It is possible that the reagents (antibodies) may cross-react or that the genes are locally co-expressed. These 2 pairs of proteins are not correlated with any other protein. The implications of protein correlations and disease state are investigated.

Session Title: Omics Technologies Poster Session III

PB3456 Proteomics of Urinary Exosomes Reveals Impaired Protein Trafficking in HPS-1 Renal Epithelial Cells.

Authors:

D. Maynard[1], B. R. Gochuico[1], H. Pri Chen[1], C. K. E. Bleck[2], P. Zerfas[3], W. A. Gahl[1], M. C. V. Malicdan[1]; [1]NHGRI (NIH), Bethesda, MD, [2]NHLBI (NIH), Bethesda, MD, [3]OD (NIH), Bethesda, MD

Background: Hermansky-Pudlak syndrome type 1 (HPS-1) is a rare, autosomal recessive disorder caused by defects in BLOC-3 (biogenesis of lysosome-related organelles complex), leading to oculocutaneous albinism, a bleeding diathesis, granulomatous colitis, and pulmonary fibrosis (PF). Renal involvement in HPS-1 patients remains poorly understood. Exosomes, small endocytic vesicles (40-100 nm), play a crucial role in various cellular functions; urinary exosomes (UrEx) contain proteins from renal tubule epithelial cells and provide a non-invasive method for studying changes in protein expression. In this study, we investigated the protein composition of UrEx obtained from HPS-1 patients compared to controls to understand the mechanisms underlying renal manifestations in HPS-1. Methods: Twenty-four-hour urine samples were collected from five HPS-1 patients and five unaffected controls. Exosomes were pelleted by ultracentrifugation and treated to remove Tamm-Horsfall protein (THP), a major contaminant. 1D-gel electrophoresis and LC-MS/MS were used to analyze UrEx protein composition. ELISA was utilized to measure serum ApoA1 levels in a subset of HPS-1 patients enrolled in a natural history protocol. We generated a murine *Hps1* knockout using genomic engineering to investigate the impact of HPS1 deficiency on kidney function. Results: THP-depleted UrEx exhibited intact morphology and expected size. Proteomic analysis identified 1029 proteins, including those involved in kidney disease and fibrosis. Ingenuity Pathway Analysis revealed disruptions in mitochondrial proteins and the LXR/RXR pathway. Our *Hps1* knockout mice exhibited characteristic deposits and vacuolation in the proximal tubule epithelial cells, along with distinct morphological differences in ultrastructure. Compared to controls and HPS-3/HPS-5 patients, HPS-1 patients had decreased serum ApoA1 concentrations. Moreover, serum ApoA1 levels positively correlated with estimated glomerular filtration rate (eGFR), an indicator of kidney function. Conclusions: We optimized the preparations of THP-depleted UrEx and identified protein abundance differences in UrEx from HPS-1 patients compared to control, implicating mitochondrial dysfunction, alterations in lipid/lipofuscin metabolism, and kidney fibrosis. These findings contribute to our understanding of the pathophysiology of HPS-1.

Session Title: Omics Technologies Poster Session I

PB3457 Quality assessment of DNA and RNA for NGS libraries using Agilent Automated Electrophoresis Systems.

Authors:

M. Meister[1], **S. Siembieda**[2]; [1]Agilent Technologies, Waldbronn, Germany, [2]Agilent Technologies, Inc., Ankeny, IA

DNA and RNA sequencing enables the study of the genomic and transcriptomic background of cells or tissues and has become an established part of scientists' toolkit for cancer research. Critical for robust results of next-generation sequencing (NGS) applications is the accurate qualification and quantification of the samples used. To compare the performance of automated electrophoresis systems such as the Agilent 2100 Bioanalyzer, TapeStation and Fragment Analyzer systems, equivalent assays for the respective systems were used for quantitative and qualitative analysis of DNA libraries and RNA samples. For NGS, it is important to ensure that the starting material has been fragmented into DNA molecules of a size suitable for the sequencer and that the library contains enough material for sequencing. The final libraries were prepared and subsequently analyzed on the Bioanalyzer, TapeStation, and Fragment Analyzer systems. Each system reported comparable smear size and concentration. For RNA sequencing, the input quality of the RNA material is critical for obtaining reliable results. Total RNA samples were analyzed with the Bioanalyzer and TapeStation systems and showed similar concentrations and quality metrics on both systems. Finally, the analysis of total RNA with the Bioanalyzer and the Fragment Analyzer systems provided comparative RIN and RQN results as well as similar concentration values for the two systems. The RNA quality scores, as well as the quantification and electropherogram results obtained on the Bioanalyzer system were comparable to the results obtained with the TapeStation or Fragment Analyzer systems, illustrating the equivalent performance of the Agilent automated electrophoresis systems.

Session Title: Omics Technologies Poster Session II

PB3458 Quantification of the escape from X chromosome inactivation with the million cell-scale human blood single-cell RNA-seq datasets reveals heterogeneity of escape across immune cells

Authors:

Y. Tomofuji[1], R. Edahiro[1], Y. Shirai[1], K. Sonehara[2], Q. Wang[3], Asian Immune Diversity Network, A. Kumanogoh[1], Y. Okada[4]; [1]Osaka Univ. Graduate Sch. of Med., Suita, Japan, [2]Graduate Sch. of Med., the Univ. of Tokyo, Bunkyo-ku, Japan, [3]Osaka Univ., Osaka, Japan, [4]Osaka Univ. / The Univ. Tokyo / RIKEN, Osaka, Japan

One of the two X chromosomes of females is silenced through X chromosome inactivation (XCI) to compensate for the difference in the dosage between sexes. There exist several genes which escape from XCI, which could contribute to the differential gene expression between sexes. The differences in the escape across cell types are still poorly characterized, especially across immune cell types, even though a large number of immune-related genes are on the X chromosome. Escape in disease conditions and non-European people have also been poorly investigated. scRNA-seq analysis is useful for investigating the escape. However, most current methods require plate-based scRNA-seq data. To investigate escape with a wide variety of datasets, we need a new method to evaluate the escape with widely used droplet-based scRNA-seq technology. Here, we investigated the escape across immune cell types utilizing the largest scale scRNA-seq datasets (1) ~1,000,000 peripheral blood mononuclear cells (PBMCs) from 489 East Asian participants, (2) ~1,000,000 PBMCs from 147 Japanese (72 COVID19 patients and 75 healthy participants) (3) ~1,000,000 PBMCs from 162 systematic lupus erythematosus patients and 99 healthy participants including East Asian and European. Differential expression gene (DEG) analysis revealed that female-biased expression of escapee genes was stronger in lymphoid cells than in myeloid cells (median log2 fold change was 0.35 in CD4+ T cells and 0.15 in monocytes, P = 0.018). To investigate whether these differences originated from the differences in the escape, we developed a method, **s**ingle-**c**ell **L**evel **ina**ctivated **X** chromosome mapping (**scLinaX**), which directly quantify relative gene expression from the inactivated X chromosome (Xi) with droplet-based scRNA-seq data. The scLinaX successfully detected the escape for the previously reported escapees (median ratios of expression from Xi were 0.24 and 0.013, respectively for escapees and genes subjected to XCI; $P = 3.1 \times 10^{-14}$). The scLinaX revealed the stronger degree of escape in lymphocytes than in myeloid cells suggesting that difference in escape makes the heterogeneity of sex-associated differential gene expression across cell types. We also found several genes which had female-biased expression while showing little evidence of the escape in scLinaX analysis, emphasizing the limitation of DEG analysis and the importance of complementary analysis, such as the scLinaX. Finally, we systematically evaluated the association between diseases and escape. The scLinaX is applicable to a wide variety of scRNA-seq datasets. This study would contribute to understanding of the XCI and escape.

Session Title: Omics Technologies Poster Session III

PB3459 Quantitative profiling of hepatitis B virus transcripts in human liver transcriptome data

Authors:

Q. Cao[1], W. Cui[2]; [1]The first affiliated Hosp., Zhejiang Univ. school of medicine, Hangzhou, China, [2]Zhejiang Univ., Hangzhou, China

The existence of viral sequence in human High-Throughput Sequencing (HTS) data, potentially indicating viral infection in human cells, has been recognized for some time. Some studies have utilized the detection of these viral sequences to infer the viral infection status of the original sample. However, a quantitative analysis of such viral sequences in transcriptome data, based on which the host-virus interaction could be explored, is still lacking. Hepatitis B Virus (HBV) is a hepatotropic DNA virus, and in certain East Asian countries, chronic HBV infection remains the primary etiology of hepatocellular carcinoma. We performed de novo analysis of RNA-seq data from over 200 liver tissue samples, including those obtained from public domain database as well as in-house collected samples. By linearizing the circular genome of HBV based on transcriptional start/stop sites, incorporating it to the reference sequence set, and employing alignment-free transcript quantification algorithms, viral transcripts were quantified and distinguished in a manner similar to human transcripts. We observed the presence of highly abundant viral transcripts specific to HBV infected samples. The abundance of these viral transcripts correlated with the clinical infection phenotypes, as well as the expression of cell cycle and proliferation related human genes. Individuals with high abundance of viral transcripts also showed elevated levels of viral-human gene fusions and mutation rates. Furthermore, in nearly 70% of infected samples, the complete HBV genome could be assembled solely from viral fragments obtained from the human transcriptome. These results demonstrate that quantitative analysis of HBV sequences in liver transcriptome data provides a unique perspective for studying the pathogenic interactions between HBV and the human host. They also highlight the significance of exploring viral information within host transcriptome data as a research avenue.

Session Title: Omics Technologies Poster Session I

PB3460 Rapid single -omic interpretation and analysis is a precondition for synthesis of large scale multiomics conclusions

Authors:

A. O'Hara, P. Vishwanath, E. Louie, D. Schneweis, L. Turner, H. Latif, C. Mozdzierz, G. Zhou; Azenta Life Sci., Burlington, MA

The -omics era expands the repertoire of approaches available for researchers and clinicians to unravel human health's complexity. Next Generation Sequencing (NGS) approaches characterize genomes, epigenomes, transcriptomes, and proteomes from a single sample source. Peripheral blood mononuclear cells (PBMCs) offer a window into the immune system that, when combined with omics tools, can provide a near holistic view of immune processes across patient cohorts. This proof-of-principle workflow utilizes a single blood draw to rapidly produce a diverse set of datatypes interrogating the genome, epigenome, transcriptome, and proteome. In 10 Hispanic men aged 50-60, five with diagnosed cardiovascular disease (CVD) and five unaffected healthy individuals, a single blood draw using a heparin tube was collected and PBMCs were isolated within 24 hours of collection to ensure high viability and yield of PBMCs, along with simultaneous plasma separation and collection. These samples were then aliquoted and simultaneously processed for whole exome sequencing, single cell RNA sequencing (scRNA-Seq), epigenetic characterization, and Olink proteomic assays.
Here we present the scRNA-Seq results using the rapid first look analysis tool known as the ROSALIND platform. ROSALIND is an interactive sequencing analysis platform, offering a powerful analysis rich downstream method connecting experimental design, quality control and pathway exploration. scRNA cluster analysis was utilized for grouping of cell types both within and between samples to identify trends and differences between the affected and control individuals. Comparative analysis between the CVD affected and healthy control individuals indicate several statistically significant differentially expressed RNA transcripts and pathways. Conclusions from this standalone analysis are then aggregated into downstream multiomics analyses, including Olink proteomics, showing that integrative downstream analyses can drive greater insights and innovation in human health applications.

Session Title: Omics Technologies Poster Session II

PB3461 RDAP: a web-based pipeline to analyze RNA-seq data

Authors:

B. Lee; Korean Bioinformation Ctr., Daejeon, Korea, Republic of

The complexity of transcriptomes and their regulative pathways make RNA-Seq one of the most complex NGS applications, addressing several aspects of the expression process. RDAP (RNA-seq Data Analysis Pipeline) is an integrated transcriptome analysis service allowing users to analyze their own RNA-seq data. RDAP consists of analytical pipelines covering the entire spectrum of RNA-seq analysis, quality assessment, aligning reads, differential gene expression analysis, and visualization. RDAP is a fully automated and user-friendly cloud-based platform that does not require application installation or testing steps. Instead, the user simply uploads RNA-seq data and chooses an appropriate analysis pipeline and parameters. RDAP aims at a broad group of users, from beginners with little command-line experience to advanced users with the ability to customize the parameters used by their chosen tools.

Session Title: Omics Technologies Poster Session III

PB3462 Refining Telomere-to-Telomere Human Genome References: Insights from a novel high accuracy sequencing-by-binding platform

Authors:

M. Fleharty, L. Fleming, M. Coole, S. Low; Broad Inst., Cambridge, MA

Advancements in genome sequencing technologies have enabled more accurate and comprehensive representations of the human genome. We used a high-fidelity sequencing platform which leverages a novel sequencing-by-binding chemistry (Onso, Pacific Biosciences, CA) to further refine two Telomere-to-Telomere (T2T) human genome references, CHM13-T2T and HG002-T2T, with a focus on reducing homopolymer errors.

For CHM13, the sequencing-by-binding data showcased superior homopolymer accuracy relative to other technologies, thus enabling the identification and potential amendment of approximately 12,000 homopolymer sites in the CHM13-T2T reference. This represents a significant step towards further improving the already high-quality CHM13 reference.

In parallel, we sequenced the Ashkenazi trio (NIST samples HG002, HG003, and HG004) using the same sequencing platform. Here, we employed a unique approach that leveraged pedigree information alongside Mendelian inheritance assumptions to spotlight regions of the HG002-T2T reference that could benefit from further refinement. Our results are highly concordant with orthogonal techniques to identify homopolymer errors using assembly-based phasing. Our results have been submitted to the T2T-Consortium as proposed polishing sites for inclusion in the HG002-T2T v0.8 reference.

Our work collectively exhibits how novel sequencing chemistries combined with novel analytical strategies can significantly contribute to improving the fidelity of human genome references. These enhancements will invariably facilitate more accurate genetic research, contributing to the development of personalized treatments, better understanding of genetic diseases, and the exploration of human genetic diversity.

Session Title: Omics Technologies Poster Session I

PB3463 RNA-seq data analysis identifies blood-based biomarkers for diagnosis and disease progression of Alzheimer's disease.

Authors:

A. Yamakawa[1], R. Mitsumori[1], M. Suganuma[1], S. Akiyama[1], S. Niida[2], K. Ozaki[1,3], D. Shigemizu[1,3]; [1]Med. Genome Ctr., Res. Inst., Natl. Ctr. for Geriatrics and Gerontology, Obu, Aichi, Japan, [2]Core Facility Admin., Res. Inst., Natl. Ctr. for Geriatrics and Gerontology, Obu, Aichi, Japan, [3]RIKEN Ctr. for Integrative Med. Sci., Yokohama, Kanagawa, Japan

Mild cognitive impairment (MCI) patients have an increased risk of developing Alzheimer's disease (AD). Biomarkers for early detection of individuals at high risk for MCI-to-AD conversion are required. Here, we performed RNA-sequencing data analysis of 1227 Japanese blood samples, representing 424 AD patients, 543 MCI subjects, and 260 cognitively normal (CN) adults, to identify blood-based biomarkers associated with MCI-to-AD conversion. Differentially expressed genes (DEGs) between CNs and MCIs and between MCIs and ADs were detected, and subsequent gene set enrichment analysis using the DEGs revealed biomarker candidates for the disease diagnosis. The DEGs were enriched in 4 pathways (coronavirus disease, ribosome, salmonella infection, and ubiquitin mediated proteolysis) for CNs and MCIs, where 54 unique genes were used: 6 pathways (phagosome, salmonella infection, yersinia infection, necroptosis, coronavirus disease, and pathogenic *Escherichia coli* infection) for MCIs and ADs, where 73 unique genes were used. Retrospective prediction models were constructed with logistic regression by using the combination of these unique genes. The best model in CNs and MCIs (CN-MCI) achieved an AUC of 0.866 (95% confidence interval [CI]: 0.861 to 0.871) when using 12 candidates: MCIs and ADs (MCI-AD), AUC = 0.763 (95% CI: 0.757 to 0.767) with 8 candidates. Common two genes (*CDC42: Cell division cycle 42* and *SKP1: S-phase kinase associated protein 1*) were used in both retrospective prediction models of CN-MCI and MCI-AD. *CDC42* protein levels are increased in the frontal cortex of AD patients, whereas *SKP1* levels are decreased in patients with sporadic Parkinson's disease. Based on these two candidates, we further constructed a prospective prediction model with a Cox proportional hazards model by using 318 MCI cohort data followed over one year (175 MCI patients converted to AD and the remaining 143 patients remained stable with MCI). The Kaplan-Meier curves showed a statistically significance in the log rank trend test (P<0.001) for AD conversion-free survival. These results suggest that *CDC42* and *SKP1* can be potential biomarkers for early diagnosis of AD. Further understanding of these potential biomarkers will contribute to the elucidation of the AD pathogenesis mechanism, and integrated analysis including clinical and genomic information may facilitate clinical application of early diagnosis of AD.

Session Title: Omics Technologies Poster Session II

PB3464 RNAseq driven diagnosis of *NBAS* deficiency expands the phenotypic spectrum of disease.

Authors:

S. Silverstein[1,2], T. Cassini[3], J. Fu[1], B. Pusey[1], E. Macnamara[1], C. Tifft[1], M. Malicdan[1], Undiagnosed Diseases Network, D. Adams[1]; [1]NHGRI, Bethesda, MD, [2]NINDS, Bethesda, MD, [3]Vanderbilt Univ. Med. Ctr., Nashville, TN

Rare diseases affect 1 in 10 individuals, and for those evaluated with exome and genome sequencing, typical diagnostic rates average 30%. RNA sequencing can be used to augment genome analysis and improve diagnosis. Open-source computational tools such as OUTRIDER provide a standard framework for RNAseq analysis and may uncover diagnoses missed by genome driven approaches.

We present a young woman with global developmental delay, poor growth, characteristic facial features, osteopenia, premature ovarian insufficiency, and ocular abnormalities who had non-diagnostic genome sequencing. RNAseq performed on patient derived skin fibroblasts was included in a cohort of 72 undiagnosed patients and run through OUTRIDER. *NBAS* gene expression was reported as significantly reduced for this patient compared with controls. Manual inspection of the BAM files revealed compound heterozygous variants: a rare deep intronic variant NM_015909.4 c.2423+403G > C which creates a hypomorphic pseudoexon not seen in control samples (gnomad AF 0.000006572) and a rare premature termination codon (PTC) c.4753C > T; p.Arg1585Ter (gnomad AF 0.000006572). Both variants are predicted to cause nonsense mediated decay of transcripts, as the pseudoexon contains a PTC.

Biallelic variants in *NBAS* are associated with two phenotypes: infantile liver failure syndrome 2 (MIM # 616483) and short stature, optic nerve atrophy, and Pelger-Huet anomaly (MIM # 614800). More recent literature has detailed a wide clinical spectrum of disease. Re-evaluation of the patient's clinical presentation revealed that it is consistent with the *NBAS* phenotypic spectrum, in a more severe form than previously reported. Published cases of *NBAS* related disease are caused by either biallelic missense variants or a missense and a PTC. To our knowledge this would be the first report of a patient with biallelic PTC variants in *NBAS*. We hypothesize based on reported genotype-phenotype data that biallelic missense variants likely cause infantile liver failure while PTC variants likely cause developmental defects, explaining our patient's unusually severe form of the disease.

We provide an example demonstrating the utility of incorporating early use of RNAseq to generate diagnostic candidates. RNAseq has the potential to further expand genotype-first molecular diagnostics, helping to extend the phenotypic spectrum of identified diseases. We recommend consideration of this approach as part of diagnostic pipelines, especially for exome or genome negative patients.

Session Title: Omics Technologies Poster Session III

PB3465 Saliva as a sample type for the evaluation of miRNA biomarkers using small RNA sequencing

Authors:

C. Dillane, M. Tayeb; DNA Genotek, Ottawa, ON, Canada

Sequencing of salivary microRNA (miRNA) is rapidly becoming a means to explore disease biomarkers. In this study we demonstrate that stabilized saliva samples provide RNA of sufficient quality for detection of miRNA biomarkers including miRNAs with reported links to cancer, head injury and mental health disorders.
Saliva was collected from 14 participants using the ORAcollect™·RNA device (DNA Genotek). RNA was extracted using the miRNeasy Serum and Plasma kit (QIAGEN) and sequencing libraries were generated with the QIAseq miRNA Library Kit with UDIs (QIAGEN) without normalization of the input RNA. XpressRef Universal Human RNA Reference (QIAGEN) was used as a positive control. Libraries were sequenced on the NextSeq 550 (Illumina) targeting a mean of 20 million reads per sample. FastQ files were uploaded to QIAGEN GeneGlobe Design & Analysis Hub and annotated using miRbase and piRNAdb databases.
The mean number of reads per sample was 20.7 million and Q scores of Unique Molecular Index (UMI) reads were above 30 demonstrating a high sequencing performance in all libraries.
On average 47% of the reads per sample did not contain a UMI; observed across both the saliva and control samples. Of the reads containing UMIs, a range of 10-51% of the reads in each sample were annotated with miRBase or piRNAdb, while the XpressRef Control contained 69% of annotated UMI reads. The range of annotated reads in saliva samples is expected given the variable miRNA input that is inherent of saliva samples.
Notably, saliva samples covered 35-50% of the total number of Annotation Records in miRbase despite the variable input of miRNA into library prep. Comparatively the XpressRef RNA Control prepared from 20 different human adult and fetal major organs, contained 58% of the total Annotation Records.
Recently a retrospective cross-sectional miRNA sequencing study utilizing 1225 saliva samples was published and gives greater insight into the expected expression of salivary miRNA's from multiple cohorts (Sullivan R. et al., 2022). We used a cohort of expression profiles from the Sullivan et al. study and compared them to our data to discern concordance between the two. After accounting for sequencing depth we found a 3.5-fold increase in annotated reads in our samples, demonstrating that the QIAseq miRNA Library Kit with UDIs performs well with saliva samples collected using the ORAcollect™·RNA device. The most highly-expressed miRNAs in each study were distinctive, though there was a 50% concordance between the top 20 most-expressed miRNAs in the two studies.
These results demonstrate that saliva is a suitable sample type for miRNA biomarker detection using small RNA sequencing.

Session Title: Omics Technologies Poster Session I

PB3466 Scalable Nanopore long-read sequencing to resolve complex regions in Parkinson's disease

Authors:

A. Miano-Burkhardt, P. Alvarez Jerez, K. Daida, L. Malik, M. Abdelhalim, B. Baker, C. Blauwendraat, A. Singleton, K. Billingsley; NIH, Bethesda, MD

Parkinson’s disease (PD) is the second most common neurodegenerative disorder worldwide, and genetics is one of the major underlying causes of this disease. The majority of PD genetic studies to date have focused on single nucleotide variants. However, current risk loci only explain 16-30% of the heritability of PD, leaving a large portion of the genes driving PD unknown. Structural variants contribute substantially to genetic variation in the human genome, yet little is known about the role they play in PD. Long-read sequencing is required to accurately identify these complex regions of the human genome, but until recently, it was expensive, error-prone, and low throughput. Here, we take advantage of recent advancements in long-read technology and develop a scalable wet lab protocol for Oxford Nanopore Technologies (ONT) long-read sequencing of whole human blood samples at the population level (dx.doi.org/10.17504/protocols.io.ewov1n93ygr2/v1). This protocol yields high-quality ONT data with average N50s of approximately 30kb and 30X coverage. As part of the Global Parkinson's Genetics Program (GP2), we applied this framework to hundreds of PD samples. This new genetic resource will allow us to not only resolve existing complex PD risk haplotypes but also unlock the impact of novel genetic variants that have been previously inaccessible using short-read and array-based datasets.

Session Title: Omics Technologies Poster Session II

PB3467 Scalable Nanopore sequencing of human genomes provides a comprehensive view of haplotype-resolved variation and methylation

Authors:

M. Kolmogorov[1], K. Billingsley[1], M. Mastoras[2], M. Meredith[2], J. Monlong[3], R. Lorig-Roach[2], M. Asri[2], P. Jerez[4], L. Malik[4], D. Ramita[1], X. Reed[5], R. Genner[6], D. Kensuke[1], S. Behera[7], K. Shafin[8], T. Pesout[9], T. Pesout[10], J. Prabakaran[1], P. Carnevali[11], J. Yang[12], A. Rhie[1], S. Scholz[4], B. Traynor[13], K. Miga[14], M. Jain[15], W. Timp[16], A. Phillippy[17], M. Chaisson[12], F. Sedlazeck[18], C. Blauwendraat[1], B. Paten[14]; [1]NIH, Bethesda, MD, [2]Univ. of California, Santa Cruz, Santa Cruz, CA, [3]UC Santa Cruz, Santa Cruz, CA, [4]NIH, Bethesda, MD, [5]The Johns Hopkins Univ. Sch. of Med., Baltimore, MD, [6]Georgetown Univ., Washington, DC, [7]Baylor Coll. of Med., Houston, TX, [8]Google LLC, Mountain View, CA, [910]DNAnexus, Santa Cruz, CA, [11]Chan Zuckerberg Initiative, San Jose, CA, [12]Univ. of Southern California, Los Angeles, CA, [13]Natl. Inst Aging, Bethesda, MD, [14]UCSC, Santa Cruz, CA, [15]Univ. of California Santa Cruz, Santa Cruz, CA, [16]Johns Hopkins Univ., Baltimore, MD, [17]NIH/NHGRI, Bethesda, MD, [18]Baylor Coll. Med., Houston, TX

Long-read sequencing technologies substantially overcome the limitations of short-reads but to date have not been considered as feasible replacement at scale due to a combination of being too expensive, not scalable enough, or too error-prone. In this work, we designed an ONT sequencing protocol that produces over 100Gb of ONT reads using a single PromethION flow cell. We developed new workflows and adapted and updated existing tools, combined into an end-to-end open-source pipeline implemented in WDL. Using the sequencing data from 3 human cell lines and 14 post-mortem brain tissue samples, we showed that our pipeline produces state-of-the-art SNP, structural variant and methylation calls. The protocol is currently being used to sequence thousands of brain genomes as a part of the NIH CARD initiative.

Our new Hapdup method generates de novo diploid assemblies from ONT sequencing only. Outside of centromeres and segmental duplications, our assemblies are structurally highly concordant with the current best human de novo assemblies produced from the more expensive combination of Pacbio HiFi and trio sequencing (structural variant F1-scores 0.95-0.97). The assembly representation allowed us to construct pangenomes of the MHC and IGH regions that are highly polymorphic and are otherwise difficult to analyze with linear references.

Our SNP calls produced with PEPPER-Margin-DeepVariant were similar or better than calls produced by the state-of-the-art Illumina-based methods (SNP F1-scores 0.997 and 0.996 respectively). The most noticeable improvement was associated with the regions of low short-read mappability. Small indels inside homopolymers and low-complexity repeats remain the major source of the residual errors (indel F1-score 0.87). However, our evaluation of the R10 sequencing protocol showed substantial improvements, in particular inside protein-coding regions (indel F1-score 0.997) and exons (indel F1-score 0.985).

We also added a new functionality to phase small and structural variant calls into megabase-scale haplotypes and reduce phasing switch error. Methylation calls produced from ONT data were highly concordant with the standard bisulfite sequencing, but in addition had haplotype-specific resolution, highlighted by our analysis of differentially methylated promoters in brain genomes.

Session Title: Omics Technologies Poster Session III

PB3468 scEmbed: Single-cell ATAC-seq clustering via representation learning of genomic regulatory regions.

Authors:

N. LeRoy, J. P. Smith, G. Zheng, E. Gharavi, J. Rymuza, A. Zhang, N. Sheffield; Univ. of Virginia, Charlottesville, VA

BACKGROUND: Thanks to the efforts of large consortia and thousands of individual labs, there is an explosion in the amount of scATAC-seq data available. However, critical challenges have emerged: handling the 1) inherent sparsity; and 2) high dimensionality of this data. Typically, this is addressed through dimensionality reduction techniques such as PCA, SnapATAC, or SCALE. These approaches cast cells into a low-dimensional representation, or embedding, for downstream tasks such as clustering or cell-type annotation. Existing methods typically use a one-step approach that embeds individual cells directly from a binary accessibility matrix. This requires significant compute power and resources. Here, we propose a two-step approach called scEmbed that first embeds genomic regions, and then uses these to build cell embeddings.
METHODS: For the first step, we use Word2Vec, an unsupervised machine learning framework, to learn low embeddings of genomic regulatory regions. For the second step, we use the learned region embeddings to compute embeddings of individual cells for downstream clustering tasks. We trained scEmbed on publicly available scATAC-seq data and then assessed it with two tasks: First, we compared scEmbed's cell clustering performance to existing analysis methods. Second, we tested the ability to apply scEmbed models pre-trained on reference data to unseen datasets.
RESULTS: First, we found that scEmbed performs as well as alternatives at cell clustering tasks. Our method consistently achieves high scores (>=0.4) for AMI, ARI, and homogeneity clustering metrics across three different clustering techniques (K-means, HC, and Louvain). Second, a major advantage of scEmbed over existing approaches is that pre-trained scEmbed models can embed unseen scATAC-seq data. Indeed, a pre-trained scEmbed model from a reference hematopoetic progenitor cell dataset could effectively cluster entirely independent hematopoetic and PBMC scATAC-seq data it had never seen. The clustering results from pre-trained models remain highly competitive against current methods (ARI > 0.4). Moreover, by using pre-trained models, we were able to generate the embeddings of >10,000 cells with >100,000 features in only a few minutes on standard desktop hardware.
SUMMARY: scEmbed will greatly improve the accessibility and throughput of scATAC-seq data analysis. scEmbed reduced the computation time of traditional scATAC-seq workflows from hours to minutes, while achieving comparable performance. This work sets the stage for large, general-purpose pre-trained models capable of generating embeddings of any scATAC-seq samples.

Session Title: Omics Technologies Poster Session I

PB3469 † ScISOr-ATAC reveals convergent and divergent splicing and chromatin specificities between cell types across cortical regions and disease states.

Authors:

C. Foord[1], W. Hu[1], J. Hsu[1], L. Fan[1], M. Corley[1], T. N. Bhatia[2], S. Xu[3], N. Belchikov[1], F. He[1], A. Pang[1], S. N. Lanjewar[2], J. J. Jarroux[1], A. Joglekar[1], T. A. Milner[1], L. Ndhlovu[1], J. Zhang[3], S. A. Sloan[2], L. Gan[1], H. U. Tilgner[1]; [1]Weill Cornell Med., New York, NY, [2]Emory Univ., Atlanta, GA, [3]Univ. of California Irvine, Irvine, CA

Multimodal measurements have become widespread in genomics, however measuring chromatin accessibility and splicing simultaneously in frozen brain tissues remains unconquered. To this effect, we have devised Single-Cell-ISOform-RNA sequencing coupled with the Assay-for-Transposase-Accessible-Chromatin (ScISOr-ATAC). First, we applied ScISOr-ATAC to Rhesus macaque (Macaca mulatta) prefrontal (PFC) and visual cortex to compare splicing and chromatin accessibility across brain regions and investigate how they may correlate. We found a higher number of excitatory neurons (ExN) compared to our human samples and identified 3 distinct subtypes characterized by layer-specific markers; RORB marks L3-L5; CUX2 marks L2/3 and L6; both CUX2 and RORB mark L2-L4. We profiled splicing and chromatin accessibility patterns in each ExN subtype between regions and showed that splicing is highly brain-region specific for RORB ExN, moderately specific in CUX2.RORB ExN and unspecific in CUX2 ExN. At the chromatin level, however, CUX2.RORB ExN show the highest brain-region specificity compared to other subtypes. These results indicate that some excitatory subtypes exhibit variable transcriptomic and chromatin accessibility patterns across tissues, while others do not. This suggests that splicing and chromatin can identify differences in excitatory subtypes within and across multiple regions, likely due to differences in morphology and signaling patterns. To compare chromatin accessibility and splicing patterns in disease, we also performed ScISOr-ATAC on Human Alzheimer's Disease (AD) and control PFC samples. We found a lower number of excitatory neurons in AD compared to controls, likely caused by disease-associated cell death. Due to this discrepancy in cell numbers, we developed a downsampling approach which samples from an equal number of reads across RNA and ATAC data. With this method, we found that oligodendrocytes are most affected by splicing and chromatin accessibility changes, suggesting that myelination may play an important role in AD. We additionally found an enrichment in spliced genes related to neurotransmitter receptors and synaptic transmission indicating that these changes play a role in neuronal signaling and function. Altogether, these results indicate that chromatin and splicing can show convergent or divergent dynamics depending on specific cases and justify the need for multimodal measurement to investigate complex systems and disease states. Here, ScISOr-ATAC allows for the characterization of single-cell patterns splicing and chromatin accessibility and the comparison of sample groups in frozen brain samples.

Session Title: Omics Technologies Poster Session II

PB3470 Screening Copy Number Variation with an Autoencoder

Authors:

K. Im, P. Ng, P. Shah, A. Kumar; MyOme, Inc, Menlo Park, CA

Background: Copy number variants are known causes of Mendelian diseases such as 22q11.2 microdeletion syndrome, spinal muscular atrophy, or hereditary breast and ovarian cancer syndrome and have also been shown to be associated with more complex phenotypes such as autism. Typically, CNVs are detected using chromosomal microarrays, array comparative genomic hybridization (aCGH), SNP arrays or, more recently, next-generation sequencing. We use whole genome sequencing to identify copy number variation (CNV) using a suite of algorithms such as Smoove, CNVnator, and Manta. The candidate CNVs are visually inspected to assess breakpoints and abnormal read depth, which is time-consuming, prone to error, and not scalable.
Methods: We implemented an autoencoder to efficiently remove technical artifacts and identify putative clinically relevant CNVs. An autoencoder is a deep learning model that learns to reconstruct input data, enabling anomaly detection by measuring the reconstruction error between the input and output. For each candidate CNV, we trained an autoencoder on reference samples without known disease-causing CNVs. We then tested whether the autoencoder's reconstruction error from a clinical sample was similar to reference samples (no CNV) or significantly different (CNV present). We evaluated the autoencoder on clinical samples with known pathogenic CNVs and rare (&lt0.01) CNVs from Genome in a Bottle's HG002.
Results: We observed improved sensitivity with the autoencoder compared to using z-scores based on average read depth. The autoencoder showed sensitivity of 100% (22/22) and 99.85% (667/668) on clinical samples and HG002, respectively. Using z-scores based on the average read depth in the candidate CNV region compared to a reference population, sensitivity was 100% (22/22) and 99.4% (664/668) on clinical samples and HG002, respectively. There was a small trade-off between sensitivity and specificity as the autoencoder had lower specificity. Specificity for the autoencoder and average read depth is 87.5% (385/440) and 88.2% (388/440), respectively.
Conclusions: The autoencoder is platform-specific and classifies polymorphisms and difficult-to-sequence regions as normal with an increase in sensitivity compared to a z-score based method. Implementing such a solution saves the need for manual sequence review of these regions while still preventing wasted time on clinical interpretation of sequence artifacts

Session Title: Omics Technologies Poster Session III

PB3471 Sequence-based deep learning for predicting gene expression from personal genomes

Authors:

A. Spiro, S. Mostafavi; Univ. of Washington, Seattle, WA

Sequence-to-function deep learning models trained on genomic DNA sequence can predict a variety of functional properties, including gene expression, with state-of-the-art accuracy. Currently, common models are trained on regions from the Reference Genome and are evaluated on their ability to predict population average features. Our recent work has shown that although these models are good predictors of population mean gene expression, they fall short when predicting gene expression from personal genomes. To address this gap in prediction performance, we propose a new approach to model training where we incorporate personal genomes into the training process. Our algorithm consists of two steps: training models on population average features (e.g., mean gene expression), followed by a fine-tuning step that utilizes personal-genomes paired with gene expression values. The effectiveness of the second step depends on the ability of the model to decompose gene expression per gene and per individual into its deviation from its mean value. We develop a composite loss function that enforces models to explicitly consider this decomposition. To assess our approach, we use paired Whole Genome Sequencing and gene expression data from n=190 individuals in the ROSMAP cohort where data is available for 18,776 protein-coding genes. We create personal genomic sequences for each individual and each gene, thus multiplying the size of our data by the number of individuals in the cohort (x190 fold for this experiment). We evaluate our model on individuals and genes that were not seen by the model during training. We assess model performance based on three correlation metrics between observed and predicted expression value: correlation across all data points, mean correlation per individual, and mean correlation per gene. As a baseline, we consider a model that has been trained only on the Reference Genome and population mean gene expression. We observe an increase in correlation across all data points from 0.530 to 0.606 and an increase in mean per individual correlation from 0.531 +- 0.011 to 0.608 +- 0.007 (+- indicates standard deviation). We do not observe an increase in mean per gene correlation. From these results, we conclude that training on personal genomes has the potential to increase the accuracy of sequence-based prediction of gene expression, however the differences between genes rather than between individuals dominate model learning when the prediction tasks are combined. We are currently exploring other fine-tuning strategies to better focus models on learning per gene expression differences across individuals.

Session Title: Omics Technologies Poster Session I

PB3472 SEQUIN: interactive web app for rapid, reproducible bulk and single cell RNA-Seq analysis

Authors:

C. Weber[1,2], M. B. Hirst[3], B. Ernest[3], N. J. Schaub[2], K. M. Wilson[2], K. Wang[2], H. M. Baskir[2], P-H. Chu[2], C. Tristan[2], I. Singec[2]; [1]Natl. Eye Inst., Rockville, MD, [2]Natl. Ctr. for Advancing Translational Sci., Rockville, MD, [3]Rancho BioSci.s, San Diego, CA

SEQUIN is a web-based R/Shiny application that allows fast and intuitive analysis of RNA-sequencing data derived for model organisms, tissues, and single cells. Integrated app functions enable uploading datasets, quality control, differential gene expression analysis (DGE), gene set enrichment (GSE), and data visualization. We developed the iPSC Profiler, a practical gene module scoring tool that helps to measure and compare pluripotent and differentiated cell types. Freely available to the public, SEQUIN empowers scientists using interdisciplinary methods to investigate and present transcriptome data firsthand with state-of-the-art statistical methods. SEQUIN help democratize and increase the throughput of analysis using next-generation sequencing data with single-cell resolution. Here, we analyzed a new pair of unpublished datasets from human iPSCs of cell lines cultured with differing media, in order to determine optimal conditions for growth and self-renewal: E8, E8+Albumin, and mTeSR. Cell lines were: LiPSC GR1-1, GM25256, NCRM5, and WA09. Bulk RNA-Seq had 27 samples in triple technical replicates, while the scRNA had a total of 24,186 cells. We performed pairwise DGE and PCA-based clustering, followed by GSE. In both datasets we found samples were most tightly clustered by growing media first, then by cell line. In PCA plots, E8 and mTeSR media were more similar by overall expression than E8 and Albumin. After performing DGE with media as contrast (E8 vs. E8-Albumin) and evaluating iPSC Profiler gene module scores, pluripotency was equally high across media types and cell lines. By the intersection of significant DE genes in bulk and scRNA-Seq, for the same media comparison, there were 90 genes (adj. p-value < 0.001, abs. value log2 fold change >= 0.5). We performed GSE with this gene set in SEQUIN and found enriched Gene Ontology terms relating to prostaglandin biosynthesis and cell membrane raft organization. These terms included ANXA (annexin) gene family which relates to phospholipid binding. Since the media differ by lipid concentration, this could explain the upregulation of ANXA. These experiments exemplify the use of SEQUIN for rapid and reproducible analysis leading to actionable insights.

Session Title: Omics Technologies Poster Session II

PB3473 Serum sphingolipid levels are modified by *HLA-DRB1*15:01* and *HLA-A*02:01* in multiple sclerosis.

Authors:

E. Misicka[1], W. Tobin[2], M. Fereidan-Esfahani[2], J. Sagen[2], S. Liu[1], J. Oksenberg[3], S. Gregory[4], D. Crawford[1], F. Briggs[1]; [1]Case Western Reserve Univ., Cleveland, OH, [2]Mayo Clinic, Rochester, MN, [3]Univ California San Francisco, San Francisco, CA, [4]Duke Molecular Physiology Inst., Durham, NC

Multiple sclerosis (MS) is a demyelinating autoimmune disease, and *HLA-DRB1*15:01* and *-A*02:01* are the most prominent risk and protective alleles that encode MHC class II and class I molecules, respectively, involved in antigen presentation. Examining metabolites in the context of genetic predisposition may highlight dysregulated processes relevant for the pathology of MS. Here, we sought to identify serum metabolites whose levels varied by *HLA-DRB1*15:01* and *-A*02:01* allele status in relationship with MS (gene-metabolite interactions). Metabolon's untargeted metabolomic profiling was completed for serum from 105 healthy controls (HCs) and 73 relapsing remitting (RR) MS cases who were DMT naïve/free, ≤5yrs from onset and ≤2yrs from diagnosis. All subjects were non-Latinx White. After QC, there were 1119 metabolites. *HLA-DRB1*15:01* and *-A*02:01* carriers were defined by tagging SNPs. Logistic regression models included an interaction term between each metabolite and allele carrier status in separate models, adjusting for age, sex, BMI, smoking status, and population substructure. Stratified models then characterized the direction of effects. 53% and 44% of MS cases and 35% and 37% of HCs were *HLA-DRB1*15:01* and *-A*02:01* carriers, respectively. In *HLA-DRB1*15:01* models, 75 metabolites had significant interactions, revealing an enrichment of sphingomyelins (fold change [FC]=9.3, $p=2x10^{-16}$) and ceramides (FC=5.6, p=0.01). In carriers, there were several notable associations with MS status, e.g., 3-hydroxybutyroylglycine (β=2.6, p=0.003), N2-acetyl,N6-methyllysine (β= -1.4, p=0.003), and kynurenine (β= -2.4, p=0.02). In non-carriers, sphingomyelins and ceramides were exclusively elevated in MS vs HCs, e.g. sphingomyelin (d18:1/17:0) (β=3.1, p=0.001). In *-A*02:01* models, there were 27 significant interactions, with an enrichment of tryptophan metabolites (FC=6.9, p=0.002), sphingomyelins (FC=4.3, p=0.03), and nucleotides (FC=3.6, p=0.05). Amongst carriers, several metabolites were associated with MS, e.g. tryptophan betaine (β= -1.1, p=0.008) and methyl indole-3-acetate (β=1.6, p=0.01). In non-carriers, there were also many associations, e.g. dopamine 3-O-sulfate (β= -1.9, p=0.003) and cholate (β= -1.2, p=0.008) both involved in processes previously associated with MS. In conclusion, there are differential serum levels of multiple metabolites in MS cases vs HCs, whose levels appear to be modified by the lead risk and protective alleles. Further investigation of the influence of these variants on the serum levels of serum lipids and amino acids may reveal novel biological mechanisms for the development of MS.

Session Title: Omics Technologies Poster Session III

PB3474 Single cell dissection of smoking and drinking addiction associated loci in multi-ancestry genetic studies.

Authors:

S. Gao, L. Wang, X. Wang, GWAS & Sequencing Consortium of Alcohol and Nicotine use; Penn State Coll. of Med., Hershey, PA

Genome-wide association studies (GWAS) are successful in detecting associations between genetic variants and tobacco and alcohol use. However, most of the associated variants lie within the non-coding regions with unknown functions for tobacco and alcohol use. To explore non-coding genetic signals, a common approach is to associate them with gene expression levels through expression quantitative trait loci. The majority of eQTL studies utilize bulk RNA-seq samples, which average the effect across cell types and hence remain underpowered to dissect cell type-specific effects. Thanks to improved single cell sequencing and multiplexing technology, scRNASeq datasets with 100s individuals start to emerge, which allow us to identify cell type specific eQTLs (sc-eQTLs) in the brain. To further identify risk genes in brain cell types, we propose a new method to integrate bulk and single-cell RNASeq data, improving cell type-specific eQTL detection and integrating it with multi-ancestry genetic studies. The method capitalizes the insights that bk-eQTLs are weighted averages of sc-eQTLs. By jointly analyzing bk- and sc-eQTLs, we can leverage large sample sizes of bk-eQTLs and enhance the cell type specific effect estimation. In this method, we jointly analyze the bk-eQTL from Cortex region in the MetaBrain data with the single cell RNASeq data from Bryois et al. (n = 192) to improve the sc-eQTL detection. To validate the improved cell-type-specific effect estimates of brain eQTLs, we seek replicate the microglia-specific signals in an independent microglia eQTL dataset. We replicate 1609 deconvoluted cell-type-specific eQTLs ($p < 1e-6$), which is 14% more than cell-type-specific eQTLs identified by single cell data alone. We then conduct co-localization analysis with *coloc* to exaimne whether European-specific local ancestry genetic variants related to smoking and drinking traits potentially act through brain cell type cis-eQTLs. Among 620 Smoking initiation associated loci, the deconvoluted cell-type-specific eQTL results identified 57 co-localized loci in Astrocytes, 130 in Excitatory neurons, 34 in Microglia, 30 in OPCs...COPs, 21 in Endothelial cells, 50 in Inhibitory neurons, 86 in Oligodendrocytes, and 17 in Pericytes using a H4-posterior probability cutoff of 0.9. Overall, the co-localization analysis identifies genes in brain cell types at 28-39% of the GWAS loci with deconvoluted results and only 3-10% using single cell data alone. Our ongoing research in deconvoluting brain cell-type-specific eQTL effect will narrow down the risk genes and enhance our understanding of the genetic basis underlying addiction traits in distinct brain cell types.

Session Title: Omics Technologies Poster Session I

PB3475 Single cell long read mRNA isoform regulation is pervasive across mammalian brain regions, cell types, and development.

Authors:

H. Tilgner[1], A. Joglekar[1], W. Hu[1], B. Zhang[2], O. Narykov[3], M. Diekhans[4], J. Balacco[5], L. C. Ndhlovu[1], T. A. Milner[1], O. Fedrigo[5], E. Jarvis, D[5], G. Sheynkman[6], D. Korkin[7], M. Ross[1]; [1]Weill Cornell Med., New York, NY, [2]spatialgenomics, inc, Pasadena, CA, [3]Worcester Polytechnic Inst., New York, NY, [4]UCSC, Santa Cruz, CA, [5]Rockefeller Univ., New York, NY, [6]Univ. of Virginia, Charlottesville, VA, [7]Worcester Polytechnic Inst., Worcester, MA

RNA isoforms influence cell identity and function. Until recently, technological limitations prevented a genomewide appraisal of isoform influence on cell identity in various parts of the brain. Our prior work developed single cell isoform sequencing for fresh and frozen tissues, spatial isoform sequencing and accurate long read interpretations[1-4]. Here, we present results under review[5] mapping isoforms accross multiple mouse brain regions, cell subtypes, and developmental timepoints from postnatal day 14 (P14) to adult (P56). For 75 percent of genes, full length isoform expression varies along one or more axis (brain region, cell subtype or age), underscoring the pervasiveness of isoform regulation across multiple scales. As expected, splicing varies strongly between cell types. However, certain gene classes including neurotransmitter release and reuptake as well as synapse turnover, harbor significant variability in the same cell type across anatomical regions, suggesting differences in network activity may influence cell type identity. Glial brain region specificity in isoform expression includes strong poly(A) site regulation, whereas neurons have stronger TSS regulation. Furthermore, developmental patterns of cell type specific splicing are especially pronounced in the P21 to P28 transition. The same cell type traced across development shows more isoform variability than across adult anatomical regions, indicating a coordinated modulation of functional programs dictating neural development. As most cell type specific exons in P56 mouse hippocampus behave similarly in newly generated data from human hippocampi, these principles may be extrapolated to human brain. However, human brains have evolved additional cell type specificity in splicing, suggesting gain of function isoforms. Taken together, we present a detailed single cell atlas of full length brain isoform regulation across development and anatomical regions, providing a previously unappreciated degree of isoform variability across multiple scales of the brain. [1]. Gupta*,Collier* et al, Nature Biotechnology, 2018 [2] Hardwick*,Hu*,Joglekar* et al, Nature Biotechnology, 2022 [3] Joglekar et al, Nature Communications, 2021 [4] Prjibelski*,Mikheenko* et al, Nature Biotechnology, 2023 [5] Joglekar et al, biorxiv, 2023

Session Title: Omics Technologies Poster Session II

PB3476 Single cell RNA sequencing reveals heterogeneity across the intestines in treatment naïve individuals diagnosed with Crohn's Disease.

Authors:

S. Washburn[1], V. Kolachala[2], A. Dodd[2], S. Maddipatla[2], S. Murthy[2], Y. Hwang[2], J. Matthews[2], G. Gibson[1], P. Qiu[1], S. Kugathasan[2]; [1]Georgia Tech, Atlanta, GA, [2]Emory Univ Sch. of Med., Atlanta, GA

Crohn's Disease (CD) is distinguished by chronic, relapsing transmural inflammation throughout both the small and large intestine. Multiple components can influence risk for developing CD such as environmental factors, genetic susceptibility, and microbial interactions within the intestines. Currently, the exact cause of CD remains unknown, and the specific cellular contribution to disease pathogenesis and therapeutic response remains to be elucidated. The aim of this study is to investigate the heterogeneity of underlying disease mechanisms and identify personalized disease signatures that are cell type or tissue specific. Treatment naïve subjects (n = 20) diagnosed with CD were recruited for this study, and droplet based single-cell RNA sequencing on 10X Genomics Chromium platform was performed on ileal (n = 19), colon (n = 21), and rectal (n = 17) mucosal tissue from each subject. After quality control, 142,245 cells were included in downstream analysis, analyzed using the R package, Seurat. Single cell transcriptomic profiles across tissue compartments were contrasted to validate the hypothesis that there are underlying cellular function differences across tissue. Within each tissue, immune, epithelial, and stromal cells were identified. The cellular proportion profiles for individuals were heterogeneous both within and across tissue compartments. Differentially expressed gene analysis was performed using MAST to investigate genetic differences of cell types across tissue. Pathway analysis, using ToppGene, revealed that oxidative phosphorylation pathways, which might affect cellular redox imbalance, and immune cell activation pathways are upregulated in the colon goblet cells relative to ileum goblet cells, supporting our hypothesis. Enrichment analysis was performed by overlaying Inflammatory Bowel Disease (IBD) Genome Wide Association Study (GWAS) genes with differentially expressed genes to inform upon potential pathogenic cell types within the mucosa. IBD GWAS genes were enriched in both immune and epithelial cells in the colon, highlighting the importance of investigating cellular interactions within tissue. Overall, our findings of marked differences in cellularity during disease presentation explains in part the differential treatment responses/non-responses observed in initial therapy in these individuals. This emphasizes the need for personalized medicine, perhaps based on initial mucosal single cell profiles in CD. Further, enrichment analysis can be leveraged to identify potentially pathogenic cell types for disease across tissue and inform upon therapeutic targets.

Session Title: Omics Technologies Poster Session III

PB3477 Single-cell analysis of the human infant airway epithelium reveals heterogenous cell subpopulations during RSV infection

Authors:

B. Bera[1], S. Bhattacharya[1], E. Chorvinsky[1], K. Salka[1], A. Welham[1], M. Rujana[1], D. Thompson[1], H. Morizono[1], M. Gutierrez[2], J. Jaiswal[1], G. Nino[1]; [1]Childrens Natl. Hosp., Washington, DC, [2]Johns Hopkins Univ. Sch. of Med., Baltimore, MD

Rationale: Respiratory Syncytial Virus (RSV) is the most common cause of severe viral respiratory infection in infants. The infant airway epithelium is a pseudostratified cell barrier that serves as the RSV point of entry and site of initial host defense. Dissecting the heterogeneity of virus-host interactions at the single cell level in the infant airway epithelium is essential to identify novel molecular pathways involved in viral infection and ultimately pave the way for targeted therapy. The goal of this study was to use to use Single Cell RNA seq (scRNA seq) to characterize the transcriptomic program of human infant airway epithelium during RSV infection. **Methods:** We performed scRNA seq of approximately 8037 nasal AECs from a single human infant donor cultured in the presence and absence of an eGFP tagged RSV-A2 virus (MOI 1). CellRanger was used to convert raw reads to gene expression, Seurat v3 and DoubletFinder were used for quality check, UMI counts were normalized using scTransform, Seurat package was used for downstream analysis. Differential expression (DE) was done by wilcox test (log FC > |0.25| & adj. pval < 0.05) in Seurat. Gene set enrichment analysis (GSEA) was used for functional annotation. **Result:** Based on Single cell eGFP expression, we identified subpopulations of AECs with high levels RSV infection above 75th percentile (RSV High) and those having expression below 75th percentile RSV Low). We found 491 differentially expressed (DE) genes between RSV high and RSV Low out of which 183 are upregulated genes in RSV High and 308 upregulated genes in RSV Low. Our data reveals subpopulations that have susceptibility and resistance genes during RSV infection which highlights functional roles that AECs have during infection. Integrated functional analyses identified that the Low RSV group (resilient AECs) demonstrate upregulation of interferon-inducible antiviral effectors (e.g., MX1) and High RSV group have RSV-induced upregulation of metalloproteinases (e.g., MMP10) and pro-inflammatory chemokines (e.g., CXCL8). **Conclusion:** Using our single-cell analysis of the human infant airway epithelium we have identified two subpopulations with susceptibility and resistance genes during RSV infection. By providing mechanistic insight surrounding the overall modulation of the inflammatory response, these results will be important in offering new target concepts for helping the pediatric population during RSV infection.

Session Title: Omics Technologies Poster Session I

PB3478 Single-cell long-read sequencing in human cerebral organoids uncovers autism-associated exons

Authors:

Y. Yang, R. Yang, B. Kang, S. Qian, X. He, X. Zhang; The Univ. of Chicago, Chicago, IL

Alternative splicing significantly increases transcript diversity and dysregulation of alternative splicing has been repeatedly associated with neurodevelopmental disorders. Single-cell analysis revealed unprecedented cellular diversity in the central nervous system but the extent of cell-type-specific splicing remains largely uncharted. Here we performed single-cell long-read sequencing in the human cerebral organoids and identified over 32,000 uncatalogued isoforms and 3,622 cell-type-specific splicing events. Long reads uncovered coordinated splicing and cell-type-specific intron retention events which were challenging to study with short reads. Retained introns were enriched in RNA splicing and translation and showed weaker 5′ splice sites. We used this rich dataset to explore the biological processes underlying neurodevelopmental disorders, focusing on autism. We found that the exons enriched in autism cases versus controls tend to have higher inclusions in progenitor cells than neurons, suggesting that the splicing program in autism is closer to the progenitor state than differentiated neurons. Furthermore, cell-type specific exons harbor significantly more *de novo* mutations in probands than siblings, highlighting the importance of these exons to autism genetics. We identified 118 autism genes displaying cell-type-specific isoforms, including an alternative exon in *UBE3A* that introduced an alternative translation start site in excitatory neurons. Overall, our results indicate pervasive splicing changes across neural cell types and highlight the importance of cell-type-specific splicing in autism.

Session Title: Omics Technologies Poster Session II

PB3479 Single-cell multi-omics analysis of mucosal immune cell profiles in ulcerative colitis

Authors:

S. Feng[1,2], E. Heidrich-O'Hare[1], Z. Xu[1,2], H. Hu[1], W. Chen[1], R. Duerr[1]; [1]Univ. of Pittsburgh, Pittsburgh, PA, [2]Tsinghua Univ., Beijing, China

INTRODUCTION: Ulcerative colitis (UC) and Crohn's disease (CD) are the two forms of inflammatory bowel disease (IBD). When UC involves only part of the colon, a sharp demarcation usually separates inflamed distal mucosa and uninflamed proximal mucosa. GWAS have successfully identified genetic risk loci associated with UC; however, the functional interpretation of these loci and their implications in disease pathogenesis remain incompletely understood. **RATIONALE:** Cells and molecules enriched in inflamed mucosa are candidate targets for UC therapy. Therefore, we compared single-cell assays for gene expression (GEX) and antibody-derived cell surface protein tags (ADTs) (CITE-seq), or GEX, ADTs and transposase accessible chromatin (ATAC) (DOGMA-seq) in immune cells from macroscopically inflamed and non-inflamed mucosa in UC patients. **METHODS:** Biopsies of inflamed and non-inflamed mucosa in UC patients were cryopreserved, thawed in batches, dissociated, and flow-sorted to enrich live immune cells. ~45,000 cells from 13 UC patients were profiled with CITE-seq, and ~4,000 cells from 6 UC patients were profiled with DOGMA-seq. Downstream analyses include trajectory inference by RNA velocity, identification of differentially expressed genes (DEG) or activated peaks (DA peaks), analysis of chromatin activity at GWAS loci, and genome-wide peak-to-gene linkages. **RESULTS:** We delineated 19 CD4+ T cell subtypes with CITE-seq. DOGMA-seq identified similar major clusters and precisely defined helper T cell subtypes based on transcription factor (TF) activities inferred from ATAC peaks. We also identified the POU family as marker TFs for the recently defined CXCR5- Tfh-like cells. Analysis of cell proportions and trajectory corroborated that CD127+ CD4+ T cell development dominates in non-inflamed mucosa, while that of CD279+ CD4+ T cells and Treg dominates in inflamed mucosa. Enrichment of UC GWAS lead SNPs in ATAC peaks was observed in CD127+ CD4+ T cells in non-inflamed mucosa, CD279+ CD4+ T cells in inflamed mucosa, and CD8+ T cells in non-inflamed mucosa. DA peaks between inflamed and non-inflamed mucosal immune cells that overlap UC or IBD (UC+CD) GWAS SNPs were linked to DEGs, suggesting regulatory roles of these UC/IBD GWAS SNPs. To further elucidate regulatory mechanisms in CD4+ T cells, we constructed a TF-gene regulatory network and identified immune cell ligand-receptor interactions that impact downstream signaling targets. **CONCLUSION:** We provide valuable insights into the phenotypic and regulatory landscape and intercellular communication networks of mucosal immune cells in UC, with a particular focus on the role of CD4+ T cells.

Session Title: Omics Technologies Poster Session III

PB3480 Single-cell spatial multiomics uncovers new mechanisms of myelin deforming diseases

Authors:

Y. Wang, A. Wang; Univ. of California Davis, Sacramento, CA

Demyelinating diseases, including conditions like multiple sclerosis, acute disseminated encephalomyelitis, and Balo's disease, afflict over a million people. Yet, the exact processes that cause myelin damage in these diseases remain largely elusive. In this study, we examine the brain sections of a demyelinating mouse model using cutting-edge single-cell spatial transcriptomics technologies like Stereo-seq, CosMx, Xenium, Visium, GeoMx, along with advanced epigenetic and metabolomics tools such as sc-spatial ATAC and imaging mass spectrometry and unveil the roles of several signaling pathways in the disease development like such as cell cycle arrest at G2 phase, dysregulated lipid metabolism and vesicle transport. Integration of the data from different sc-spatial omics offers a panoramic view to deepen our understanding of these diseases.

Session Title: Omics Technologies Poster Session I

PB3481 Single-nucleus RNA-seq reveals key contributors in Duchenne muscular dystrophy

Authors:

E. Jeon[1], J-H. Chae[2], M. Choi[1]; [1]Seoul Natl. Univ., Seoul, Korea, Republic of, [2]Seoul Natl. Univ. Hosp., Seoul, Korea, Republic of

Duchenne muscular dystrophy (DMD) is a devastating X-linked disorder caused by mutations in the dystrophin gene. Slowly deteriorating muscle forces patients to lose ambulation in their early teens and to die in their 20s, mostly due to respiratory or cardiac complications. Deflazacort, a glucocorticoid derivative, has been used to control inflammation and delay muscle weakness. Despite recent progress in understanding of the genetics and pathogenesis of the disease and attempts to treat it, DMD still remains a major pediatric muscle dystrophy without cure due to an incomplete characterization of the molecular and cellular interactions responsible for muscle maintenance and functions. Here, we present single-nucleus RNA-sequencing (snRNA-seq) results of muscles from DMD, milder-form Becker muscular dystrophy (BMD), and healthy controls, followed by spatial transcriptomics analysis of DMD and healthy muscles. In addition, to examine the effect of deflazacort, we administered the drug to *D2-mdx* mice and subjected their muscles to snRNA-seq. Integrated analysis of human and mouse muscles revealed pathogenic features in the patients and palliative effects of deflazacort. Meanwhile, it also highlighted the perturbation of proliferating satellite cells, leading to an increased signal transduction pathway involving EZH2, NR3C1, and cell cycle progressor proteins in the patient cells, which was confirmed by EZH2 ChIP-seq in satellite cells. We also demonstrate a therapeutic effect by perturbation of these pathways in *D2-mdx* mice through improved muscle phenotype. Our analysis of DMD patient muscles reveals pathogenic mechanisms that can be readily targeted by pre-existing therapeutic options.

Session Title: Omics Technologies Poster Session III

PB3483 Spatially-resolved, single-cell transcriptomics detects cell types and novel cell states in kidney diseases

Authors:

J. Reeves[1], E. Metzger[1], R. Ivy[2], H. Ira[2], N. Brousaides[2], S. Kim[1], S. Murphy[1], R. Smith[2], R. Colvin[3]; [1]NanoString Technologies, Seattle, WA, [2]Massachusetts Gen. Hosp., Boston, MA, [3]Massachusetts Gen. Hosp., boston, MA

Spatially-resolved, single-cell tissue atlases can provide mechanistic insights not appreciable by bulk RNAseq or spatially-agnostic scRNAseq. We have begun to harness the power of space in understanding mechanisms in kidney diseases, beginning with pathogenetically relevant transplant glomerulopathy in chronic antibody mediated rejection (CAMR), the major cause of late kidney graft loss.

A necessary first step is to generate the spatial atlases themselves. To do this, gene expression was measured in FFPE biopsy sections from 7 CAMR cases, 4 TCMR cases, and 3 cases with no evidence of rejection (NER) with 1000 gene probes using the CosMxTM Spatial Molecular Imager. Cells were detected by IF staining of B2M/CD298 and DAPI with the machinery in the AtoMxTM Spatial Informatics Platform. The analyzed dataset consisted of 295,937 cells and a total of 51,142,869 transcripts. Cell typing was accomplished by semi-supervised clustering with the Kidney Cell Atlas database.

With the disease-inclusive spatial atlas, we report on 3 tests related to kidney rejection and cell states. First, we examined expressional differences of glomerular endothelial cells (GECs) within glomeruli between CAMR and controls. In the CAMR, GECs showed upregulation of genes related to repair and angiogenesis (e.g., COTL1: log2FC=1.1, FDR=2.9e-5); protective genes (e.g., CD59: log2FC=0.96, FDR=9.2e-7), and a downregulation of genes involved in VEGFR2 signaling including (e.g.,EFNB2: log2FC=1.25, FDR=1.8e-31). Second, we further interrogated the in-glomerulus CAMR GECs. We found two distinct classes of GECs with one class containing genes related to inflammation and angiogenesis and the other class previously identified in normal kidneys with genes related to cell adhesion and growth arrest. Third, we sought to understand the role of NK cells in CAMR. We found that NK cells tended to co-localize with GECs in CAMR (p=0.08). Moreover, differential expression showed evidence that CAMR-residing NK cells were activated relative to NER NK cells.

We found profound transcript changes in GEC affecting the VEGFA pathway, known to be required for normal GEC differentiation. NK cells were in proximity to glomerular endothelium and expressed increased activation transcripts in CAMR. Our data show that pathological phenotypes of cells may not be represented in normal cell atlases and these atlases can be improved with spatially-resolved, single cell approaches.

Session Title: Omics Technologies Poster Session I

PB3484 Spot-level cell type deconvolution methods using Visium immunofluorescence data on the human anterior hippocampus.

Authors:

M. Tippani[1], E. Nelson[1,2], A. Ramnauth[1,3], H. Divecha[1], S. Kwon[1,3,4], S. Bach[1], J. Yao[5], E. Pattie[1], T. Hyde[1,6,7], L. Collado Torres[1], K. Maynard[1,6], K. Martinowich[1,3,6,8], S. Page[1], S. Hicks[5]; [1]Lieber Inst. for Brain Dev., Baltimore, MD, [2]Cellular and Molecular Med., Johns Hopkins Sch. of Med.,, Baltimore, MD, [3]The Solomon H. Snyder Dept. of NeuroSci., Johns Hopkins Sch. of Med., Baltimore, MD, [4]Biochemistry, Cellular, and Molecular Biology Graduate Program, Johns Hopkins Sch. of Med., Baltimore, MD, [5]Dept. of Biostatistics, Johns Hopkins Bloomberg Sch. of Publ. Hlth., Baltimore, MD, [6]Dept. of Psychiatry and Behavioral Sci., Johns Hopkins Sch. of Med., Baltimore, MD, [7]Dept. of Neurology, Johns Hopkins Sch. of Med., Baltimore, MD, [8]The Kavli NeuroSci. Discovery Inst., Johns Hopkins Univ., Baltimore, MD

Background: The hippocampus (HPC) is a 3-layered archicortex that plays key roles in plasticity, learning, and memory. It is composed of heterogeneous cell types localized across the major subregions like dentate gyrus (DG), CA1-4, and subiculum. Many important HPC functions are reflected in its spatial organization, and hence molecular profiling approaches that retain spatial information of gene expression are highly useful. The 10x Genomics Visium platform combines transcriptome-wide gene expression with high-resolution imaging, but not at the single-cell level. Methods like SPOTlight, Tangram, and cell2location have been developed to perform cell-type deconvolution of individual Visium spots, and the Visium Spatial Proteogenomics (SPG) platform provides orthogonal data by allowing for protein detection on immunofluorescence images. Methods: We used Visium-SPG to label 4 broad cell types in tissue sections at the level of the anterior hippocampus from 2 neurotypical donors. Tissue was immunostained for DAPI, GFAP, NeuN, OLIG2, and TMEM119 to identify nuclei, astrocytes, neurons, oligodendrocytes, and microglia, respectively. Multispectral imaging was performed followed by on-slide cDNA synthesis and Illumina sequencing. Data from Visium-SPG was paired with existing HPC Visium and snRNA-seq (150,917 spots and 81,123 nuclei from 10 donors) from spatially-adjacent tissue sections. Individual nuclei in the DAPI channel were segmented using Cellpose, and ~1,000 cells were manually classified. A decision-tree classifier was then trained to classify the remaining cells from all samples. Results: We investigated the performance of Tangram, cell2location, and SPOTlight, for cell type deconvolution. These tools mapped individual cell type annotation onto individual spots measured in the Visium-SPG expression data and all the results were compared to the orthogonal SPG-derived cell proportions. The best-performing method was used to perform spot deconvolution on the remaining Visium samples. Conclusion: Given the sparsity of gene expression data in the Visium platform, spot deconvolution is challenging in complex regions like HPC. While current methods can generate results that are satisfactory for downstream analyses, improved methods are needed for specialized neuroanatomical structures with unique cellular compositions. The data generated here provides a valuable resource for benchmarking future spot deconvolution methods by providing an archival dataset for the cellular composition in the anterior HPC.

Session Title: Omics Technologies Poster Session II

PB3485 Statistical inference and downstream analysis of cell-type-specific co-expressions from single cell RNA-sequencing data

Authors:

C. Su[1], X. Shan[2], Z. Xu[3], B. Cai[2], H. Zhao[2], J. Zhang[1]; [1]Emory Univ., Atlanta, GA, [2]Yale Univ. Sch. of Publ. Hlth., New Haven, CT, [3]Univ. of Washington, Seattle, WA

The advancement of single cell RNA-sequencing (scRNA-seq) technology has enabled the direct inference of co-expressions in specific cell types, facilitating our understanding of cell-type-specific biological functions. For this task, the high sequencing depth variations and measurement errors in scRNA-seq data present two significant challenges, and they have not been adequately addressed by existing methods. We propose a statistical approach, CS-CORE, for estimating and testing cell-type-specific co-expressions, that explicitly models sequencing depth variations and measurement errors in scRNA-seq data. Systematic evaluations show that most existing methods suffered from inflated false positives as well as biased co-expression estimates and clustering analysis, whereas CS-CORE gave accurate estimates in these experiments. When applied to scRNA-seq data from postmortem brain samples from Alzheimer's disease patients/controls and blood samples from COVID-19 patients/controls, CS-CORE identified cell-type-specific co-expressions and differential co-expressions that were more reproducible and/or more enriched for relevant biological pathways than those inferred from existing methods. Based on CS-CORE, we further developed two pipelines for downstream analyses that utilize cell-type-specific co-expressions for biological discoveries: co-expression QTL analysis and module-trait association analysis. Using population-scale scRNA-seq data, the co-expression QTL analysis identifies genetic variations which correlate with co-expressions in specific cell types, revealing dysregulation in functional pathways that may be caused by genetic variations. Based on scRNA-seq data with rich phenotypes, the module-trait association analysis estimates the correlation between groups of genes and phenotypes, uncovering the involvement of biological pathways and molecular mechanisms in disease pathogenesis in cell types of interests.

Session Title: Omics Technologies Poster Session III

PB3486 † STDCC: Supervised Tensor Decomposition tool for studying Cell-cell Communication using single-cell RNA-seq data

Authors:

Q. Dai[1], M. Epstein[2], J. Yang[1]; [1]Emory Univ., Atlanta, GA, [2]EMORY Univ., Atlanta, GA

Motivation: Knowledge of communication patterns among different types of cells is crucial for studying biological processes and diseases. Multiple computational tools have been developed to infer such cell communication patterns using multi-subject single-cell RNA-seq (scRNA-seq) data. However, existing tools in this space are unsupervised and infer communication patterns without considering confounding factors and phenotypes of interest.
Methods: To fill this significant gap, we propose a Supervised Tensor Decomposition tool for studying Cell-cell Communication (STDCC) using scRNA-seq data, which simultaneously infers cell communication patterns, estimates phenotypic effects, and adjusts for confounding covariates. STDCC first constructs a 4-dimensional communication score tensor representing subjects, sender/receiver cell types, and ligand-receptor pairs per dimension, then employs a supervised tensor decomposition technique to obtain low-dimensional factors representing distinct communication patterns. Specifically, STDCC decomposes the communication score tensor within the defined feature space of confounding covariates and the phenotype of interest, and thus can identify phenotype-associated cell-cell communication patterns while adjusting for confounding covariates.
Results: Simulation studies based on communication patterns obtained from real data demonstrated that STDCC outperformed unsupervised methods with improved decomposition accuracy and reduced bias of phenotypic effects. Applying STDCC to scRNA-seq data profiled for 13 Autism Spectrum Disorder (ASD) subjects and 10 controls, we found that dysregulations of interactions between neurexin-neuroligin, neural cell adhesion molecules, and extracellular matrix-receptor in AST-FB, AST-PP, endothelial, L2/3, L4, L5/6, and L5/6-CC cells were most strongly associated with ASD. We also found that non-biological confounding covariates like batch effects were strongly associated with the decomposed communication patterns in ASD data, emphasizing the importance of accounting for confounding covariates to obtain robust and interpretable cell-cell communication patterns. We also applied STDCC to real scRNA-seq data profiled for studying COVID-19 severity (n = 60) and SLE (n = 221) and observed interesting cell-cell communication patterns as well.
Conclusion: We provide a useful tool STDCC for studying cell-cell communication using scRNA-seq data, which can account for confounding factors and accurately estimate phenotypic effects. A Python package for implementing the proposed STDCC tool will be publicly available through GitHub.

Session Title: Omics Technologies Poster Session I

PB3487 Streamlined RNA-Seq workflow reveals extracellular vesicle transcriptome insights.

Authors:

T. Kainth, M. Isa, S. Perike, E. Dreskin, S. Okino, C. Jiang, L. Lingelbach, N. Kumar; Bio-Rad Lab., Hercules, CA

Extracellular vesicles (EVs), produced naturally within cells, are membrane-bound particles that contain DNA, RNA, protein, and other bioactive molecules. EVs released from parent cells deliver cargo to recipient cells, triggering vital biological responses such as cell proliferation, differentiation, and apoptosis. EVs thus play a critical role in cell-to-cell communication, an essential process that helps maintain the health of complex organisms. Current EV RNA research focuses primarily on miRNAs which, as a class, have regulatory effects on biological pathways and can be involved in cell-to-cell communication. Many of these studies used miRNA-specific reagents for RNA isolation and RNA-Seq library preparation; this approach likely limited the findings of non-miRNA transcripts. Here we present a novel workflow that overcomes these limitations, enabling comprehensive assessment of the entire EV transcriptome. This streamlined approach allows for the isolation of EV RNA from only 1.4 mL of conditioned cell culture media using a standard microcentrifuge and two separate, one-hour sessions over two days. Our 10-minute RNA extraction method efficiently removes protein and DNA from the lysate while preserving RNAs of all sizes. We employ a polyadenylation/template switch strategy for RNA-Seq library construction, which captures RNAs of all sizes. Our findings confirm that miRNAs are present in the EV transcriptome and correlate with the miRNAs found in cellular RNA. Additionally, we identify many small non-coding RNAs in the EV transcriptome, known components of ribonucleoprotein (RNP) complexes that have enzymatic activity and are implicated in cellular processes such as RNA modification, DNA replication, and RNA quality control. We evaluate EVs from a variety of cell lines to confirm the reproducibility and effectiveness of our workflow. Moreover, this novel approach includes a non-traditional data analysis technique that provides new insights into the EV transcriptome and EV biology.

Session Title: Omics Technologies Poster Session II

PB3488 Structural similarity search in a comprehensive database of the human proteome.

Authors:

P. Russell, B. Hannigan; Generate Biomedicines, Somerville, MA

The ability to search entire proteomes for structural elements can eliminate bottlenecks in the prediction of molecular behavior inside cells. For example, in the context of drug development, substantial time and resources can be saved through computational identification of candidate molecules with potential off-target binders in the human proteome.

Structural similarity search across complete target proteomes has recently become possible due to advances in computational structure prediction. With the launch of the AlphaFold Protein Structure Database (AlphaFold DB) in 2021, high confidence structure predictions for several entire proteomes are newly available. These predicted structures supplement the structures solved through low-throughput structural analysis techniques such as cryo-EM, NMR and X-ray crystallography, accessible primarily through the Protein Data Bank (PDB). Despite the availability of structural search functionality in the PDB and several third-party tools supporting structure-based search of AlphaFold DB, there is no resource that enables researchers to easily search for structural motifs across a curated database of solved human protein structures supplemented with predicted structures.

We present the Human Motif Search ProteinStore (HMS: ProteinStore), a user-friendly application that takes in a protein structure and region of interest and identifies similar structural elements across a comprehensive structural database of the human proteome. The structural database is composed of a curated set of solved protein structures found in the Protein Data Bank and supplemented with predicted structures from AlphaFold DB to cover the complete UniProt human proteome. HMS: ProteinStore consists of a backend that executes the structural search against the database, as well as a frontend that takes queries in the form of a protein structure and region of interest. Results are returned as top structural matches with alignment scores.

Session Title: Omics Technologies Poster Session III

PB3489 Study of multi-omic variation across diverse peoples of rural Cameroon

Authors:

A. Harris[1], M. A. Rubel[2], M. E. Hansen[1], W. Beggs[1], M. Kachman[3], C. Fokunang[4], A. K. Njamnshi[4], S. A. Tishkoff[1]; [1]Univ. of Pennsylvania, Philadelphia, PA, [2]The Janssen Pharmaceutical Companies of Johnson & Johnson, San Diego, CA, [3]Univ. of Michigan, Ann Arbor, MI, [4]Université de Yaoundé I, Yaoundé, Cameroon

Many studies highlight the importance of the gut microbiome (GM) in shaping health. The GM is proposed to impact host metabolism, providing an axis of interaction between the digestive system and other organ systems. Associations with specific gut microflora have been identified in a spectrum of conditions, including diabetes, autism, colorectal cancer, and liver disease. However, knowledge about the human GM is disproportionately based on studies of urban cohorts from the Global North. These results might not transfer to the more than 40% of people worldwide living rurally—including over half of sub-Saharan Africans, per World Bank estimates. To better understand the composition of GMs from ethnically diverse rural populations, we generated shotgun metagenomic profiles for 178 study participants living across rural Cameroon, practicing hunter-gatherer (Baka and Bagyeli), pastoralist (Fulani), and agriculturalist (Bantu-speakers) lifestyles. We integrated GM profiles with targeted plasma lipidomic and host genomic data over the same individuals. GMs were highly variable across the cohort: permutation testing of non-omic features via PERMANOVA revealed that geography, lifestyle, sex, and gut parasite infection (all $p \leq 1.7 \times 10^{-3}$; $R^2 \geq 0.012$) were significantly correlated with GM β-diversity. In order to measure the contribution of host genetic and lipidomic variation to the GM abundance (genus) variation, we used Principal Components (PC) analysis. Upon regressing each GM PC on the host genetic PCs, host lipidomic PCs, and non-omics covariates, we find that the host genetics, lipidomics, and non-omics covariates explained 2.7%, 3.1%, and 4.1% of the GM variation, respectively. To find which components of the GM and lipidome interact, we used MiRKAT to detect lipids significantly associated with variation in GM β-diversity, and LASSO to select bacterial genera most associated with those lipids. We find that 30 of 1070 measured lipids (2.8%), primarily triacylglycerols, and 180 of 1385 genera (13%), 110 of which lack a standardized genus name, underlie the microbiome-lipidome correlations. Our work thus provides a more complete understanding of healthy GM diversity, and the factors that shape it, in rural sub-Saharan Africa. This work was supported by NIH grant 1R35GM134957-01 and American Diabetes Association grant ADA 1-19-VSN-02 to SAT, and NHGRI training grant T32 HG 9495 to AMH.

Session Title: Omics Technologies Poster Session I

PB3490 Systematic analysis of the impact of short tandem repeats on gene expression

Authors:

A. Goren, S. Benton, L. Zhang, Q. Gong, M. Maksimov, S. Shleizer-Burko, E. Wang, C. Wang, M. Lamkin, M. Gymrek; UC San Diego, San Diego, CA

Short tandem repeats (STRs, 1-6bp) are an intriguing type of genetic variation given the diverse repeat units (e.g. $(A)_n$, $(AC)_n$) or the polymorphism in the number of repeats. We and others have linked STRs to variation in gene expression and complex traits. These studies suggested STRs may impact transcription via multiple mechanisms. To study STR involvement in transcription regulation, we optimized a massively parallel reporter assay (MPRA) with random barcoding to handle low complexity sequences. We focused on 3-4 variants each of ~30,000 unique human promoter-proximal STRs for a total of 100,000 fragments. Our MPRA captured ~83% of the library, with >2.1M unique barcodes. We transfected this array to three cell lines (HEK293, RPE1 and HeLa), each in triplicate. The reporter expression was highly correlated between the triplicates in the same cell line (R=0.98), and to a lesser extent between different lines (R=0.62).

Of the ~30K promoter STRs, 8K drove reporter expression ("active"). Motif analysis of the genomic context of the active STRs showed enrichment of key transcription factors (e.g. SP family (p=1e-55) and KLF family (p=1e-47) motifs near GC-rich STRs). Of active STRs, 6.7K (~84%) showed an association between repeat unit number and reporter expression ("eSTRs"). eSTRs were enriched for GC-rich repeats and were strongly biased toward showing positive associations between copy number and reporter expression (4K, ~60%, p=4.6e-57). Our MPRA results aligned with population level data, with ~500 active and ~120 eSTRs of the ~1.1K variants that overlapped STRs significantly linked ($p<0.05$) to transcription (Geuvadis) or blood traits (UKB).

To study the role of STR composition in regulating transcription we designed a second array where we perturbed each of the top 300 active STRs (e.g. by altering repeat sequence or strand orientation). Transfection of this array allowed us to identify key properties of active STRs, such as how directionality can be impacted by repeat unit composition or strand orientation. Our results suggest certain repeat units (e.g. GC-rich) consistently drive positive associations with expression, whereas others (e.g. GTTT) consistently drive negative association, while some repeats show more context-dependent effects.

Our study provides a first in depth interrogation of how sequence characteristics of STRs and their contexts impact transcriptional regulation. Notably, while our initial experiment focuses on promoter STRs, building on new advances in MPRA design (e.g. genomic integration) would allow extensions of our approach to study STRs in other genomic regions (e.g. enhancers) which may work by alternative mechanisms.

Session Title: Omics Technologies Poster Session II

PB3491 Systematic discovery of gene-environment interactions underlying the human plasma proteome

Authors:

B. Sun[1], R. Hillary[2], D. Gadd[2], Z. Kuncheva[3], T. Mangelis[4], T. Lin[1], K. Ferber[1], H. mclanghlin[1], H. Runz[5], Biogen Biobank Team, R. Marioni[2], C. Foley[6]; [1]Biogen, Cambridge, MA, [2]Univ. of Edinburgh, Edinburgh, United Kingdom, [3]Optima Partners Ltd, London, United Kingdom, [4]Optima Partners, London, United Kingdom, [5]Biogen Inc., Cambridge, MA, [6]Optima Partners, Edinburgh, United Kingdom

Understanding how gene-environment interactions (GEIs) influence the circulating proteome could foster biomarker discovery and validation. The presence of GEIs can be inferred from single nucleotide polymorphisms that associate with phenotypic variability - termed variance quantitative trait loci (vQTLs). Here, we conducted the first genome-wide vQTL association studies on plasma levels of 1,463 proteins in UK Biobank (N=53,752). We detected 683 vQTLs for 571 proteins, all newly described. Effect sizes were correlated 96% with a multi-ancestry independent subset of UK Biobank. We then tested for possible interactions that underlie each vQTL association using 114 candidate environmental exposures. We uncovered 1,400 GEIs between 142 proteins and 101 lifestyle and metabolic exposures. The proportion of vQTLs that participated in a GEI was five times greater than conventional protein QTLs (i.e. additive effects, 20% vs. 4%), highlighting the strength of this approach. Stratified analyses within UK Biobank clearly illustrated how age, sex and genotype alter associations between plasma proteins and environmental exposures. Specific examples included a GEI between glycodelin (a reproductive glycoprotein) and body weight, where women under 55 years of age showed negative correlations between protein and phenotype. Women over 55 years showed positive correlations. The opposing signed-effects would mask each other in non-stratified analyses, and the association was entirely absent in males. Furthermore, using knowledge from our GEIs, we pinpointed biological mechanisms that explain why some sites are vQTLs only and lack conventional additive effects on protein levels. This study leverages an unrivalled resource of open-access proteomics and genomics to establish the most comprehensive database yet of GEIs for the human proteome. Importantly, others will be able to use our catalogue of vQTL effects to investigate GEIs with environmental exposures of interest to them, which will expedite the identification of genetic mechanisms underlying a diverse array of disease states.

Session Title: Omics Technologies Poster Session III

PB3492 Systematic identification of silencers in the human genome

Authors:

B. Pang[1], M. Snyder[2]; [1]Leiden Univ. Med. Ctr., Leiden, Netherlands, [2]Stanford Dept. of Genetics, Palo Alto, CA

The majority of the human genome does not encode proteins. Many of these noncoding regions contain important regulatory sequences that control gene expression. To date, most studies have focused on activators such as enhancers, but regions that repress gene expression—silencers—have not been systematically studied. We have developed a system that identifies silencer regions in a genome-wide fashion on the basis of silencer-mediated transcriptional repression of caspase 9. We found that silencers are widely distributed and may function in a tissue-specific fashion. These silencers harbor unique epigenetic signatures and are associated with specific transcription factors. Silencers also act at multiple genes, and at the level of chromosomal domains and long-range interactions. Deletion of silencer regions linked to the drug transporter genes ABCC2 and ABCG2 caused chemo-resistance. Overall, our study demonstrates that tissue-specific silencing is widespread throughout the human genome and probably contributes substantially to the regulation of gene expression and human biology. <!--EndFragment-->

Session Title: Omics Technologies Poster Session I

PB3493 Taming the Titans: Building an Ecosystem for Massive Scale Single-Cell RNA-Seq Data Analysis and Visualization

Authors:

Z. Ouyang; BioInfoRx, Inc, Madison, WI

The advent of single-cell (single-nuclei) RNA sequencing (sc/sn RNA-Seq) has revolutionized our understanding of cellular diversity, rare cell populations, and dynamic cellular processes. However, with the advancement of modern sc/sn RNA-Seq technologies, profiling millions of cells in a single experiment has become easily achievable, leading to massive datasets that pose challenges in storage, processing, and analysis. To address these challenges, the development of computational methods specifically tailored for analyzing massive-scale sc/sn RNA-Seq data is crucial to unlock the hidden biological insights within.

Previously, we published scRNASequest and CellDepot, an ecosystem for sc/sn RNA-seq data analysis, visualization, and publishing. This ecosystem has provided benefits to researchers across different domains, enabling biologists to gain insights, analysts to improve turnaround time, core facilities to host data projects, and publishers to examine data integrity. Here we present the next-generation development of this ecosystem to handle massive scale single-cell data.

One major hurdle encountered when analyzing massive scRNA-Seq datasets is the limitation of R sparse matrices, which do not support matrices with over 2 billion non-zero entries. To overcome this limitation, we have developed an updated solution that allows the analysis and visualization of tens of millions (and above) cells. The pipeline employs a novel split-combine approach. In this method, batch integration features were extracted after preprocessing all samples in multiple chunks, including quality control (QC), filtering and doublet prediction. Subsequently, various integration methods are applied on each sample chunk using these features, followed by harmonizing the data across all samples.

Effective visualization of massive sc/sn RNA-Seq data is equally important for interpreting complex biological insights. Building upon the cellxgene VIP data visualization platform, we introduce cellxgene VIP meta-cells. The new platform provides a simplified multi-level visualization on tens of millions of cells.

Through the continuous improvement of capabilities to analyze and visualize massive-scale scRNA-Seq data, the scRNASequest ecosystem empowers researchers to overcome computational challenges and extract valuable biological insights from these large complex datasets.

Session Title: Omics Technologies Poster Session II

PB3494 † Targeted long-read sequencing of *C9orf72* in multiple human tissues.

Authors:

E. Udine[1], N. A. Finch[1], M. DeJesus Hernandez[1], J. Jackson[1], T. Gendron[1], M. Ebbert[2], D. Dickson[1], L. Petrucelli[1], R. Rademakers[3], J. Shah[1], B. Oskarsson[1], M. van Blitterswijk[1]; [1]Mayo Clinic, Jacksonville, FL, [2]Univ. of Kentucky, Lexington, KY, [3]VIB Ctr. for Molecular Neurology, Antwerp, Belgium

The most frequently observed genetic cause of amyotrophic lateral sclerosis (ALS) and frontotemporal lobar degeneration (FTLD) is a non-coding hexanucleotide repeat expansion (GGGGCC) in the gene *C9orf72*. Little is known about how variation in the *C9orf72* expansion contributes to clinico-pathological heterogeneity across this disease spectrum. Advancements in long-read sequencing technologies now allow us to study repeat expansions at a greater resolution than traditional methods, like Southern blotting. Therefore, we aim to use long-read sequencing to uncover variation in the length, sequence content, and methylation levels of the *C9orf72* repeat expansion. Previously, we have used long-read sequencing to examine the *C9orf72* repeat expansion in the cerebellum and found that expansion length is associated with survival time and levels of dipeptide repeat (DPR) proteins, one of the hallmarks of *C9orf72*-related diseases. We are currently extending this targeted long-read DNA sequencing study that utilizes No-Amp sequencing to accurately examine the *C9orf72* repeat expansion in multiple brain regions and blood. No-Amp sequencing is being performed on genomic DNA and utilizes CRISPR-Cas9 to capture a specific genomic region. For this, we are extracting high-quality DNA from the human brain (n~40 per region), including frontal cortex, temporal cortex, motor cortex, and whole blood (n~30) and sequencing samples using the Sequel II (PacBio) at the Mayo Clinic Genome Analysis Core, enriching for the *C9orf72* repeat expansion. Raw data is being analyzed using PacBio's RepeatAnalysisTools. We have selected ALS and FTLD samples with detailed clinical and pathological information available, such as age at onset, age at death, *C9orf72* promoter methylation, *C9orf72* transcript levels, and measures of RNA foci and DPRs. Our work demonstrates that No-Amp sequencing can be used to sequence the entire *C9orf72* repeat expansion in single, long-reads. Thus far, we have detected a wide range of expansion sizes, ranging from less than 1kb to greater than 24kb. Initial analyses of longitudinal blood samples confirm that some individuals have expansions that are stable over time, while others are more dynamic. We have also validated the presence of paternally inherited contractions of the expansion. Our initial estimates of the sequence indicate that approximately 80% of the expansion is comprised of the pure GGGGCC motif and is methylated. Ongoing efforts are focused on further optimizing methylation analyses and identifying clinical and pathological associations.

Session Title: Omics Technologies Poster Session III

PB3495 Targeted sequencing of *ATM* for the genetic diagnosis of Ataxia Telangiectasia - A pilot evaluation of oxford nanopore technology

Authors:

H. Yeung, K-H. Tse; The Hong Kong Polytechnic Univ., Hung Hum, Hong Kong

Introduction Rare diseases (RDs) in childhood often exhibit catastrophic consequences with major neurological deficits. The infrequency of rare neurological disorders makes genetic screening difficult and clinical specialists scarce. Ataxia Telangiectasia (A-T) is one such disease with hundreds of possible pathogenic mutations on *ATM* (Ataxia Telangiectasia Mutated) gene. Targeted sequencing on the next generation sequencing (NGS) platform emerges as a cost-effective genetic diagnosis for A-T. Here, we evaluate if oxford nanopore technology (ONT)-based third generation sequencing may improve genetic diagnosis for A-T. **Methodology** Genomic DNA of A-T (n = 10, mean age = 26.2 years) and healthy control (n = 10, mean age =24.3 years) were isolated from frozen postmortem brain tissues kindly provided by NeuroBioBank, NIH. For Illumina-based NGS, fragmented gDNA (50 ng, 150 bp) was captured with TruSight Inherited Disease Sequencing Panel, and pooled library was sequenced on NextSeq 500 platform with established pipeline. For ONT, fragmented gDNA (1 µg, 1500 bp) was captured with Agilent ClearSeq Inherited Disease Panel and each captured library sequenced by a R9.4 nanopore flow cell on MinION Mk1b device. Real-time basecalling was conducted by Guppy (v6.5.7) on GPU. Reads from two platforms were then aligned to human reference genome (hg19) with genetic variants called by GATK (v1.6) and Clair3 (v1.0.2) respectively and annotated by ClinVar, COSMIC and dbSNPs database. **Results** For NGS, 2.71M reads (N50 = 103.6 bp) were obtained per sample with 1.03Gb reads aligned with targets with an average of 181.6x coverage. Among the A-T samples, 13 pathogenic variants were identified with missense indels or frameshift mutation. For ONT, an average of 3.72M reads (N50 = 1.55kb) was detected with 3.51Gb bases were called per sample. The average read depth aligned with targeted panel region was 67.5x. ONT was able to pin down the identical pathogenic variants in A-T found in Illumina-based NGS. The single nucleotide polymorphism identified in the *ATM* (including missense and frameshift variants) reported was consistent in A-T (80%) and control (66.7%) across the two platforms with comparable average read depth (ONT, 8x - 99x; NGS, 46x - 570x). Ongoing analyses are underway for genes responsible for RDs with deficits in DNA repair. **Conclusion** ONT-based targeted sequencing could identify pathogenic ATM variants in A-T. Targeted sequencing with NGS and/or ONT is useful for genetic screening of childhood RDs but warrants further investigations and standardization. **Funding** Health and Medical Research Fund (HMRF06173836), Hong Kong SAR Government; PolyU Start-up Fund: P0030307

Session Title: Omics Technologies Poster Session I

PB3496 Targeted single-cell transcriptome sequencing characterizes 1 million healthy and leukemic bone marrow mononuclear cells.

Authors:

B. Hariadi[1], E. Papalexi[1], G. Kim[1], A. Malinkevich[2], C. Roco[2], A. Rosenberg[2]; [1]Parse BioSci.s Inc., Seattle, WA, [2]Parse BioSci.s, Seattle, WA

In the thirteen years since its inception, single-cell RNA-sequencing (scRNA-seq) has rapidly spread across multiple fields of research, leading to many new discoveries. As technologies have matured, the number of cells that can be processed in a single experiment has seen exponential growth with workflows now assaying up to one million cells in an individual experiment. While high throughput sequencing methods have facilitated the discovery and characterization of various cell types, sequencing costs can be prohibitively high for routine use. Many applications of scRNA-seq are focused on cell type identification, gene regulatory networks, or biomarker discovery. These applications often do not require surveying the entire transcriptome, but rather require the interrogation of specific sets of well-characterized genes. In these cases, sequencing the entire transcriptome may be adding unnecessary costs. To increase throughput and minimize sequencing costs, the development of a targeted gene enrichment method is required. Here, we extend our whole transcriptome (WT) split-pool combinatorial barcoding technology to enrich a subset of genes representing hundreds of thousands of human bone marrow mononuclear cells (BMMCs) from three acute myeloid leukemia (AML), one acute lymphocytic leukemia (ALL), and four healthy donors. We used our Immune1000 panel to enrich genes representing canonical immune cell markers and pathways. Our method increased the percent of reads on target from as low as 7% in the whole transcriptome libraries to 75% in the targeted libraries. This enabled a nearly ten-fold reduction in sequencing reads between unenriched and enriched libraries, with the resulting clustering yielding high concordance of cell type identities and preserving leukemia-specific signatures such as FLT3, MKI67, and CD19. Overall, we demonstrate our modular enrichment strategy preserves biological structure and allows for deep characterization of gene signatures in health and disease. We envision our approach will enable researchers to simultaneously reduce sequencing costs while drastically scaling up the number of cells and samples across experiments.

Session Title: Omics Technologies Poster Session II

PB3497 Technical comparison between short-read and long-read RNA sequencing.

Authors:

Y. Song[1,2], Z. Chong[1,2]; [1]Dept. of Genetics, Heersink Sch. of Med., Univ. of Alabama at Birmingham, Birmingham, AL, [2]Informatics Inst., Heersink Sch. of Med., Univ. of Alabama at Birmingham, Birmingham, AL

The RNA sequencing technology, developed over a decade ago, has become an indispensable tool for analyzing the transcriptome. Next Generation Sequencing (NGS) typically produces short reads, which can present challenges in deciphering the full complexity of transcriptomes, particularly in repetitive regions. In contrast, long-read sequencing platforms, including Pacific Biosciences (PacBio) and Oxford Nanopore Technologies (ONT), have emerged as valuable tools for studying transcriptomes. These platforms provide full-length transcript sequences and enable more comprehensive analyses. This project investigated the transcriptome differences between long-read sequencing platforms and the Illumina short-read platform, focusing on gene coverage, comparing gene expression, and detecting isoforms. 12 samples from Human Genome Structural Variation Consortium (HGSVC) containing both PacBio Iso-seq and Illumina high-coverage RNA-seq data were selected. The top six regions from GRCh38.p13 annotation file were selected including protein-coding genes, lncRNA, retained introns, protein-coding CDS not defined, nonsense-mediated decay, and processed pseudogenes. The transcript numbers of these regions are 89305, 56049, 33947, 26468, 21234, and 10150, respectively. Long-read sequencing achieved nearly full coverage of selected regions based on the percentage of base pairs, while short-read sequencing fell short of reaching 80% coverage. When considering the extent of transcripts covered at over 95% of length, long-read sequencing demonstrated coverage rates of 21.6%, 4.3%, 13.6%, 6.6%, 10.5, and 3.7%; while short-read sequencing achieved coverage rates of 15.1%, 4.2%, 13.3%, 11.6%, 8.8%, and 5.58%. The correlation of gene expression between the long-read and short-read platforms was approximately 59%. Long- and short-read RNA sequencing can both identify known isoforms, but long-read RNA sequencing provides clearer results without noisy alignment. In conclusion, our investigation of these aspects will contribute to a better understanding of the strengths and limitations of long-read transcriptome sequencing platforms in comparison to the Illumina short-read platform, ultimately guiding researchers in selecting the most appropriate technology for their transcriptome studies.

Session Title: Omics Technologies Poster Session III

PB3498 Technical overview of machine learning pipelines used for assay predictions.

Authors:

T. Rajasekar[1], A. Musolf[2], B. Guo[2], K. Shkura[3], A. Vij[3], G. Bryman[2], K. Williams[4], C. Mann[5], J-H. Sul[2], S. Longerich[2], M. Kennedy[2], A. Tamburino[6], R. Mathew[2]; [1]Merck, Cambridge, MA, [2]Merck & Co., Inc., Cambridge, MA, [3]MSD, London, United Kingdom, [4]Merck & Co., Inc., Rahway, NJ, [5]Merck & Co., Inc., San Francisco, CA, [6]Merck & Co., Inc., West Point, PA

Machine learning (ML) in pharmaceuticals is becoming exceedingly popular, particularly for their predictive value. Through a cross-functional collaboration, our group has developed two automated machine learning pipelines. While these pipelines can be utilized to make predictions on multiple types of data, our initial validation used the results of a CRISPR assay to predict hits from the untested genome. CRISPR screens are useful in drug development for identifying targets and gene modulators, however, the screening process is time consuming. Our ML pipeline can streamline this process, enabling targeted genes (or selected targets) to be selected for testing with CRISPR assays based on results from predictive ML models.

We have developed two parallel pipelines (written in R and Python), utilizing 12 classification models in parallel on a dataset consisting of 40+ features derived from genetics, multi-omics, and known pathways. The pipelines are implemented on AWS using Databricks/MLFlow experiment tracker to optimize the pipelines and allow users to assess which model functions best at making predictions. Our implemented models were chosen as they covered a broad range of classifications algorithms, including random forest, gradient boosting, support vectors, and neural networks. An ensemble approach was chosen as it allows for data to be analyzed individually within each model, taking advantage of each model's inherent strengths/weaknesses for different datasets. It also allows for an effective way to compare multiple models using metrics such as precision, recall, and ROC. From these values, we can determine which of the models is most effective at making predictions. All models output feature importance metrics listing which attributes are most important for developing a predictive model. Simulation studies have shown this method to have high power under multiple scenarios and we present comparisons of models and the differences between the R and Python pipelines. Data science capabilities prove to be a beneficial addition to known genetic research practices, further expanding the scope in which the data can be used. This pipeline can also be leveraged for novel datasets across various therapeutic areas, including those using CRISPR assay data as well as other types of data for predictions.

Session Title: Omics Technologies Poster Session I

PB3499 Telomere dynamics in aging and cancer by nanopore long-read sequencing.

Authors:

T. Schmidt[1], C. Tyer[2], P. Rughani[2], C. Haggblom[1], J. Jones[1], J. Beaulaurier[2], X. Dai[3], S. Juul[3], S. Hickey[2], J. Karlseder[1]; [1]Salk Inst. for Biological Studies, La Jolla, CA, [2]Oxford Nanopore Technologies, Inc, Alameda, CA, [3]Oxford Nanopore Technologies, Inc, New York, NY

Telomeres are the protective, nucleoprotein structure at the ends of linear eukaryotic chromosomes. The accurate measurement of both telomeric length and composition of individual telomeres in mammalian cells has been challenged by the length and repetitive nature of telomeres. With Oxford Nanopore Technologies sequencing, it is now technically possible to sequence entire telomeres and map them to individual chromosome arms. Here, we report a reliable method to enrich, sequence and analyze human telomeres using native nanopore sequencing. To enrich for telomeric sequences, we combine the ligation of adapters complementary to the telomeric G-overhang with restriction enzyme digestion to sequence the telomeric C-strand and part of the adjacent subtelomere. The subtelomeric information is harvested to map individual telomeric reads to specific chromosome arms. We have measured bulk and chromosome-arm specific telomere length dynamics during cellular aging of cultured primary cells and in a patient-derived aging cohort. To address the impact of the telomere maintenance mechanism on telomere length and composition, we have sequenced five well-established telomerase- and ALT-positive cancer cell lines. Our results suggest that, based on nanopore telomere long-read sequencing, ALT-positive cells can be easily discriminated from normal and telomerase-positive cancer cells.
In summary, nanopore sequencing of telomeres grants a deeper understanding of individual telomere composition through telomere length measurement and mapping to specific chromosome arms. As such, telomere sequencing using our Oxford Nanopore enrichment method is a valuable tool to study telomere biology in aging and cancer.

Session Title: Omics Technologies Poster Session II

PB3500 The expression analysis using RNA-seq on human cerebellar tissues.

Authors:

K. Ishikawa; Tokyo Med & Dental Univ, Tokyo, Japan

Many neurological diseases are known to be caused by the expansion of RNA repeat sequences, and the number of such diseases is expected to increase in the future. In such repeat expansion disorders, abnormally expanded repeat sequences attract specific binding proteins, and in pathological states, binding of the specific protein to RNA repeats is accelerated that interferes protein functions. If the binding protein acts as a splicing factor, its binding to the repeat sequence may cause splicing dysregulation. In this study, we performed RNA-seq analysis in the cerebellum of SCA31, a repeat expansion disorder, where critical splicing factors such as TDP-43 are known to bind to the causative RNA repeat. We compared the result with the cerebellum of healthy subjects and SCA6 cerebellum as a control. The RNA-seq was performed using a 150 base-pair pair-end method as well as a long read sequencing. Data analysis was focused on genes such as *ITPR1* and *CACNA1A* that are particularly important in Purkinje cells, and mutations of both cause spinocerebellar ataxia in human. Of note that these genes are large in mRNA size (ITPR1: 9719 bp; CACNA1A: 8660 bp). In addition, it is easy to assume that the splicing abnormality causes severe damage to Purkinje cells, resulting in cerebellar ataxia. We mapped the RNA-seq results of healthy subjects and one subject each for SCA6 and SCA31, and compared their transcript counts. We first found that expression levels of *ITPR1* and *CACNA1A* were clearly lower in SCA31 than in healthy subjects, which is reasonable because the number of remaining Purkinje cells was lower in SCA31. On the other hand, the splicing pattern was very complex for both genes in the healthy subjects, and no significant difference in the splicing pattern was observed so far in an SCA31 patient. We are now checking long read sequencing results to identify if there is any splicing change is present in SCA31

Session Title: Omics Technologies Poster Session III

PB3501 The molecular signature of restrictive cardiomyopathy - a multi-omics approach to diastolic heart failure.

Authors:

N. Hanafi[1], A. Turinsky[1], L. E. Wybenga-Groot[2], M. F. Moran[2,3,4,5], S. Mital[6,7]; [1]Ctr. for Computational Med., Hosp. for Sick Children, Toronto, ON, Canada, [2]SPARC BioCtr., Hosp. for Sick Children, Toronto, ON, Canada, [3]Program in Cell Biology, Hosp. for Sick Children, Toronto, ON, Canada, [4]Dept. of Molecular Genetics, Univ. of Toronto, Toronto, ON, Canada, [5]Princess Margaret Cancer Ctr., Univ. Hlth.Network, Toronto, ON, Canada, [6]Dept. of Pediatrics, Univ. of Toronto, Toronto, ON, Canada, [7]Ted Rogers Ctr. for Heart Res., Toronto, ON, Canada

Introduction: Diastolic dysfunction, the central cause of heart failure in restrictive cardiomyopathy, is a predictor of adverse clinical outcomes. However, it is under-diagnosed in cardiomyopathy, difficult to treat, and little is currently known about disease mechanism.
Methods: We did RNA-sequencing and tandem mass spectrometry on myocardial samples from 49 pediatric patients with cardiomyopathy, and performed differential gene expression and protein abundance analyses. We applied a range of unsupervised and supervised computational methods, including machine learning classifiers to clinical, echocardiographic and multiomic data to identify patterns associated with diastolic dysfunction.
Results: We identified molecular patterns associated with restrictive cardiomyopathy in both the transcriptomic and the proteomic components. Supervised principal component analysis of the transcriptomic data revealed two distinct clusters of patients, both of which were enriched for patients with diastolic dysfunction. The first cluster primarily included patients with either restrictive cardiomyopathy or arrhythmogenic cardiomyopathy involving the left ventricle and correlated with upregulation of lipid biosynthesis genes. Cluster 2 primarily included patients with hypertrophic cardiomyopathy and severe diastolic dysfunction and correlated with upregulation of the gene *CA3*. Proteomics was also able to cluster patients with diastolic dysfunction, but was correlated with upregulation of proteins in the coagulation regulation pathway. We used machine learning to generate a prediction model for diastolic dysfunction using clinical, imaging and 'omics' data. The model incorporating multi-omics with clinical components had a stronger classification accuracy for diastolic heart failure compared to a clinical only classifier.
Conclusion: Our study identified a unique molecular phenotype of restrictive cardiomyopathy and diastolic heart failure, which presents exciting opportunities for the development of new biomarkers and therapies for diastolic heart failure targeted to the dysregulated lipid metabolic pathways.

Session Title: Omics Technologies Poster Session I

PB3502 † The Office of the National Coordinator's Sync for Genes Program is delivering on standards-based genomics-EHR integration.

Authors:

B. Dolin[1], P. Paul[2], S. Garcia[2], R. Freimuth[3], S. Chamala[4]; [1]Elimu Informatics, Richmond, CA, [2]US Dept. of Hlth.and Human Services, Washington, DC, [3]Mayo Clinic, Rochester, MN, [4]Children's Hosp. Los Angeles, Los Angeles, CA

Standardizing genomic information and making it interoperable for use in health and research systems is a complex yet crucial endeavor. The Sync for Genes project, launched in 2017 by the Office of the National Coordinator for Health Information Technology (ONC) in partnership with the National Institutes of Health (NIH), aims to enable the sharing of standardized genomic information among laboratories, providers, patients, and researchers. From the onset of this project, it was evident that the complexities of precision medicine and genomic data presented many challenges in adopting, implementing, integrating, and using such data. Over several years, and with the help of stakeholders, demonstrations, and contributions to the Health Level Seven (HL7®) Fast Healthcare Interoperability Resources (FHIR®) standard, the Sync for Genes project has created a well-vetted set of publicly available resources to help health and research organizations tackle these challenges.

One focus area was standardizing genomic variant sharing and interpretation for clinical knowledge. Annotating variants with various pieces of information (e.g., diagnostic or therapeutic implications, predicted molecular consequences, population allele frequencies) is critical in next-generation sequencing (NGS) analysis and interpretation. Standardizing the sharing of annotated variants utilizing FHIR and the knowledge structures defined by the Global Alliance for Genomics and Health (GA4GH) standards organization resulted in a set of resources produced by the Sync for Genes project currently deployed in a large healthcare setting.

Utilizing the Sync for Genes resources to standardize methods for sharing statically annotated variants (e.g., those coming in a lab report) and building on these resources to develop methods for sharing dynamically annotated variants by applying GA4GH-encoded knowledge into a clinical health record resulted in a deployed solution in which healthcare providers can access this information at the point of care. Additionally, this information is provided with context and clinical relevance, making it easier for a caregiver to consider the information as part of patient care and more informed decision-making. By decoupling the update of knowledge from the reporting of sequencing results, the pipeline presents the most current annotations to clinicians.

Tools developed under the Sync for Genes program to support the communication of dynamically annotated variants can be found here [https://github.com/FHIR/genomics-operations].

Session Title: Omics Technologies Poster Session II

PB3503 The Personal Genome: Evaluating Individual Sample *de novo* Genome Assemblies and Reporting Performance Benchmarks

Authors:

N. Duro[1,2], P. Schreiner[3,2], Y. Shi[2], X. Gong[2], A. Srour[2], S. Gautam[3], B. Gorman[3,2], S. Pyarajan[4,5]; [1]Booz Allen Hamilton, Tampa, FL, [2]VA Boston Hlth.care System, Jamaica Plain, MA, [3]Booz Allen Hamilton, Lexington, MA, [4]Partners Hlth.Care, Jamaica Plain, MA, [5]Harvard Med. Sch., Boston, MA

Individual human genomics is the key that will unlock the future of personalized diagnosis, prognosis, healthcare intervention, and precision therapeutics. Although, even after high-impact efforts have been made to better represent human diversity in a reference genome, significant challenges persist in the field that hinder personalized healthcare insights. We aim to produce new data, alongside novel analytics, to address some of these challenges by establishing high-quality references and QA/QC benchmarks for de novo genome assemblies.

We report the findings from our pilot study using PacBio HiFi reads that includes the assembly of 24 individual genomes in comparison to well-curated samples like HG002. Individual genomes are assembled using 1 to 4 SMRT cells on the PacBio Sequel IIe to provide insights into the breadth and depth of coverage as a function of the number of HiFi reads. We develop a fingerprinting method that is used to validate the identity of individual long read samples using signature SNPs. These same, deidentified individuals also have WGS data available that we use to compare short-read, long-read, and hybrid genomes. Further, we assess the amount of additional information achieved in a trio-based genome assembly. From these analyses, we observe de novo genome assembly metrics that are particularly descriptive and robust in the definition of high-quality genome assemblies.

We further detail benchmarks and relative performance describing long read-derived small & large variant identification and genome-wide DNA methylation. For small and large variant identification, we report benchmarks using DeepVariant and pbsv software, respectively. PacBio data also includes 5mc genome-wide methylation profiling, which we assess the relative concordance with respect to the Illumina Infinium MethylationEPIC microarray. The validity and consistency of targeted tandem repeat genotyping for these genomes is assessed using TRGT software. We anticipate that our work will provide insight that will help to establish standards, inform future research, and broadly drive continued research efforts in the field long-read sequencing.

Session Title: Omics Technologies Poster Session III

PB3504 The rate of RNA isoform discovery since 2015 and their expression in human frontal cortex brain tissue based on long-read sequencing.

Authors:

K. Dikobe[1], B. Aguzzoli Heberle[2], J. Brandon[2], K. Nations[2], M. Page[2], M. Wadsworth[2], S. Goodwin[3], E. Ghiban[4], R. Wappel[4], S. Mavruk Eskipehlivan[4], P. Nelson[2], J. Miller[2], J. Fryer[5], M. Ebbert[2]; [1]Univ. of Kentucky, Lexingthon, KY, [2]Univ. of Kentucky, Lexington, KY, [3]Cold Spring Harbor Lab., Cold Spring Harbor, NY, [4]Cold Spring Harbor Lab, Cold Spring Harbor, NY, [5]Mayo Clinic, Jacksonville, FL

Intro: Most human genes produce multiple RNA isoforms, averaging 7 across all genes. These distinct mRNA transcripts result in different protein sequences, and potentially distinct functions. For example, two isoforms from the BCL-X gene have opposite functions: BCL-XL is anti-apoptotic while BCL-XS is pro-apoptotic. Due to inherent limitations in short read sequencing, RNA isoform expression has been understudied. Recent advances in long-read sequencing allow researchers to explore the full breadth of RNA isoform diversity. Here, we compare Ensembl annotations since 2015 to our deep long-read RNAseq from human brain to: (1) quantify the recent dramatic increase of annotated transcripts; and (2) highlight RNA isoform diversity and its importance in human health and disease.
Methods: We sequenced frontal cortex tissue from 12 postmortem human brain samples (50% female) using one Oxford Nanopore PromethION flow cell per sample. We analyzed the data using guppy, pychopper, minimap2, and bambu. We quantified the number of newly annotated human transcripts in Ensembl by year from 2015-2022 and measured the RNA expression in frontal cortex brain tissue for transcripts annotated between 2019-2022.
Results: Ensembl annotations gained 8075 transcripts from 2015-2018, but there was a much more significant increase from 2019-2022 with 45000+ newly annotated transcripts. This significant increase coincides with the expanded adoption of long-read sequencing, though it is unclear whether this increase was, in fact, directly because of long-read data. We further found that 2797 of these transcripts are expressed in our long-read human frontal cortex data with a median Counts per Million (CPM) > 1. Of the 2797 transcripts, 868 came from medically relevant genes with 598 having a new protein coding sequence. One example is KIF5A-a gene associated with amyotrophic lateral sclerosis (ALS)-that had 80% of its total gene expression originating from a transcript with newly annotated protein coding sequence since 2019.
Conclusions: Our results demonstrate that the number of annotated human transcripts in Ensembl has grown significantly since 2019, where a large proportion of them are from known medically relevant genes. Researchers studying a given disease need to be aware of all isoforms for a given gene, and their relative expression patterns, to properly understand the gene's involvement in disease and how to target the gene. Intriguingly, thousands of transcripts annotated between 2019-2022 are expressed in human frontal cortex, including brain disease relevant genes, making these results especially important for researchers studying brain related diseases.

Session Title: Omics Technologies Poster Session I

PB3505 Tilted-CCA: Quantifying common and distinct information in multi-modal single-cell data via matrix factorization

Authors:

K. Lin[1], N. Zhang[2]; [1]Univ. of Washington, Seattle, WA, [2]Univ Pennsylvania, Philadelphia, PA

Paired multi-modal single-cell data profile multiple modalities for each cell simultaneously, such as the transcriptome alongside either the surface antibodies or epigenome. This new type of data has been growing in popularity in many areas of biomedical research and provides unique opportunities to learn how different modalities coordinate within each cell. In this work, we develop the Tilted-CCA to learn this coordination via dimension reduction. This novel method estimates low-dimensional embeddings that separate the axes of variation shared between both modalities (i.e., the "common geometry," capturing the coordination between both modalities) from the axes of variation unique to a particular modality (i.e., the "distinct geometry"). This task fundamentally differs from existing methods, which capture all the axes of variation from either modality instead. Methodologically, Tilted-CCA combines ideas from Canonical Correlation Analysis (CCA) and density clustering. Our method first uses the nearest-neighbor graphs from each modality to infer the common geometry between both modalities and decomposes the canonical scores from CCA to approximate this geometry. Biologically, we demonstrate that Tilted-CCA can enable many downstream analyses for CITE-seq (measuring the transcriptome alongside surface antibodies) and 10x Multiome (measuring the transcriptome alongside accessible chromatin regions) datasets on various biological systems. We focus on two detailed analyses in this work. The first shows that Tilted-CCA aids surface antibody panel design for finding a concise set of antibodies that best compliments the transcriptome for CITE-seq instruments. The second shows that Tilted-CCA can unveil cellular dynamics in developmental systems based on the proportion of variation between the common and distinct embeddings for RNA-ATAC multiome datasets.

Session Title: Omics Technologies Poster Session II

PB3506 Trinity of chromatin architects - the coordination of CTCF, RNAPOL2 and Cohesin proteins in shaping the spatial landscape of human genomes

Authors:

D. Plewczynski[1,2], A. Agarwal[3], M. Chilinski[2], S. Korsak[2], A. Halder[2]; [1]Univ. of Warsaw, Warsaw, Poland, [2]Warsaw Univ. of Technology, Warsaw, Poland, [3]CENTRE OF NEW TECHNOLOGIES, Univ. OF WARSAW, Warsaw, Poland

The higher-order organization of chromatin within the nucleus can be deciphered through advanced high-throughput next-generation sequencing techniques like Hi-C, ChIA-PET, and HiChIP, providing insights into the spatial landscape of human genomes at the population scale. Structural proteins, including the CCCTC-binding factor (CTCF), RNAPOL2, and the cohesin complex, are instrumental in stabilizing the spatial arrangement of chromatin. These proteins play critical roles in establishing and maintaining the long-range chromatin interactions, forming chromatin loops and topologically associating domains, and coordinating communication between genes with their regulatory elements, such as enhancers. In our work, we explore the precise contributions of CTCF, RNAPOL2, and cohesin in shaping the multi-scale three-dimensional architecture of chromatin exploiting 3C-type experimental methods combined with the advanced Artificial Intelligence (AI) computational models. Specifically, we answer the question: how the static architecture defined by CTCF micro-condensation is transformed by the dynamic activity of cohesin via the loop extrusion model (LEM) and reorganized during transcriptional activity by RNAPOL2. Finally, we enhanced the previously proposed 3d-gnome algorithm by training a deep learning transformer model (BERT) on chromatin long-range interactions from ChIA-PET experiments mediated by all three protein factors. The Artificial Intelligence powered model of chromatin looping use only DNA sequence as an input, and can be used as the powerful tool to predict the 3D structure of chromatin at the population scale. This allowed us to study the impact of structural variations on genome topology for various human populations by exploiting the 1000 genomes project DNA sequences. By integrating high-throughput sequencing data with advanced computational techniques, 3DNBERT enables the identification and characterization of genomic structural variations, such as copy number variations (CNVs), insertions, deletions, and inversions, across diverse individuals. These structural variations have a profound influence on the three-dimensional organization of the genome, leading to alterations in chromatin looping, topologically associating domains (TADs), and higher-order chromosomal architecture.

Session Title: Omics Technologies Poster Session III

PB3507 TRuc: A GraphAligner based method to characterize tandem repeat expansions from Nanopore sequencing.

Authors:

M. Khan[1,2], S. Chaudhary[3], C. Pearson[1,2], J. Simpson[3,2], R. Yuen[1,2]; [1]The Hosp. for Sick Children, Toronto, ON, Canada, [2]The Univ. of Toronto, Toronto, ON, Canada, [3]Ontario Inst. for Cancer Res., Toronto, ON, Canada

Approximately 7% of the human genome is comprised of tandem repeats (TRs) - stretches of DNA sequence repeated adjacent to one another. These repeats are polymorphic in sequence and length across the population, and tandem repeat expansions (TREs) beyond a pathogenic length are associated with over 60 neurological, neurodevelopmental, and neuromuscular disorders. Larger TREs lead to earlier onset, increased severity, and faster progression of disease. DNA methylation and sequence purity within a repeat can also affect disease outcomes. Traditional PCR diagnostic methods are unable to effectively report sizes for large, GC-rich repeats, while southern blotting offers a lower resolution. Neither approach provides information on methylation, or specific sequence context. While next-generation sequencing overcomes some of these challenges, it is unable to correctly parse large, repetitive sequences.

Long-read sequencing by Oxford Nanopore Technologies (ONT) overcomes these limitations by providing base-pair sequence resolution and methylation information. Here, we present TRuc (Tandem Repeat unit counting), a novel GraphAligner-based method to estimate tandem repeat sizes from Nanopore sequencing data, given the repeat motif and coordinates. We demonstrate the utility of this method on sizing TREs in 11 repeat-containing plasmids, and 7 patient cell lines harbouring known disease-associated TREs of various unit sequence. This method can be used to accurately size repeats in low-coverage datasets, and identify mosaic expansions. Comparison of TRuc repeat size estimates to molecularly validated 'truth' sizes in our sample set results in Pearson correlation coefficients from R=0.82 to R=0.99 for plasmid and patient samples. We compare the accuracy of TRuc to three other existing TR-sizing methods, and demonstrate that TRuc performs as well as, or better than, other methods. We also demonstrate that sequencing with R9.4.1 ONT chemistry results in strand-bias and sequence errors, particularly in GC-rich repeats, and show that these are resolved with higher accuracy and modified ONT basecalling configurations.

This work highlights the benefit of Nanopore sequencing for characterization of TRs. We reveal how different ONT basecallers and configurations affect the accuracy of tandem repeat analysis from Nanopore sequencing data. We also demonstrate the utility of TRuc in sizing heterogenous and mosaic TREs; this can be similarly applied to study complex diseases, or microsatellite instability in cancer. Implications of such methods may also include more accurate prognosis, and faster clinical diagnosis of tandem-repeat associated diseases.

Session Title: Omics Technologies Poster Session I

PB3508 Two-stage Bayesian network (BN) causal structure learning for discovering ncRNA-gene regulatory network.

Authors:

H. Ke, T. Ma; Univ. of Maryland, Coll. Park, College Park, MD

Despite the fact that non-coding RNAs (ncRNAs) make up the majority of the human genome, their annotation and functional understanding are still generally under-studied compared to the coding region. Increasing evidence suggests that ncRNAs play important regulatory roles in various biological processes, such as gene expression. Nevertheless, the precise underlying regulatory mechanisms of most ncRNAs remain elusive.
The interaction between ncRNAs and genes is a complex phenomenon: it has various regulatory patterns, including up-regulation and down-regulation, and mixed-regulation; a single ncRNA can potentially regulate multiple genes simultaneously, while a single gene can also be regulated by multiple ncRNAs at the same time; the regulatory mechanisms are likely to vary under different physiological conditions. In addition, there also exists high ncRNA-ncRNA and gene-gene correlations that might create spurious correlation between ncRNAs and genes. While high-throughput technologies enable us to examine these relationships on a genome-wide scale, it also poses a challenge due to the high-dimensionality of both ncRNA data and gene expression data. In this paper, we propose a novel two-stage hybrid Bayesian network (BN) structure learning method for discovering ncRNA-gene regulatory network with both high-dimensional ncRNA and gene expression data. In the first stage, we employ fast iterative constraint-based screening algorithms to filter out irrelevant ncRNA-ncRNA, gene-gene as well as ncRNA-gene edges, thereby mitigating the confounding bias and reducing the search space of causal structure. In the second stage, we develop a score-based Markov Chain Monte Carlo (MCMC) sampling approach that fully utilizes the unique bipartite backbone structure of ncRNA-gene regulation and incorporate prior knowledge from previous studies or external databases to guide the final structure learning of ncRNA-gene regulatory network. Comparing to existing BN structure learning methods which deal with only one single type of node (e.g., genes only), our method considers two types of nodes and their unique structures and uses existing biological knowledge to guide structure learning and improve the computation efficiency.
Comprehensive simulation studies and real data applications have demonstrated the advantage of our method on both correct causal structure identification and computation efficiency. We also developed a user-friendly software with GUI interface to implement our method with interactive ncRNA-gene regulatory network visualization.

Session Title: Omics Technologies Poster Session II

PB3509 Uncovering isoforms modulating early endothelial cell differentiation via a system-scale approach

Authors:

M. Mehlferber[1], V. Pavelko[1], D. Wissel[2], E. D. Jeffery[1], M. D. Robinson[2], G. M. Sheynkman[1]; [1]Univ. of Virginia, Charlottesville, VA, [2]Univ. of Zurich, Zürich, Switzerland

Isoform variation is known to contribute to the earliest establishment of cell fates; however, the role of isoforms in the developing vascular system remains understudied. To characterize isoforms involved in the establishment of endothelial cells, we utilized an induced human pluripotent stem cell line to model early endothelial cell (EC) differentiation. To gain insight into temporal isoform expression patterns, in collaboration with PacBio, we collected cells on five days of differentiation for long-read sequencing; obtaining bulk MAS-Seq data across five days with biological triplicates for the first (Day 0) and last (Day 5) timepoints. We also performed mass-spectrometry analysis on the differentiated cells, thus forming the basis for an integrative long-read proteogenomic approach. MAS-Seq libraries were sequenced on the Revio SMRT Cell system, giving ~60 million full-length reads per sample, representing the most comprehensive repertoire of full-length transcripts from early primordial endothelial cell differentiation. To characterize the isoform drivers of EC differentiation, we surveyed differential isoform usage from Day 0 and 5, identifying 2,159 genes with statistically significant differential isoform usage from DRIMSeq. The top genes showing dramatic isoform switches were RBM7, PIAS1 and GNAS, previously implicated in endothelial processes. To confirm the expression of these isoforms at the protein-level, we utilized a deep coverage MS-based proteomics approach, collecting data from multiple proteolytic digestions (Trypsin, AspN, Chymotrypsin, and Lys-C) to expand the identification of protein isoforms within our collected mass-spectrometry samples. Focusing on the proteomics data collected from Day 0, we identified proteins from an average of 9,370 genes. Through the enhanced resolution of the MAS-Seq Revio system, in tandem with the proteomic data, we present one of the most comprehensive surveys of protein isoform expression evidence during early developmental pathways. The findings from this work lay the foundation to build a more comprehensive framework for defining functional isoform regulatory modules during development, and technical assessment of parallel RNA and protein data collection for iPSC and other tissue resources.

Session Title: Omics Technologies Poster Session III

PB3510 Unraveling disease-driving genes and cell types through optimized integration of single-cell transcription and genome-wide association studies

Authors:

X. He, Q. Li, X. Zhu; Penn State Univ, University Park, PA

How disease-associated SNPs confer disease risk remains unknown. Recent studies have begun to integrate GWAS with scRNA-seq to uncover genes and cell types by which SNPs drive diseases. These studies often used a single method to identify cell-type-specific genes from scRNA-seq (e.g., differential expression) or link them to SNPs in GWAS (e.g., proximity). However, relying solely on one method could be misleading, because a method that excels in one study might fall short in another, and determining the optimal method for a new study is often uncertain. Hence, there is a need for a framework that synergizes various approaches to automate the optimal integration of scRNA-seq and GWAS.

Here, we develop an automatic framework that optimizes the integration of scRNA-seq and GWAS to find disease-driving genes and cell types. First, we use 6 metrics of expression specificity (ES) to identify genes of the transcriptional signature for a cell type on scRNA-seq. Next, we link these genes to their regulatory SNPs through 5 strategies based on SNP-to-gene (S2G) proximity, bulk and single-cell epigenomes. Lastly, we quantify heritability enrichment of these SNP annotations in GWAS via stratified LD score regression, and we normalize the results to be comparable across datasets, ES metrics and S2G strategies. By optimizing the normalized results, our framework automatically selects the best combination of ES metric and S2G strategy for a given study, which advances over previous studies.

We applied this framework to 4 scRNA-seq studies of healthy human hearts (n=40; 741,907 cells) and GWAS for 20 cardiovascular traits. Our results recapitulated known disease-associated cell types, including cardiomyocytes (CMs) in atrial fibrillation (AF) and endothelial cells in coronary artery disease (CAD). Our results identified previously unexplored but plausible cell-disease relations, such as pericytes in CAD (9.9-fold, p=1.4e-4), demonstrating the hypothesis-generating potential. Our results showed the benefit of S2G via cell-type-matched single-cell epigenomes. For example, AF heritability enrichment in CM-specific genes increased from 21.3-fold (p=2.5e-7) to 41.8-fold (p=3.4e-7) when we switched S2G from bulk epigenomes of heart tissues to single-cell epigenomes of CMs. Given a disease, our results prioritized the same cell types on European and East Asian GWAS, highlighting the cross-population transferability.

While focusing on heritability, our framework identifies cell-type-specific gene sets that can aid gene prioritization and polygenic prediction. Though our framework was tested mainly on cardiovascular GWAS, it can be applied to study other diseases.

Session Title: Omics Technologies Poster Session I

PB3511 Unraveling dynamic enhancer-promoter interactions across neural differentiation using network analysis models

Authors:

A. Kreimer, **W. DeGroat**; Rutgers Univ., Piscataway, NJ

There is an emerging consensus that a substantial portion of the mutations associated with human diseases in genome-wide association studies (GWAS) likely occur in cis-regulatory elements, especially gene-distal enhancers, that mediate their effects by causing a change in the expression of a target gene. However, our understanding of how the variation in this 98.5% of the sequence that constitutes the human non-coding genome contributes to human disorders remains largely unknown. Here, we collected epigenetic and gene expression data from seven early time points during the differentiation of human embryonic stem cells (hECSs) into neural progenitor cells (NPCs). Focusing on this model system, we constructed an intricate network of enhancer-promoter (E-P) interactions which demonstrated characteristics suggesting dynamic network patterns across time points. Temporal analysis suggests timepoint-specific characteristics. Next, we collapsed all the time points into one network and performed enrichment analysis which showed relevant genomic annotations. We applied bi-clustering techniques by grouping enhancer and promoter regions together using a modularity metric, which unraveled various network sub-structures of interactions suggesting different transcriptional regulation programs. Finally, we identified overlap with variants from various neuropsychiatric and neurodevelopmental disorders within our constructed network with matching disease-associated target genes. Analysis of the transcription factors (TFs) predicted to bind to the enhancers in our network suggests a mechanism by which these variants can exert their effect. Overall, our results suggest a general framework for exploring gene regulation programs and their dynamics across developmental processes and a comprehensive approach to exploring the effect of disease-associated genetic variation on transcriptional networks.

Session Title: Omics Technologies Poster Session II

PB3512 Unraveling Layers of Proteogenomic Complexity in Cancer through Multiomic Exploration of Archived FFPE Tissue

Authors:

H. Latif[1], Y. Han[1], D. Corney[2], Y. Fan[1], W. Wang[1], J. Matthew[1], Z. Ying[1], C. Mozdzierz[1], G. Zhou[1]; [1]Azenta, South Plainfield, NJ, [2]Azenta, South Planfield, NJ

Formalin Fixed Paraffin Embedded (FFPE) tissues are a staple in clinical diagnostics associated with solid tumor. The preserved tumor tissue can be interrogated using classical histopathology and complemented with molecular approaches to elucidate cancerous SNPs, indels, structural variants, repeat expansion, copy number expansions, and mutational burden. Increasingly, approaches are being deployed to interrogate tissue not only through individual datasets but through integrated approaches that provide spatial information that defines the tumor microenvironment. Here, we illustrate using FFPE blocks from various tissues a multiomic workflow that allows for deep exploration of the molecular underpinnings of cancerous tissue. FFPE blocks were serially sliced into various FFPE slides with a single slide H&E stained. Individual slides were then utilized to explore the genome, epigenome, single cell RNA-seq, and digital spatial profiling. Genome information was captured using hybrid capture based approaches followed by deep NGS on an Illumina platform and analyzed for a variety of variants and tumor mutational burden. DNA methylation was detailed using target capture probes targeting DNA methylation sites. Single cell approaches were applied to explore the transcriptome using 10X Genomics scRNAseq Flex kit, and digital spatial profiling was done using the NanoString GeoMx Whole Transcriptome Atlas and immunostaining. Aggregating these data illustrates the expanding landscape of information that can be extracted from FFPE derived tissue and the potential for novel discovery and diagnostic power integrating these complex data types holds.

Session Title: Omics Technologies Poster Session III

PB3513 Unravelling Myoblast Potential: A Breakthrough in CRISPR/Cas9 Efficiency.

Authors:

A. Rihoux[1,2], A. Gagné[1,2], L. K. Hamilton[1,2], M. Tétreault[1,2]; [1]Univ. of Montreal, Montreal, QC, Canada, [2]CRCHUM, Montreal, QC, Canada

Until now, the use of CRISPR/Cas9 in myoblast research has been challenging and has yielded low editing efficiency. In this study, we present a transformative protocol, effectively integrating Neon Transfection system with Single Cell Cloning on immortalized myoblasts. This cohesive approach magnifies the yield of viable clones nine-fold and triples the occurrences of successful knockout mutations. More clones translate to an increased likelihood of obtaining colonies with desired mutations, thereby significantly improving the success rate of CRISPR manipulations.

Our method's success is underpinned by the integration of the high-resolution melting (HRM) curve technique for clone screening, replacing the time-consuming and expensive sequencing processes. HRM reduces screening time from days to just a couple of hours and significantly cuts costs by minimizing cell culture resources.

This innovative protocol not only speeds up myopathy research but also proposes an alternative to patient biopsies. It eliminates lengthy paperwork and patient discomfort associated with biopsies, making the process both faster and more humane. The unified approach of Neon Transfection, Single Cell Cloning, and HRM screening promises a higher success rate while avoiding bulk effects, thus maintaining the purity of the clones. This breakthrough holds substantial potential to redefine myoblast research and offers exciting possibilities for future myopathy studies and treatments.

Session Title: Omics Technologies Poster Session I

PB3514 Unravelling the interplay between type 2 diabetes, genetics and metabolite levels

Authors:

O. Bocher, A. Singh, A. Arruda, P. Kreitmaier, A. Barysenka, W. Rayner, E. Zeggini; Helmholtz Zentrum Munich, Neuherberg / Munich, Germany

Type 2 diabetes (T2D) represents a major health burden and is forecast to increase dramatically in the upcoming years. The genetics of the disease have been successfully investigated in large GWAS but the remaining challenge lies in fully understanding the role of these variants in the biology giving rise to the disease. Metabolomics offers an opportunity to answer this question by giving insights into these biological mechanisms and into why certain patients progress to specific complications. We sought to investigate the interplay between genetics, metabolomics and T2D risk in the UK Biobank cohort. We first conducted a bidirectional Mendelian Randomization (MR) study using the UK Biobank to assess the effects of metabolites on T2D risk and the DIAMANTE GWAS meta-analysis results to assess the effects in the opposite direction. We found suggestive evidence for some metabolites to be causal of T2D, including glucose and metabolites linked to low-density lipoproteins. In the reverse direction, a lot more metabolites showed significance, possibly due to the higher statistical power reflecting the larger study size of the DIAMANTE meta-analysis. We find changes in half of the 164 absolute metabolite levels tested to be caused by T2D (with p-value down to 10^{-61}), including an increase in amino acids and glucose levels, and a decrease in metabolite levels from cholesterol classes. Some of these metabolite levels are also seen to be associated with specific T2D complications in the UK Biobank cohort such as HDL cholesteryl esters showing lower values in T2D individuals with kidney complications compared to T2D individuals without complications (β=-0.55, $p=3.66x10^{-8}$). Secondly, we assessed the interaction between T2D status and genetic variants through a differential metabolite QTL analysis. We find 22 metabolites that are differentially genetically regulated in individuals with and without T2D, including glycine (β=0.41, $p=5x10^{-25}$ for the most significant SNP) and low-density lipoproteins (β=0.53, $p=1.61x10^{-12}$) which show the most prominent signals. More than a third of these 22 metabolites were found to be caused by T2D. This work provides a better understanding of the metabolic changes induced by the occurrence of T2D. While further work is needed to confirm these results and better disentangle the underlying biology, they provide potential directions to investigate T2D consequences and subsequent complications.

Session Title: Omics Technologies Poster Session II

PB3515 Unveiling novel actionable targets: Proteomic GWAS insights into 1790 largely unstudied proteins.

Authors:

J. Kuliesius, P. Timmers, P. Navarro, L. Klaric, J. Wilson; Univ. of Edinburgh, Edinburgh, United Kingdom

Genome-wide association study (GWAS) power lies in testing hundreds of thousands of single nucleotide polymorphisms (SNPs) across many genomes for associations with a trait. Recent technological developments in high-throughput proteomic assays allow simultaneous quantitative measurement of thousands of proteins in a single sample. In this study we investigated 6432 plasma proteins using the SomaLogic aptamer-based technology in the Viking Health Study - Shetland, an endogamous population with a relatively low genetic diversity.
A total of 505 significant independent protein quantitative trait loci (pQTL) were found for 455 proteins in human blood plasma (382 cis ($P < 5x10-8$), 123 trans ($P < 6.6x10-12$)). Of these, 31 cis-pQTL were for proteins previously unstudied in large-scale proteomic GWAS. We leveraged this new resource to perform causal inference using bidirectional Mendelian Randomization and Colocalization against complex traits of biomedical importance. Causality was established for 18 little-studied proteins, with hitherto undiscovered links to type-2 diabetes, prostate cancer, depression, and other disease outcomes.

Session Title: Omics Technologies Poster Session III

PB3516 Unveiling the Genomics of Refractory Epilepsy: Human Diploid Genome Assembly using Long-Read Sequencing Technologies (PacBio HiFi).

Authors:

J. Diaz[1,2], L. Guio[3], R. Naranjo[3], J. Gúzman[3], J. Pérez[3], Ó. Zorro[3], D. García[2], P. Siauchó[2], S. Maradei[2,3], J. Duitama[1]; [1]Univ. de los Andes, Bogotá, Colombia, [2]Biotecnología y Genética SAS Biotecgen, Bogotá, Colombia, [3]Fundación Hosp. Pediátrico De La Misericordia HOMI, Bogotá, Colombia

Refractory epilepsy (RE) is a debilitating condition which severely impacts the quality of life, and healthcare costs on patients. Epidemiological evidence reveals that 20% to 40% of individuals diagnosed with epilepsy will ultimately develop resistance to conventional treatments. Understanding the genomic basis of this condition is crucial to enhance the clinical diagnosis of RE. In this study, we employed PacBio HiFi sequencing, to build a phased genome assembly for a patient diagnosed with RE. A de novo assembly of the error-corrected reads using Hifiasm generated a primary genome assembly with median contig length (NG50) of 60 Mbp, and total length of 3.1 Gpb. The total lengths of the individual haplotypes generated by HifiAsm were found to be similar to those obtained for the primary assembly. (3.04 Gpb and 3.0 Gpb, respectively). However, the NG50 values reduced to about 30 Mpb, and 24 Mpb, respectively. All assemblies covered approximately 90% of the GrCh38 reference genome, according to statistics obtained with the Inspector software. Evaluation of the assemblies using BUSCO revealed a high level of completeness, ranging from 93.2% to 95.8% of conserved genes in the primate lineage. Moreover, we were able to recover and annotate ~96% of protein coding genes within each haplotype assembly. Among these genes, approximately 3.6% were identified as duplicated genes. Taking into account that structural variants (SVs) could have causal effects on neurological and psychiatric disorders, we identified and genotyped 33,360 SVs, comparing the reads with the T2T genome assembly. A total of 27,684 of these SVs (83%) were heterozygous and intersect 4,483 protein coding genes. Through the utilization of long-read sequencing techniques, this study seeks to enrich the growing body of knowledge pertaining to this multifaceted condition, with the potential to unveil novel insights into its genetic underpinnings.

Session Title: Omics Technologies Poster Session I

PB3517 Use of long read sequencing and optical genome mapping to solve unsolved rare Mendelian diseases.

Authors:

E. Wohler[1], C. Antonescu[1], M. Schatz[1], R. Martin[1], S. Raskin[2], N. Sobreira[1]; [1]Johns Hopkins Univ., Baltimore, MD, [2]GENETIKA, Curitiba, Brazil

Chromosomal disorders are the most frequent cause of intellectual disability in our population, however, deletions and duplications shorter than 300 kb and longer than 50 bp, inversions, and translocations are not easily identified by FISH, arrays or by short-read sequencing methods. Consequently, structural variants represent the largest unexplored reservoir of genetic variation. With average read lengths now over 10 kb and some reads exceeding 100 kb, the long-read technology has dramatically improved sensitivity over short-read sequencing, FISH, or arrays to detect structural variants. The new optical genome mapping (OGM) methodology also has the potential to address this important class of genomic variation and contribute to cases that fail to be solved by short-read sequencing. Here, we performed OGM on 5 probands. Three probands had known germline duplications of uncertain clinical significance identified by SNP array. SNP array could not detect if these duplications were in tandem or inserted elsewhere in the genome. Using OGM we found that all three duplications were in tandem. Next, we will use RNAseq to determine which affected genes are disrupted by these duplications and if they could be the cause of the phenotypes being investigated. We also performed OGM in two families with rare Mendelian diseases unsolved by WES, short-read and long-read WGS. In Family 1, 8 males in 2 generations were affected with an apparently X-linked Treacher-Collins-like phenotype. OGM identified a ~368 kb hemizygous tandem duplication with a partial inversion at Xq26.2 in the proband and his unaffected mother. SNP array is being used to confirm the segregation of this rearrangement in other family members. Family 2 is a non-consanguineous family segregating an apparently X-linked myopathy. The proband and 3 other males over 3 generations have a congenital progressive myopathy. OGM identified a tandem hemizygous duplication of ~807 kb at Xq22.1q22.2 in the proband and his affected maternal uncle. Further analysis of the long-read sequencing also identified this duplication, but the size was ~832 kb and now it included the *PLP1* gene. Duplications of this gene have been associated with Pelizaeus-Merzbacher disease that matches the phenotype of this family. Further SNP array analysis showed that this duplication is also present in the maternal great uncle. In summary, a combination of methodologies is necessary to precisely identify possible causative structural variants, determine their familial segregation, and understand the effect of these rearrangements on the transcription of the genes involved and their relationship with the diseases being investigated.

Session Title: Omics Technologies Poster Session II

PB3518 Using CUT&RUN/Tag with a portable nanopore sequencing device

Authors:

P. Hook, W. Timp; Johns Hopkins Univ., Baltimore, MD

Introduction: Many methods have been developed using DNA sequencing to measure protein-DNA interactions. Two recently developed methods, Cleavage Under Targets and Release Using Nuclease (CUT&RUN) and Cleavage Under Targets and Tagmentation (CUT&Tag), have become extremely popular. These assays are typically used with short-read sequencing platforms, but those platforms have limitations including high cost, limited read lengths, and time needed for sequencing runs. We sought to overcome these limitations by testing sequencing CUT&Tag and CUT&RUN libraries with Oxford Nanopore Technologies (ONT) minION single-molecule sequencer. Sequencing with ONT devices is ideal due to a small physical footprint, lack of read length limitations, the ability to perform real-time analysis, and the ability to measure primary sequence and DNA modifications simultaneously.
Methods: We performed both CUT&RUN and CUT&Tag on the GM12878 cell line including experiments targeting histone marks (H3K27me3 and H3K4me) and a transcription factor (CTCF) as well as negative control experiments (IgG). Libraries were prepared using PCR amplification as if they were to undergo sequencing on an Illumina platform. Additionally, purified DNA fragments from CUT&RUN experiments were directly sequencing without amplification. All libraries were sequenced on a minION device from ONT. Sequencing was compared to publicly available CUT&RUN and DNA methylation data.
Results: Amplified CUT&Tag and CUT&RUN libraries sequenced with ONT showed concordance with short-read datasets. Expected patterns were observed including H3K4me3 enrichment at promoters with negative controls displaying no such enrichment. With unamplified CUT&RUN, we found sequenced fragments displayed nucleosomal laddering including tetra- and pentanucleosomes (>600 base pairs), which cannot be sequenced on short-read platforms. Unamplified H3K4me3 and H3K27me3 samples showed high concordance with amplified profiles as well. DNA modifications (5-methylcytosine and 5-hydroxymethylcytosine), called on unamplified libraries, showed expected patterns such as hypomethylation in active promoters.
Conclusion: By combining CUT&RUN and CUT&Tag with nanopore sequencing, we have enabled affordable benchtop sequencing and measurement of protein-DNA interactions. We have also demonstrated that CUT&RUN can be transformed into a real-time, multimodal, single-molecule assay without the need for additional molecular biology manipulations. These advancements will expand the portability of protein-DNA binding assays, increasing accessibility to these assays in a range of environments.

Session Title: Omics Technologies Poster Session III

PB3519 Using interpretable machine learning to understand variation in the human genome.

Authors:

N. Dwarshuis[1], P. Tonner[1], N. Olson[1], F. Sedlazeck[2], J. Wagner[1], J. Zook[1]; [1]NIST, Gaithersburg, MD, [2]Baylor Coll. Med., Houston, TX

The Genome in a Bottle (GIAB) consortium generates variant benchmarks for a set of human genomes to enable evaluation and comparison of sequencing technologies and variant detection methods. While these technologies have advanced significantly, correctly calling variants in complex or repetitive regions remains a challenge. We generally understand that sequencing biases and short read lengths can lead to incorrectly-called variants; however, we lack a data-driven model that uses GIAB benchmarking metrics to link variant caller performance to quantifiable features of the context surrounding a given variant.

We made such a model using explainable boosting machines (EBMs) which fit data using a linear combination of arbitrary univariate and bivariate functions (a generalized additive model with interaction terms). Despite being flexible, the simplicity of EBMs allows a human to easily understand the functional relationship and relative contribution of each feature. We exploit this transparency by fitting EBMs to variant call errors as a function of genomic context, which enables understanding when and to what extent a given feature will impact performance.

We demonstrated this strategy with two use cases. First, we compared false positive rates for Illumina PCR-free and PCR-plus sequencing technologies. Within homopolymers, INDEL false positives were much more prevalent in PCR-plus, and the error rate for both methods began to increase after 8 bp for both A/T and G/C homopolymers. SNP errors in contrast were not different between the two technologies; however, the error rate increased beyond 10 bp for A/T homopolymers and any length in the case of G/C homopolymers. For our second use case, we compared the likelihood of callsets from Illumina PCR-free or PacBio HiFi data to miss clinically relevant variants. While most false negatives occurred in hard-to-map regions, some errors occurred in homopolymer regions.

Ultimately, this will provide a data-driven mechanism for understanding sources of error within different variant caller methods and sequencing technologies within difficult genomic regions. We also plan to use this model to improve genome stratifications for creating and using GIAB benchmarks. The code is available online.

Session Title: Omics Technologies Poster Session I

PB3520 Using single cell transcriptomics to identify endothelial cell senescence signature genes that correlate with atherosclerotic disease.

Authors:

P. Hejazi[1,2], M. Mehdizadeh[1,3], É. Thorin[1,4], M. Sirois[1,5], S. Nattel[1,2,3,5], G. Lettre[1,2]; [1]Res. Ctr. of Montreal Heart Inst., Montreal, QC, Canada, [2]Dept. of Med., Université de Montréal, Montreal, QC, Canada, [3]Dept. of Pharmacology and Therapeutics, McGill Univ., Montreal, QC, Canada, [4]Dept. of Surgery, Université de Montréal, Montreal, QC, Canada, [5]Dept. of Pharmacology and Physiology, Faculty of Med., Université de Montréal, Montreal, QC, Canada

Background- Endothelial dysfunction is one of the hallmarks of atherosclerosis and results in arterial stiffness and the weakening of vascular endothelium integrity. Cellular senescence is one of the many consequences of endothelial dysfunction in which cells remain metabolically active but do not die or divide. The accessibility of the skin circulation system makes it an attractive tissue to do an in-depth analysis of endothelial function. We hypothesized that the skin vasculature can be used, in combination with single-cell transcriptomics, as an easily accessible biosample to monitor endothelial cell dysfunction and senescence and that this information can be used to monitor atherosclerosis. **Aims and Methodologies** - We aimed to investigate the impact of senescence on endothelial cell gene expression in the skin of mice. We used INK-ATTAC mice for these experiments, in which senescent cells can be selectively eliminated upon treatment with the drug AP20187. The mice received twice-a-week treatment throughout the period from 12 months to 18 months of age. We used three different groups of this mouse model: (1) young and untreated animals (3 months), (2) old and vehicle-treated animals (18 months), and (3) old and AP-20187-treated animals (18 months). Afterward, we collected and snap-froze skin samples from three mice in each group. We used 10X Genomics to construct sequencing libraries with a target of 10,000 cells/animal. We applied standard software such as CellRanger and Seurat for scRNA-seq for data analysis. **Results**- The validation of p16 gene expression using qPCR did not reveal significant differences among the three groups, prompting us to employ single-cell RNA sequencing for a more comprehensive examination. Single-cell RNA-seq analysis of 47,441 cells identified relevant cell clusters, including endothelial cells, but no differentially expressed genes (DEGs) were found across the groups. Pathway analysis on approximately 300 genes with nominally significant expression at p-value< 0.05 did not yield significant results. We speculate that the aged group (18 months old) may not have been old enough to exhibit a striking difference in gene expression related to endothelial cell senescence. Future studies should consider alternative approaches, such as investigating older mice, to better understand the senescence signature in the vascular system and atherosclerotic disease.

Session Title: Omics Technologies Poster Session II

PB3521 Using single-molecule analysis for a deep investigation of Tau proteoform heterogeneity and proteome-wide changes in Alzheimer's Disease.

Authors:

G. Kapp[1], S. Guha[1], S. Tan[1], J. Joly[1], D. Park[1], R. Wang[1], B. Nortman[1], T. Gillies[1], V. Budamagunta[1], J. K. Robinson[1], T. E. Rinker[1], J. Sherman[1], T. Wiederhold[2], D. Kirkpatrick[3], J. Lipka[4], T. Wendorff[4], N. J. Pandya[4], A. Rohou[4], J. Egertson[1], P. Mallick[1]; [1]Nautilus Biotechnology, San Carlos, CA, [2]Cell Signaling Technologies, Danvers, MA, [3]Interline Therapeutics, South San Francisco, CA, [4]Genentech, South San Francisco, CA

Introduction: The Tau protein is known to be extensively post-translationally modified at a large number of sites on the protein. The protein isoforms (proteoforms) that result from splicing and post-translational modification and the molecular heterogeneity of Tau are critical contributors to Alzheimer's disease (AD) progression and pathology. Currently, very little is known about the prevalence or impact of the diversity of Tau proteoforms. Here, we detected and analyzed Tau proteoforms in the context of the broader proteome in model systems using a novel single-molecule proteomic analysis platform.

Methods: We used a single-molecule proteomic analysis platform that leverages affinity reagents combined with novel instrumentation, single-molecule biochemistry, and machine learning bioinformatics to enable deep proteoform and broad proteome analysis. We first evaluated the proteoform distribution of Tau using commercially available antibody reagents. Proteoform heterogeneity was compared (specifically splicing and phosphorylation variants) in both healthy and diseased samples. We additionally applied Protein Identification by Short-epitope Mapping (PrISM), for broadscale proteomics analysis. PrISM leverages proprietary multi-affinity probes designed to recognize short epitopes and a machine learning algorithm that decodes binding of hundreds of multi-affinity probes into protein quantifications.

Results: The approach was evaluated by first measuring defined mixtures of recombinant Tau proteins. Next, we examined Tau enriched from induced pluripotent stem cell (iPSC)-derived neurons and tau-expressing cell lines. The platform revealed the molecular heterogeneity of Tau proteoforms missed by bulk measurements and peptide-centric proteomics approaches. The proteoform data was then supplemented with broader proteome analysis in order to identify proteome-wide changes - differences in proteins and pathways due to or in combination with specific Tau proteoforms and disease.

Conclusion: Understanding the role of proteoforms in biology uncovers a level of additional biological complexity that complements the rich genomic and proteomic data from existing analytical methods. Single-molecule tools that can analyze proteoforms in depth and the proteome broadly and link molecular signatures to disease states are essential. These tools will enable improved biomarkers, improved understanding of disease, and improved therapies for diseases like Alzheimer's disease.

Session Title: Omics Technologies Poster Session III

PB3522 Utilizing long-read sequencing to decipher the genomic architecture of mitochondrial DNA in neurodegenerative diseases

Authors:

N. Tran[1], A. I. Soto-Beasley[2], Y. Ren[1], S. Basu[2], W. Springer[2], O. A. Ross[2]; [1]Mayo Clinic Graduate Sch. of BioMed. Sci., Jacksonville, FL, [2]Mayo Clinic, Jacksonville, FL

Mitochondrial DNA (mtDNA) integrity is crucial in the overall health and physiological function of the cells. Previous studies have shown that mtDNA variations are implicated in several neurodegenerative diseases including Early-Onset Parkinson's Disease (EOPD). mtDNA is a circular 16.5 kb molecule and each mitochondrion has multiple copies. Given that there are several mitochondria per cell, there exist hundreds to tens of thousands copies of mtDNA. mtDNA variations are predominantly captured using Short-read whole-genome sequencing (SR-WGS) which allows the accurate and high-yield capture of single nucleotide variants (SNV); however, more complex structural variants (SV) such as repeats, inversions, large deletions, etc. are unable to be detected. The recent development of long-read sequencing (LRS) has allowed us to capture more complex SV and copy number variation (CNV), thus addressing the limitations of conventional SR-WGS. LRS will allow us to better characterize mtDNA variation influencing disease onset and progression. Furthermore, given PacBio LRS generates Circular Consensus Sequence with optimal read length of 15-17kb, it gives us an unprecedented opportunity to examine each copy of the 16.5 kb mtDNA genome in one read. This will allow for a true quantification of heteroplasmy levels and characterize individual mtDNA genomes for variation. In our current study, we have generated SR-WGS and PacBio LR-WGS for a subset of EOPD patients (age of onset <39). We utilized existing bioinformatics tools (GoldenHelix, PacBio, etc) to determine SNV and CNV for each dataset, and compared our findings between the two methods. Our study provides a comprehensive method that allows for the integration of sequencing approaches, capture of mtDNA CNV/SV with high resolution and in-depth quantification of the mitochondrial genome species to resolve the role of mtDNA variation in the EOPD phenotypic presentation including disease onset and progression.

Session Title: Omics Technologies Poster Session I

PB3523 VizCNV: An integrated platform for CNV detection and analysis of genome sequencing data

Authors:

H. Du[1], M. Lun[2], M. Mehaffey[2], C. M. Grochowski[1], M. Gandhi[3], S. N. Jhangiani[4], D. M. Muzny[4], D. Pehlivan[1], F. Sedlazeck[4], Z. Coban-Akdemir[5], J. Posey[1], A. Lindstrand[6], I. K. Chinn[1], R. Gibbs[4], J. Lupski[1], C. M. B. Carvalho[2]; [1]Baylor Coll. of Med., Houston, TX, [2]Pacific Northwest Res. Inst., Seattle, WA, [3]UT MD Anderson Cancer Ctr., Houston, TX, [4]Baylor Coll. Med., Houston, TX, [5]The Univ. of Texas Hlth.Sci. Ctr. at Houston, Houston, TX, [6]Karolinska Inst.t, Stockholm, Sweden

Copy number variation (CNV) is a class of structural variation that results in deviations from the diploid state and contributes to the pathogenicity of genomic disorders if dosage-sensitive genes are involved. Karyotyping was initially used to detect large CNVs (>5 Mb); however, in the past ten years, comparative genomic hybridization (aCGH) or SNP arrays, which have a resolution of 5-10 kb, have emerged. With the advancement of genome sequencing (GS) in the clinical diagnostic setting, CNV analysis using GS data has faced several challenges, including the overwhelming number of false-positive calls due to systemic biases from non-uniform read coverage and collapsed calls resulting from the abundance of paralogous segments in the human genome. To address these challenges, we developed VizCNV to facilitate CNV analysis using GS data. VizCNV normalizes read depth using the chromosomal median read depth information and leverages the shifting level model for segmentation. Using R Shiny, VizCNV generates aCGH and SNP array-like interactive graphs to facilitate the detection, visualization, and interpretation of CNVs at a resolution of 1-10 kb. Moreover, this platform incorporates tools to distinguish inherited and de novo CNVs, both of which are validated by B-allele frequency plots. VizCNV provides a genomewide view enabling pattern recognition of chromosomal abnormalities [e.g., aneuploidy, deletions, duplications] and mutational events such as chromothripsis/chromoanasynthesis, multiple de novo CNVs, and complex genomic rearrangements. To test the effectiveness of VizCNV, we analyzed 19 aCGH-positive trios using Illumina short-read GS (30x) data. This dataset includes 16 probands with MECP2 duplication syndrome and three probands with de novo marker chromosomes. VizCNV accurately identified nine dnCNVs and 10 inherited variations, including a triplication and a mosaic CNV. In contrast, combined output from three structural variant callers (Manta, Delly, and Lumpy) identified only 9 out of 19 aCGH-observed events. To further test the tool, VizCNV was utilized to screen for potentially pathogenic CNVs in 39 patients with a primary diagnosis of primary immunodeficiency disease. On average, 7 potential dnCNVs and 90 inherited CNVs >10 kb per trio passed the filter. In one family, a compound heterozygous combination of a paternal 12.8 kb exonic deletion and a maternal missense variant affecting DOCK8 explained the proband's phenotype. In conclusion, VizCNV provides a reliable method for CNV analysis at a higher resolution than traditional aCGH analysis. VizCNV is freely available online.

Session Title: Omics Technologies Poster Session II

PB3524 VPAP: A pipeline for proteomic analysis of genetic variants

Authors:

V. Magalhaes Borges[1], N. Teruel[2], A. Nato[1]; [1]Dept. of BioMed. Sci., Joan C. Edwards Sch. of Med., Marshall Univ., Huntington, WV, [2]Dept. of Pharmacology and Physiology, Faculty of Med., Université de Montréal, Montreal, QC H3T 1J4, Canada, Montreal, QC, Canada

Understanding the three-dimensional structure of proteins plays a pivotal role in unraveling the functional implications of genetic variants. Recent advancements in structural biology, led by AlphaFold, revolutionized our ability to explore the universe of protein structures and allowed us to integrate and complementarily connect genomics and proteomics. This work introduces a variant proteomic analysis pipeline (VPAP) that effectively evaluates the proteomic consequences of disease-associated genetic variants. The development of VPAP is driven by two conceptual motivations. First, the evaluation of the allele-specific impact of a nonsynonymous variant on the protein structure allows us to predict its potential effects on protein stability, folding, and interaction with other molecules. Second, the degree of interactions across residues flanking an amino acid mutation (e.g., protein-protein interaction, protein-ligand interaction) is useful in interpreting genomic data to identify disease-causing variants. Implemented in R, VPAP simplifies access to publicly available datasets (e.g., GWAS Catalog, Protein Data Bank) and incorporates a systematic approach to select the top genetic ancestry-specific variants associated with a specific complex disease. By integrating scriptable methods that are suited to high-throughput evaluations, VPAP identifies and scores the interactions between the wild-type protein (containing the reference allele) or the mutated protein (containing the alternate allele) with a ligand or another protein, employing (1) MODELLER, a software that generates the modified protein structures based on homology modeling and (2) Surfaces, a tool to quantify per-residue contributions to molecular interactions within and between proteins and ligands. Based on the interaction values calculated by Surfaces, VPAP estimates allele-specific impact scores, which may be used in downstream analyses. VPAP compares these impact scores with values from several variant annotations and prioritizers to evaluate and validate its robustness. VPAP therefore allows researchers to measure the magnitude of effect caused by a specific risk allele for a particular disease and visualize the molecular interactions occurring within these residues. This multi-omic approach leverages on the strengths of each field to elucidate the underlying mechanisms involved in the etiology of a disease and provide valuable information to explore therapeutical targets and ultimately guide drug design.

Session Title: Omics Technologies Poster Session III

PB3525 Whole-genome sequencing and metabolic profiling in Finns underscores the relevance of rare variants in blood metabolome regulation and disease risk.

Authors:

X. Yin[1], B. Ryan[2], J. Okamoto[2], J. Chiou[3], D. Bose[2], J. LI[2], A. Jackson[2], H. Stringham[2], S. Ripatti[4], A. Palotie[4], J. Morrison[2], X. Wen[5], C. Burant[2], E. Fauman[3], M. Laakso[6], M. Boehnke[2]; [1]Nanjing Med. Univ., Nanjing, China, [2]Univ. of Michigan, Ann Arbor, MI, [3]Pfizer, Cambridge, MA, [4]Inst. for Molecular Med. Finland FIMM, Helsinki, Finland, [5]Univ. of Michigan, Ann Arbor, MI, [6]Univ. of Eastern Finland and Kuopio Univ. Hosp., Kuopio, Finland

Metabolite levels reflect both environmental and genetic influences. GWAS have identified thousands of common (minor allele frequency [MAF]≥1%) genetic variants associated with metabolite levels. However, the contribution of rare genetic variation to metabolite levels is less studied. We profiled plasma levels of 1,540 metabolites in 10,188 Finnish men from the METSIM study using Metabolon HD4 untargeted mass spectrometry platform and sequenced genomes for 9,501 of the 10,188 individuals (average depth=25x). We quantified the contribution of common and rare (MAF<1%) variants to the single nucleotide variation (SNV)-based heritability (h2SNV) of blood metabolites. We found that common and rare variants accounted for on average 16.6% (median=14.1%, range=0.09-65.9%) and 9.6% (median=8.3%, range=0.17-77.5%) of the h2SNV for blood metabolites, respectively, with most of the rare variant heritability attributed to the variants with 0.1%<MAF<1%. We performed GWAS for each of the metabolites, identifying 3,890 independent association signals (metabQTLs) for 1,011 metabolites at P<5.0x10-9. To prioritize putative causal variants for each metabQTL, we conducted statistical fine-mapping analysis and identified signal-level posterior probability≥0.95 for 3,666 of the 3,890 metabQTLs. For 370 (9.5%) metabQTLs for 282 (27.9%) unique metabolites, we identified putative causal variants at 263 rare variants with variant-level posterior probability (VPIP)≥0.3. Notably, of the 263 variants with VPIP≥0.3, 193 showed MAF>10-fold greater in METSIM than in non-Finnish Europeans (NFE) for 355 metabQTLs of 291 metabolites. We performed colocalization analysis between the metabQTLs from the METSIM and GWAS for the 1,209 disease traits from the FinnGen release 9. We identified 6,561 metabolite-trait pairs having at least one locus with locus-level colocalization posterior probability (LCP) > 0.5, suggesting a pervasive role for blood metabolites in disease molecular mechanisms. For example, we suggested that Finns-enriched missense variant rs77273740 of the *DBH* gene affects the risk of type 2 diabetes through regulating plasma 3-methoxytyramine sulfate levels. We are integrating metabolites and disease GWAS with several other molecular traits including gene and protein expression to help understand disease molecular mechanisms.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4001 3D-4-BigData: 3D Printed Electronics enabled 4D Visualization of Big Data Networks in Human Genetics

Authors:

J. Kim[1], G. Manogharan[2], Y. Veturi[3,4]; [1]Schreyers Honors Coll., Pennsylvania State Univ., University Park, PA, [2]Dept. of Mechanical Engineering, Pennsylvania State Univ., University Park, PA, [3]Dept. of Biobehavioral Hlth., Pennsylvania State Univ., University Park, PA, [4]Dept. of Statistics, Pennsylvania State Univ., University Park, PA

Background: Current statistical/computational advances in network analysis lack the ability to understand the interconnectedness between network nodes in a tactile manner. Effective visualization of network graphs can not only identify underlying structure in the data but are also critical for dimension reduction and training. **Objectives:** The overall research goal is to design-develop-demonstrate-evaluate a novel data visualization method that brings together data analytics, augmented reality (AR), and 3D Printed Electronics (3DPE - embedded wireless electronics at nodes and LED strips at edges). This novel approach is demonstrated for genetic associations between neuroimaging phenotypes and diseases in electronic health records (EHR) across the phenome. Our approach can yield novel image-derived phenotype (IDP) biomarkers for complex diseases as well as provide pleiotropic gene targets for drug repurposing. **Methods:** We have established: (1) an open-source Python toolkit to drive the 3DPE-node and data network using open-source low-cost controllers (e.g., Arduino, Teensy 4.1), and (2) visualization of node-edges in mixed reality through AR lens modalities with physical model (e.g., Unity, Meta Quest Pro) to visualize 3D network models. We apply this to summary statistics from a transcriptome-wide association study on 2,124 IDPs (structural and diffusion MRI measuring cortical volume, cortical thickness, brain volumes, cortical grey-white contrast etc.) from UK Biobank (UKB) on 38K samples of European ancestry across 13 brain tissues from GTEx v8 followed by gene-based colocalization. We mapped the significant genes to fine-mapped eQTL from PredictDB (using multivariate adaptive shrinkage models). Finally, we conducted IDP-guided phenome-wide association study on 12,494 fine-mapped eQTL across 664 ICD-10 codes (n=452,595 samples). **Results:** We found that 3D-4-Big Data can seamlessly create mixed-reality 3D network models for GWAS and TWAS studies with preprocessed high dimensional data and is effective in mapping network models with multiple modalities (e.g., labeling of genes, drug pathway using AR, and dynamic networking - edge direction, LED intensity to identify strength of relationship). **Conclusions:** We have developed a novel immersive 3DPE enabled 4D visualization of big data networks to better understand the complexity of the high-density data networks in human genetics. This study demonstrates the utility of this approach as a powerful tool for interactively visualizing pre-processed high dimensional multilayered data (e.g., genomics, transcriptomics, EHR) for statistics education.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4002 A BAYESIAN NETWORK-BASED APPROACH FOR MULTI-OMICS INTEGRATION TO REVEAL UNDERLYING MECHANISMS OF HEALTHY AGING

Authors:

A. Leshchyk[1], S. Monti[1], P. Sebastiani[2]; [1]Boston Univ., Boston, MA, [2]Tufts Med. Ctr., Boston, MA

Previous studies of individuals who live to very old age have found that very old centenarians experience a significant delay in the onset of age-related diseases and disabilities such as Alzheimer's, dementia, and cardiovascular diseases compared to the general population. Genetic studies of long-lived individuals showed that carriers of the APOE e2 allele had increased odds of reaching longevity compared to the non-e2 allele carriers. Moreover, the APOE e2 allele is characterized by distinct serum proteomics and metabolomics profiles that could be useful to understand the mechanism of propagation of the genetic effect of APOE to the molecular level and eventually to phenotypes. We are developing a novel Bayesian network-based approach that integrates genetics, multi-omics, and multi-phenotypes to identify shared molecular profiles among the subjects with familial longevity that lead to compression of morbidity, disability, and mortality. This Bayesian network-based approach can be used to also make various predictions and probabilistic reasoning. For example, the preliminary analysis shows that the APOE e2 allele carriers have lower sphingolipids abundance than APOE e4 carriers and better cognitive performance (for SM:18 $-1.5<FC<1.5$ in APOE e2 carriers $P = 0.32$ for a high cognitive score in contrast to APOE e4 carriers with $P= 0.08$). Sphingolipids are fatty amino alcohols that regulate cell survival, inflammation, senescence, and apoptosis, which might impair the cognitive function of elderly adults and contribute to their disability and morbidity, thus affecting their aging.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4003 A cloud based pipeline for rare variant analysis using public summary counts for prioritizing disease predisposition genes in rare disease studies.

Authors:

S. Tithi, W. Chen, G. Wu; St. Jude Children's Res. Hosp., Memphis, TN

Sequencing data from rare disease cohorts often lack healthy matched individuals to serve as controls for association analysis to prioritize disease risk genes. In such cases, researchers often use public datasets as controls, but these public datasets are often assembled from different sources, using different pipelines, making it difficult to combine the control data with the disease cohort directly for genetic association analysis. Previously, we developed CoCoRV (consistent summary counts based rare variant burden test), which is a computational tool to help prioritize disease predisposition genes and genetic variants in rare diseases using public summary counts. CoCoRV implements consistent variant quality control and filtering, ethnicity-stratified rare variant association test, accurate inflation factor estimation and powerful FDR control, which minimize the chance of false positives due to batch effects between cases and controls. However, the full pipeline that applies CoCoRV involves multiple steps such as variant QC, annotation, and ethnicity inference, which can be either computationally intensive or demanding a large physical memory, such as preprocessing of the gnomAD control data. Here, we develop a cloud version of CoCoRV based on the popular workflow language nextflow and the docker container. Our cloud based pipeline includes already pre-processed and annotated gnomAD count data, and it is simple to run the full analyses once required parameters are specified. The nextflow based pipeline allows users to resume the execution from the last successfully executed step, so users do not need to re-run every step each time. The docker containerization of each step relieves the users from installing all dependencies for their specific platform, making it much easier to use the tool. The pipeline is deployed in a cloud platform called CAVATICA through a partnership with Seven Bridges. As CAVATICA hosts a large amount of sequencing data, including pediatric cancer data from Kids First Portal, researchers around the world can analyze these datasets using our tool to make potential discoveries on novel disease predisposition genes. The nextflow based pipeline is also portable and can be downloaded to run in a local environment, e.g., a high-performance computing cluster.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4004 A complete end-to-end workflow for implementing clinical polygenic risk scores

Authors:

G. Busby, S. Kulm, A. Bolli, J. Kintzle, P. Di Domenico, G. Bottà; Allelica, New York, NY

Polygenic Risk Scores (PRS) aggregate the effect of many variants into a single value that represents an individual's genetic liability of disease from common genome-wide variation. Over the last 5 years, interest in the clinical use of PRSs for a range of common diseases has grown, but challenges to widespread adoption and implementation remain. These include but are not limited to the attenuation of performance of the same PRS in different populations, the differential performance of different PRSs for the same disease in the same population, misunderstandings on what the output of a PRS actually represents and how it should be used, and uncertainty around how to standardize and regulate PRS tests.

Here we describe the development of a transparent and fully automated, genetic ancestry-aware bioinformatics pipeline that takes raw genotype data and turns it into a clinical assessment of an individual's disease risk via a PRS. Our pipeline is agnostic to the genetic technology used to generate the genetic data, and we discuss the need for essential quality control, genetic ancestry adjustment, clear definitions of reference risk distributions that are required to deploy PRSs, and how to convert and communicate a PRS into a measurement of disease risk. We also provide clarity on how the choice of PRS and reference distribution should be a key consideration when applying a PRS to a new population and how to match existing datasets to the intended use population. We provide metrics that can be used to objectively quantify potential discrepancies between different PRSs in order to transparently assess the application of a PRS to specific intended-use populations.

We illustrate this approach using cardiovascular disease, type 2 diabetes, breast and prostate cancer as examples. By documenting the end-to-end process of PRS generation, our study shows the key considerations necessary to generate clinical level risk predictions from genomic data and provides a framework for the reproducibility of PRSs in the clinic.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4005 A continuous local ancestry measure and efficient local-ancestry-aware association tests

Authors:

H. Sun[1], B. Yu[1], Q. Qi[2], A. Morrison[3], E. Boerwinkle[1], R. Kaplan[2], H. Chen[4]; [1]Univ. of Texas Hlth.Sci. Ctr. at Houston, Houston, TX, [2]Albert Einstein Coll. of Med., Bronx, NY, [3]Univ. of Texas Hlth.Sci. Cntr Houston, Houston, TX, [4]The Univ of Texas Hlth.Sci. Ctr. at Houston, Houston, TX

Population stratification is a well-known challenge in genetic association tests, due to the genetic heterogeneity between averaged ancestral proportions (global ancestry) and ancestry at specific regions (local ancestry), particularly for admixed populations. However, conventional approaches, like top ancestral principal components (PCs), may fail to correct the underlying population structure. Moreover, relying on reference panels, local ancestry inference is prone to misclassification with improper reference populations involved. To address the issues, we proposed Local Ancestry Coordinates (LACs), a continuous measurement aggregating population-specific local ancestry components into linear combinations. LACs can preserve local ancestry characteristics in fine-scale resolution, and is more robust against misclassification. We projected local ancestry probabilities onto a reference-based global PC-spanned space, to represent the demographic and geographical patterns of local ancestry across the genome. Specifically, we first established a standard reference panel comprising 3 African (ASW, LWK, YRI), 3 European (CEU, TSI, GBR), 3 American (PEL, CLM, MXL) populations from the 1000 Genomes project, which provided a constant projection canvas regarding their global PCs. We then simulated 5,000 independent 9-ancestry-admixed study samples and called their reference-dependent local ancestry probabilities using RFMix v1.5.4 with reference panels from 6, 9, or 12 ancestral populations. Study samples were then projected onto the same reference-based global PC-spanned space to compute their global PCs and LACs, using their global ancestry proportions and local ancestry probabilities, respectively. To evaluate LACs' efficacy and robustness, we computed the correlations between LACs inferred using various reference panels and the simulated true values. The mean correlations for the top 2 PCs across the genome for 6, 9, and 12-reference-population inferred LACs are 0.9489 and 0.9123, 0.9891 and 0.9816, 0.9879 and 0.9808, respectively. In our simulations, the LAC-adjusted association test effectively controlled type I error rates confounded by local ancestry, outperforming global PCs-adjusted tests. We also applied our local-ancestry-aware association test to real data from the Hispanic Community Health Study/Study of Latinos (HCHS/SOL). Our findings demonstrate LACs as a reliable proxy for characterizing local ancestry and mitigating the impact of local ancestry misclassification, and incorporating LACs into association tests shows promise in effectively accounting for population stratification effects.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4006 A framework to evaluate and validate polygenic risk models at scale.

Authors:

S. Moreno Grau[1,2], M. Vernekar[3], A. Lopez-Pineda[1,4], M. A. Nava-Aguilar[1,4], C. Quinto-Cortés[1], B. Moatamed[1], M. Lee[1], Z. Yu[3], K. Numakura[3], Y. Matsuda[3], J. D. Wall[1], A. Ioannidis[1,2], N. Katsanis[1], T. Takano[3], C. Bustamante[1,2]; [1]Galatea Bio, Hialeah, FL, [2]Dept. of BioMed. Data Sci., Stanford Univ. Sch. of Med., Stanford, CA, [3]Genomelink Inc, Berkeley, CA, [4]Amphora Hlth., Morelia, Mexico

Polygenic risk scores (PRS) are a promising approach to predict an individual's genetic risk to diseases. Thousands of PRS have been developed, but each new PRS is highly influenced by the characteristics of the discovery dataset. There is limited validation outside the training sets. A lack of standardized approaches to compare and select the best available PRS exists. To address this problem, we present a framework for harmonizing, testing, selecting and validating the best performing pre-existing PRS for 13 phenotypes of medical interest.

A harmonized set of unrelated European individuals from the UK biobank was used to test PRS models. Then, we validated the best performing PRS in an independent cohort of Europeans from the Biobank of the Americas-Genomelink (BBofA-GL) study (N=8,394), and tested them in non-European populations (N UKBB =18,846). Case counts ranged from 216 for T1D to 33,441 for high HDL levels in the UKBB, and from 151 for T1D to 2,283 for hypertension in the BBofA-GL. Phenotypes were extracted from UKBB phecodes (ICD10) and by self-reported questionnaires in the BBofA-GL. Genetic information was processed following standard methods. To select the best performing PRS model we estimated the variance explained by the PRS, which was calculated as the difference between AUC for the model with PRS and covariates (age, sex, 4PCs) versus the covariates-only model. Odds ratio for each outcome was estimated per percentile.

We evaluate a total of 401 PRSs downloaded from PGSCatalog for cardiac conditions (atrial fibrillation [AF], CAD, hypertension), metabolic (T1D, T2D, hypothyroidism), gastric (celiac disease [CD], ulcerative colitis [UC], hemorrhoids), inflammatory (gout), and lab values (LDL, HDL and TG measures). Our framework lets us replicate published PRS effectively. Selected PRS explained from 1.4% of disease variance for CAD to 25% for CD. The 90th percentile threshold showed 10 times more cases than the observed in the mid-percentile (40-60%) for T1D. Large case enrichment was also observed for CE (x3), CAD and TG levels (x2), conversely to other outcomes, where case enrichment was moderate (gout, T2D, hypertension, hypothyroidism and LDL) or limited (AF, UC, hemorrhoids, HDL, and VitaminD). We observed variation in the PRS distribution by ancestry, and a reduction in PRS accuracy from European sets to South, East asians and Africans.

Here we have developed a scalable framework to test, select and implement the best performing PRS, maximizing the comprehension of existing scores. We have validated PRSs in an independent cohort and highlighted conditions, where implementing risk-prioritization strategies could have a practical utility

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4007 † A framework to evaluate the utility of embeddings in genetic discovery

Authors:

S. Mukherjee[1], Z. McCaw[1], J. Pei[1], H. Somineni[1], T. Soare[1], B. Michael[2], C. O'Dushlaine[1], C. Probert[1], T. Karaletsos[1]; [1]Insitro, South San Francisco, CA, [2]Stanford Univ., Palo Alto, CA

Embeddings are a compressed representation of high content data modalities obtained through deep learning models. Embeddings have been thought to capture detailed information about disease states and have been qualitatively shown to be useful in genetic discovery. Despite their promise, embeddings are also known to have several downsides: i) they are often confounded by covariates, and ii) their disease relevance is hard to ascertain. In this work we aim to develop a framework to systematically evaluate the utility of embeddings in genetic discovery.
Our genetic validation workflow focuses on comparing phenotypes along two axes: i) heritability of the phenotypes, and ii) ability to identify 'disease relevant' variants. We use the number of genome-wide significant signals and mean/median chi-square statistic as a proxy for heritability of multivariate phenotypes. To evaluate disease relevance, we compute polygenic risk scores for each principal component (PC) of the embedding (or multivariate comparators) and evaluate their association with a held-out set of patients with high-confidence disease phenotypes. We use NAFLD as a case study and compare several embedding/non-embedding derived comparators using our framework: i) embeddings from a deep learning model trained to predict liver fat percentage from MRIs in UK Biobank, ii) 203 NAFLD-related traits comprising of circulating metabolites, blood biochemistry markers and abdominal composition variables, and iii) imputed NAFLD case/control status from biomarkers using machine learning. To keep our comparisons consistent, we perform the GWAS for each trait on the same set of samples (discovery cohort) and evaluate in a held-out set of samples with NAFLD case/control labels (test cohort).
We find that while embeddings had more genome-wide significant associations (442) than predicted NAFLD case/control status (6), a multi-trait analysis across all of the NAFLD-related traits had the greatest number of associations (4968). However, embedding-derived PRSs (p-value = 8.9e-18) show a much more significant association with gold-standard NAFLD case/control status than PRSs from non-embedding baselines (p-value for PCs of traits = 0.75, p-value for predicted case/control status = 0.15).
We propose a framework to evaluate the utility of embeddings in genetic discovery along two important axes. Using a case study and reasonable non-embedding baselines, we evaluate the utility of embeddings for genetic discovery in NAFLD. Our finding shows that embeddings derived from a model trained to predict liver fat percentage yields highly disease relevant associations when compared against our chosen baselines.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4008 A framework to improve the interpretation and prediction of variant effect sizes using non-linear functional models

Authors:

S. Cheng, S. Gazal; USC, Los Angeles, CA

Modeling the relationship between the effect sizes of common variants and their functional annotations is critical to accurately characterize the functional genetic architecture of human diseases and complex traits. Existing functional models rely on linear relationships between variants squared effect sizes (or per-variant heritability) and their functional annotations; this linear relationship is methodologically constrained by the fact that we are observing marginal effect sizes from genome-wide association studies (GWAS), and that true effect sizes are unobserved. However, linear models constrained our interpretation of functional architectures, and it is thus unclear if a non-linear structure would improve model prediction. Here, we aim to improve the interpretation and prediction of existing functional models by using machine learning approaches.Here, we developed a framework that leverage fine-mapping methods to 1) estimate the true allele effect sizes of each common variant, and 2) model per-variant heritability directly on these estimated effects (rather than marginal effects) using 96 functional annotations from the baseline-LD model (Gazal et al. 2017 Nat Genet). Using simulations, we observed that leveraging the fine-mapping method SuSiE with functional priors (estimated using an additive model of annotations) provides nearly unbiased functional enrichment estimates, even if the causal model includes interactions between annotations. We investigated decision trees and deep neural networks (DNN) on the fine-mapping results of 15 independent UK Biobank traits. Optimal hyper-parameters were estimated by looking at the mean squared error (MSE) using a leave-one-chromosome-out (LOCO) approach (to avoid overfitting). For decision trees, we observed that the functional genetic architecture of complex traits were dominated by annotations related to evolution (8/18 nodes) and continuous annotations (10/18 nodes) where we were able to optimize cutoff. The tree also highlighted constrained enhancers and recent constrained mutations as the most important leafs. By analyzing new annotations based on the 19 leaves of this tree with S-LDSC on 63 independent GWAS, we observed that 13 annotations had significant conditional effect over the existing annotations, demonstrating an improvement of model fit. Finally, we observed that DNN produced lower MSE than the linear model (used by S-LDSC), demonstrating the benefits of DNN annotations for future genetic analyses.In conclusion, we have developed and validated a framework using non-linear models that improve representation and prediction of existing heritability models.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4009 † A Functionally Informed Rare Variant Analysis Approach for Unbalanced Case-Control Ratios in Biobank-Scale Sequencing Data

Authors:

Z. Li[1], X. Lin[2], X. Li[3]; [1]Northeast Normal Univ., Changchun, China, [2]Harvard T.H. Chan Sch Pub Hlth, Boston, MA, [3]Univ. of North Carolina at Chapel Hill, Chapel Hill, NC

Introduction
Large-scale sequencing studies have enabled the investigation of rare variants (RVs) associated with complex human diseases and traits. Functionally informed RV analysis represents a powerful approach for studying RV associations through the integrative analysis of sequencing data and functional annotations that predict the biological functionality of genetic variants. The existing STAAR method and its corresponding whole-genome sequencing analysis workflow, STAARpipeline, provide robust, scalable, and powerful functionally informed association analysis for quantitative and binary traits in biobank-scale sequencing data. However, these approaches have limited capacity to account for extremely unbalanced case-control ratios, such as <1:100, for binary traits.
Methods
We propose STAAR-B-SPA, an extension of the STAAR framework using the Burden test, to improve type I error control for significantly unbalanced case-control ratios. We integrated STAAR-B-SPA into the STAARpipeline, enabling scalable, flexible, and streamlined functionally informed genetic association analysis for biobank-scale sequencing data, including gene-centric analysis and genetic-region analysis for coding and noncoding variants. STAARpipeline implemented with STAAR-B-SPA is available in both offline and cloud computing environments, such as high-performance computing clusters, the UK Biobank research analysis platform and the NHLBI BioData Catalyst ecosystem. Through extensive simulation studies, we demonstrated that STAAR-B-SPA can efficiently analyze large datasets while maintaining well-controlled type I error rates.
Results
We applied STAAR-B-SPA to detect rare variants associated with five cancer traits (breast, colorectal, lung, prostate and ovarian) using the UK Biobank 200K whole-exome sequencing data. The control to case ration varies from 20 to 200. We detected 15 genome-wide significant associations while well protecting the type I error rate in the analysis of these five cancer traits. Notably, the significant findings include not only known genetic associations, such as protein-truncating variants in *BRCA2* with breast, prostate and ovarian cancer, but also potential new associations, such as synonymous variants in *IQCJ*, *PRSS16*, and *CACHD1* with lung cancer.
Summary
In summary, STAAR-B-SPA provides a comprehensive tool for studying coding and noncoding rare-variant associations in biobank-scale sequencing data with extremely unbalanced case-control ratios.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4010 A genetic risk profile classifies disease severity in children with sickle cell anemia from birth to six years of age.

Authors:

S. Rashkin[1], J. Ding[2], J. H. Estepp[1,3], E. Rampersaud[4], L. E. Palmer[1], M. McCarville[5], S. Hwang[5], J. Hodges[1], M. Barton[1], M. P. Smeltzer[6], A. Olanrewaju[7], W. C. Wang[1], J. S. Hankins[1,3], G. Wu[4], M. R. DeBaun[8,9], G. Kang[2], M. J. Weiss[1]; [1]Dept. of Hematology, St. Jude Children's Res. Hosp., Memphis, TN, [2]Dept. of Biostatistics, St. Jude Children's Res. Hosp., Memphis, TN, [3]Dept. of Global Pediatric Med., St. Jude Children's Res. Hosp., Memphis, TN, [4]Ctr. for Applied Bioinformatics, St. Jude Children's Res. Hosp., Memphis, TN, [5]Dept. of Diagnostic Imaging, St. Jude Children's Res. Hosp., Memphis, TN, [6]Div. of Epidemiology, Biostatistics, and Environmental Hlth., Sch. of Publ. Hlth., Univ. of Memphis, Memphis, TN, [7]Dept. of Oncology, St. Jude Children's Res. Hosp., Memphis, TN, [8]Div. of Hematology/Oncology, Dept. of Internal Med., Vanderbilt Univ. Med. Ctr., Nashville, TN, [9]Dept. of Pediatrics, Vanderbilt-Meharry Ctr. for Excellence in Sickle Cell Disease, Vanderbilt Univ. Med. Ctr., Nashville, TN

Sickle cell anemia (SCA) is an autosomal recessive disease caused by one major missense mutation in the *HBB* gene, which encodes the β-globin subunit of adult hemoglobin (HbA, $\alpha 2\beta 2$). Abnormal sickle hemoglobin (HbS, $\alpha 2\beta^S 2$) polymerizes at low oxygen tension, causing red blood cells (RBCs) to become sickle shaped and triggering a complex pathophysiology with vascular occlusion, inflammation, and hemolysis. Symptoms begin in early childhood with acute pain episodes and progressive multi-organ damage leading to early mortality. The onset and progression of organ damage in SCD vary according to genetic and environmental risk factors. Variants impacting HbS polymerization in RBCs, such as α-thalassemia and predictors of post-natal fetal hemoglobin (HbF) levels, are associated with many SCA-related outcomes. Hence, we developed a genetic risk profile (GRP) including eleven single nucleotide polymorphisms (SNPs) from *BCL11A*, *HBS1L-MYB*, and the extended β-like globin locus, which account for ~20-50% of HbF variance in SCD patients, and the most common α-thalassemia allele ($\alpha^{-3.7}$). Using 327 children with SCA (HbSS/HbSβ^0; 0-6 years old; 1,952 person-years of observation) enrolled in the Sickle Cell Clinical Research and Intervention Program (SCCRIP), we evaluated the effects of $\alpha^{-3.7}$ and an 11-SNP polygenic score for HbF (PGS^{HbF} = # HbF-increasing alleles) on six hematological traits (HbF, hemoglobin [Hb], mean corpuscular volume [MCV], reticulocyte [ARC], white blood cell [WBC], and neutrophil [ANC] counts) and four clinical events (elevated transcranial doppler velocity [TCD] or stroke, acute chest syndrome [ACS], single RBC transfusion [TXN], and acute splenic sequestration [ASS]). For all outcomes where $\alpha^{-3.7}$ or PGS^{HbF} were associated individually, the effects were independent, so we constructed a GRP with 3 risk categories: high ($\alpha\alpha/\alpha\alpha$ & $PGS^{HbF} \leq 7$; N=110/327), low ($\alpha\alpha^{-3.7}/\alpha\alpha$ & $PGS^{HbF} > 7$ or $\alpha\alpha^{-3.7}/\alpha\alpha^{-3.7}$; N=53/327), and intermediate (otherwise; N=164/327).

Compared to the intermediate and high groups, at a false discovery rate adjusted P (P_{FDR})<0.05, those with low-risk profiles had higher HbF and Hb and lower MCV, ARC, WBC, and ANC, all of which predict reduced disease severity. The high-risk group had increased rates of elevated TCD or stroke (β=1.50; $P_{FDR}=2x10^{-5}$), ACS (β=0.31; P_{FDR}=0.04), TXN (β=0.74; $P_{FDR}=3x10^{-4}$), and ASS (β=0.54; P_{FDR}=0.04). Results for HbF, Hb, MCV, WBC, and ACS were validated in the Silent Infarct Transfusion (SIT) cohort. While the GRP may not always be as strongly associated as the $\alpha^{-3.7}$ or PGS^{HbF} alone, it was associated with all outcomes, so it can be used to stratify infants by risk of severe disease agnostic of specific outcome.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4011 A joint model for transcriptome-wide association study analysis to dissect gene and pathway level contribution to complex polygenic traits

Authors:

N. Wang, T. Ma; Univ. of maryland, COLLEGE PARK, MD

Genome-wide association studies (GWAS) have identified tens of thousands of single nucleotide polymorphisms (SNPs) associated with human disease or complex traits in the past 10 years, however, their biological mechanisms remain largely unknown. Transcriptome-wide association studies (TWAS) were proposed as a tool for detecting associations at gene levels and potentially investigating the gene regulatory mechanisms. Existing TWAS methods are usually implemented in two stages, which first used the reference dataset (e.g. genotype and gene expression data from GTEx) and genotype data of the study to impute the gene expression data and then correlated the imputed expression data with the phenotype in the study. However, these methods failed to account for the uncertainty inherent to the imputing step and may lead to a potential loss in statistical power. In addition, all existing TWAS methods are strictly univariate without considering the gene-gene correlation and groups of genes (e.g. biological pathways) that might be jointly predictive of complex polygenic traits. In this paper, we proposed a joint model that accounts for the uncertainty in gene expression imputation and dissects genetic contribution to traits at both gene and pathway levels. A sparse group Lasso penalty was introduced into a penalized expectation-maximization (EM) framework to select both genes and pathways associated with a trait. Our method can be regarded as an polygenic extension of an existing joint model, collaborative mixed model (CoMM). Our extension elevates the analysis to a multivariate framework, enhancing its capacity to handle complex genetic data and providing better functional insight into the genetic architecture of traits. We conducted extensive simulations with varying group structures and signal strengths to evaluate the variable selection and prediction performance of our method as compared to two-stage TWAS, univariate and additive CoMM. We also applied our method to identify genes and pathways predictive of a polygenic trait blood pressure (BP) using UK Biobank data.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4012 A map of genetic and phenotypic associations across male reproductive phenotypes.

Authors:

J. Figueredo, Estonian Biobank research team, T. Laisk, **R. Mägi**; Inst. of Genomics, Univ. of Tartu, Tartu, Estonia

Male infertility is a common, complex disease, affecting ~7-10% of men, and manifesting in diverse phenotypes ranging from morphological and functional abnormalities in sperm to severe spermatogenic impairment. Epidemiological data indicates that male infertility is often not an isolated condition, and it is not only a concern related to failed fatherhood. Recent research has revealed that many of the molecular pathways and mechanisms involved in male reproductive traits are shared with chronic diseases such as diabetes, heart disease, and cancer. Therefore, understanding the underlying causes of male infertility and finding ways to improve overall male reproductive health can advance reproductive health outcomes and help identify and prevent other health issues. Here, we carry out the most extensive effort to map the genetic background of phenotypes related to male reproductive health and provide an atlas of genetic and phenotypic correlations. In a joint analysis of three large population-based biobanks [Estonian Biobank (EstBB), FinnGen and UK Biobank (UKBB)] we perform a genome-wide association study (GWAS) meta-analysis of up to 532,376 men across 53 phenotypes defined using International Classification of Disease 10 (ICD-10) classification and encompassing diagnoses related to cancers of the reproductive tract and diseases of the genitourinary system. In addition to ICD-10 codes, 12 EstBB lab measurements were used in the analyses. Besides annotating the GWAS findings from individual phenotypes, we estimated the heritability of analysed phenotypes. We explored the genetic correlations across phenotypes using the LDScore framework and to compare genetic and phenotypic associations, we also analysed associations amongst the studies reproductive diagnoses using a logistic regression framework adjusted for year of birth and 10 genetic principal components. Altogether, our analyses identify 146 genome-wide significant ($p<5\ 10^{-8}$) variants, including numerous novel findings, many of which tag coding variants. We provide the first large-scale map of shared genetic architecture for male reproductive health phenotypes with several estimated pairwise genetic correlations. Overall, we found the genetic and phenotypic correlations to be quite similar and reflecting shared biological background. In conclusion, our work represents the largest effort to comprehensively characterize and map the genetic determinants of different male reproductive health-associated diagnoses.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4013 A meta-analysis of genome-wide association studies of high-sensitive CRP in Korean.

Authors:

K. Oh, Y. Kwak, M. Yuk, S. Yang, J. Lee, J. Youn, N. Song; Chungbuk Natl. Univ., Cheongju, Korea, Republic of

[Introduction]
Circulating hs-CRP (high-sensitive C-reactive Protein) is a representative marker of systemic inflammation. Since chronic inflammation is known as possible risk factor for various chronic diseases such as cardiovascular diseases and metabolic syndrome, the association studies between hs-CRP and chronic diseases have been conducted widely. But genetic factors of high hs-CRP as well as their biological functions and pathways were not fully understood. Therefore, we conducted a meta-analysis of genome-wide association study (GWAS) to identify genetic variants associated with hs-CRP and downstream functional analysis in Korean population.
[Method]
Epidemiological and genome-wide genotype data of 71,020 subjects was collected from HEXA, CAVAS, and Ansan and Ansung study in Korean Genome and Epidemiology Study (KoGES). We performed GWAS using multivariable linear regression model adjusted for sex, age, body mass index, history of hypertension, smoking and regular physical activity. After that, we performed meta-analysis of 3 GWAS results on hs-CRP. Based on our meta-GWAS results, we conducted functional annotation using the 3 strategies (positional mapping, eQTL mapping and chromatin interaction mapping) in Functional Mapping and Annotation (FUMA). For the functional analysis, we also performed gene-based, gene-set and gene-property analyses using the Multi-marker Analysis of GenoMic Annotation (MAGMA).
[Result]
The 188 SNPs associated with hs-CRP were found in this meta-GWAS study ($P<5\times10^{-8}$). Through the FUMA functional annotation, we found 511 mapped genes. Using the MAGMA gene-based analysis, we annotated SNPs to 18,681 protein coding genes, of which 58 were Bonferroni significant ($P<2.68\times10^{-6}$). In total, FUMA annotation and MAGMA gene-based analysis found 514 unique genes significantly associated with hs-CRP. From the MAGMA gene-set analysis of 15,480 gene sets, 19 gene sets were prioritized at FDR <0.05 and the strongest gene set were involved in metabolites or immune and inflammatory response. The gene-property analysis identifying tissue specificity of hs-CRP related gene expression showed the prominent expression levels in 3 of 54 tissues ($P<0.05$, specifically in liver, cells cultured fibroblasts and artery aorta).
[Conclusion]
We found genetic variants associated with hs-CRP and identified functional downstream pathways such as metabolic and/or immune and inflammatory response pathways linking metabolic or cardiovascular diseases.
[Keyword]
Genome-wide association study, Single-nucleotide polymorphism, high-sensitive C-reactive protein, functional analysis, meta-analysis

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4014 A method for fine-mapping time-to-event phenotype based on "sum of single effects" model

Authors:

Y. Yang, K. Tayeb, C. Peter, M. Stephens; Univ. of Chicago, Chicago, IL

With the increasing availability of time-to-event (TTE) phenotypes such as age of onset, disease progression and lifespan, researchers are developing methods for genome-wide association studies(GWAS) based on TTE outcomes. These studies have shown that modeling TTE outcomes, rather than binary outcomes, can enhance the power to detect genetic risk factors (Staley et al, 2017). Having assessed support for genetic associations with a TTE outcome, the next question is which genetic variants are more likely to be causally related to TTE outcome. However, there is a lack of powerful methods for fine-mapping genetic associations based on TTE outcomes. Here, we propose a novel fine-mapping method for TTE phenotypes called the "proportional hazard Sum of Single Effects" (phSuSiE) model. It combines the Cox proportional hazard model, a commonly used model for survival analysis, together with the SuSiE model for fine-mapping quantitative traits (Wang et al, 2020). phSuSiE inherits attractive properties of SuSiE including computational efficiency and intuitive interpretation, and uses credible sets to summarize uncertainty in which genetic variants are the candidate causal variants. We performed simulations with real genotype data to benchmark phSuSiE against two Bayesian variable selection methods for survival analysis, bvsnlp (Nikooienejad et al., 2020) and survival.svb (Komodromos et al., 2022). Our results show that phSuSiE increases power to detect TTE causal variants over the other methods, and it is the only method that provides credible sets. Having demonstrated the benefits of phSuSiE for fine-mapping TTE outcomes, next we plan to develop a software implementation of phSuSiE that can handle “biobank-scale” data sets.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4015 A method to improve GWAS by combining cohorts across genotyping platforms.

Authors:

A. Greenberg[1], H. Kang[2], D. Lee[1]; [1]Boston Children's Hosp., Boston, MA, [2]Univ Michigan, Ann Arbor, Ann Arbor, MI

Many large-scale GWAS cohorts, such as UK Biobank (UKB), are used to improve the power of studies lacking sufficient controls. However, technical differences in array platform are a major obstacle to perform combined analysis. Here, we present 2-stage imputation (2SI), a novel method for merging multiple cohorts which reduces Type I error and batch effects. In brief, (1) cohorts are imputed separately to increase potential SNP overlap (S1), (2) then merged on high-quality SNPs present in both imputed genotypes, and (3) the merged cohort is imputed again to achieve genome-wide coverage (S2). We validate this method by analyzing platform-specific bias, imputation-derived error, and impact on Type I error and GWAS results using data from the 1000 Genomes Project (1KG), UKB, and a case-control psoriatic arthritis (PA) cohort. In initial analyses, we focused on European genetic populations (EUR) and later expanded to include East Asian (EAS). First, we compared allele frequency differences (AFD) after both S1 and S2 for the same 1KG samples genotyped on two different array platforms. In S1, 0.53% of SNPs (n=8,829,116) showed greater than 3% AFD between platforms compared with 0.01% in S2 (n=8,230,518) indicating that 2SI reduces platform-specific bias when combining cohorts. We also performed AFD analysis for imputed genotypes of 1KG samples and their WGS and found similar results for both stages of imputation (S1=5.9%, S2=5.8%) suggesting there is no additional imputation-derived error introduced by S2. Moreover, we applied 2SI and conducted GWAS for controls from PA (n=1355) and a UKB subset (n=12666) using cohort as outcome and found 1 genome-wide significant locus (hit) in S2 ($p = 5\times10\text{-}8$) compared with over 500 in S1 indicating our method controls for Type I error. We also conducted a chromosome 19 GWAS on the merged set of 1KG (n=475) and UKB (n=1787) for EAS using cohort as outcome and found no hits in S2 compared with 3 in S1 ($\lambda S1 = 1.07$, $\lambda S2 = 1.02$). To assess reproducibility of 2SI, we performed GWAS on PA in EUR. We applied 2SI to only PA cases (n=1263) and our UKB EUR subset, simulating a case-only cohort scenario. S1 results showed many hits outside previously known regions. In S2, we replicated 4/5 hits from the original GWAS of PA cases and controls, with the 5th reaching suggestive significance ($p = 4.5\text{x}10\text{-}7$), and no novel loci were found. The GWAS of the full PA cohort combined with UKB achieved similar results, likely due to saturation of the case-control ratio. With the plethora of large biobanks and publicly available control cohorts, 2SI provides an avenue for cohorts of previously understudied phenotypes and populations to be investigated in the expansive GWAS framework.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4016 † A multi-ethnic reference panel to impute classical and non-classical *HLA* class I alleles: Enhancing HLA imputation accuracy in admixed populations.

Authors:

N. Silva[1,2], S. Bourguiba-Hachemi[1], S. H. Y. Knorst[3], R. T. Carmo[3], C. Masotti[3], M. Naslavsky[4], P-A. Gourraud[1], S. Limou[1], E. Castelli[2], N. Vince[1]; [1]Nantes Université, INSERM, Ecole Centrale Nantes, Ctr. for Res. in Transplantation and Translational Immunology, UMR 1064, Nantes, France, [2]São Paulo State Univ., Molecular Genetics and Bioinformatics Lab., Sch. of Med., Botucatu, Brazil, [3]Dept. of Molecular Oncology, Hosp. Sírio-Libanes, São Paulo, Brazil, [4]Dept. of Genetics and Evolutionary Biology, BioSci.s Inst., Univ. of São Paulo, São Paulo, Brazil

The MHC class I region presents several genes related to immune response. Genome-wide association studies have identified many SNP associations within the MHC region. These associations miss the immunological significance of specific *HLA* alleles. To impute these *HLA* alleles from SNPs remains a challenge especially when studying admixed populations as *HLA* genes exhibit high levels of polymorphism, and reference panels consist mainly of Europeans. The SNP-HLA Reference Consortium (SHLARC) aims to enhance HLA association studies by improving HLA imputation.

We called SNP genotypes and *HLA* alleles of 1,000 Genomes (1KG, n=2,504) and SABE (Brazilians, n=1,170) from 30X WGS using a bioinformatics pipeline developed to minimize genotype errors for HLA genes (hla-mapper). We assessed HLA imputation for *HLA-A*, *HLA-B*, *HLA-C*, *HLA-E*, *HLA-F*, and *HLA-G*. Three reference panels were created using HIBAG: 1KG, SABE, and full (1KG+SABE) reference. We extensively tested each dataset by creating several pairs of reference panel + test set: we subsetted each super population of 1KG (Africans, Americans, East Asians, and Europeans) and SABE to obtain 10 random resampling (200 samples each). Individuals of the test set were excluded from the paired reference panel. Additionally, we compared HLA imputation of our reference panels with the Michigan Server (TopMed dataset, n=20,000) for *HLA-A*, *HLA-B*, and *HLA-C* genes in an independent Brazilian population (n=192), with SNP data and *HLA* calls from commercial kits.

Accuracy was assessed by the percentage of correctly predicted alleles. The full reference consistently outperformed the 1KG and SABE alone. When compared to SABE, the 1KG reference generally exhibited better performance, except for predicting subsets from SABE. Non-classical genes achieved around 100% accuracy. We also computed the F1 score, a metric averaging coverage of a specific allele, and the proportion of correct allele calls. For classical genes, the F1 score was lower for *HLA-B* due to its high variability. Among the non-classical genes, *HLA-E* had a lower F1 score in Africans due to the presence of rare and population-specific alleles. The SABE reference was the best to impute Brazilians for *HLA-A* and *HLA-B*. Here, the accuracy for *HLA-B* ranged from 84% (Michigan Server) to 90% (SABE). In conclusion, our findings highlight the influence of population-specific references and target genes on imputation accuracy. Multi-ethnic reference panels generally yield superior results; however, the absence of admixed populations in major reference panels (including TopMed) jeopardizes the imputation accuracy, particularly for highly polymorphic genes.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4017 A multiomics Mendelian randomization study identifying novel therapeutic targets for alcohol use disorder and problem drinking in the cortical proteome

Authors:

D. Rosoff[1], J. Wagner[1], A. Bell[2], L. Mavromatis[1], J. Jung[3], F. Lohoff[4]; [1]NIH/NIAAA, Bethesda, MD, [2]NIH/NIAAA, Rockville, MD, [3]NIH / NIAAA, Bethesda, MD, [4]NIH, Bethesda, MD

Background Alcohol use disorder (AUD) is a common neuropsychiatric disorder that is a leading cause of morbidity and mortality worldwide; however, only a few pharmacological treatment options are currently available, highlighting the need for novel and safe drug development. While protein biomarkers with causal genetic evidence are promising novel drug target candidates for AUD, systematic scans of brain proteins have not been performed.
Methods We integrated genome-wide association summary statistics (GWAS) for AUD and alcohol consumption behaviors, i.e., problem items from the Alcohol Use Disorders Identification Test (AUDIT-P), binge drinking, and total drinks per week (DPW), and applied *cis*-instrument Mendelian randomization (MR) to perform a proteome-wide MR with data from >1,700 brain proteins within the dorsolateral prefrontal cortex (dlPFC) to investigate their relationships with the AUD and alcohol consumption behaviors. We performed replication and validation analyses in independent proteomic and transcriptomic datasets.
Results We identified 34 unique brain protein-alcohol associations that emerged as causal mediators of AUD and alcohol consumption behaviors. Novel proteins not previously implicated in alcohol consumption behaviors included *CAB39L*, *TESC* (P-value=1.9910^{-7} (AUD)), *ERLIN1* (P-value=2.3110^{-12} (AUDIT-P)), *CPS1* (P-value=6.910^{-6} (binge drinking)), *HDGF* (*SLC5A6* (P-value=1.5310^{-7} (DPW)). *CAB39L* was consistently associated with increased drinking across alcohol phenotypes (with P-values ranging from 8.6610^{-10} (AUD) to 1.6910^{-114} (DPW)). We were able to replicate proteins using independent dlPFC protein and gene expression datasets. 11 of the proteins also showed evidence of a shared causal variant between the brain protein and respective alcohol consumption behavior, including *CAB39L*, *HDGF*, and *SLC6A5* with DPW and *CPS1* with binge drinking. Single-cell enrichment was predominantly in excitatory neurons within the dlPFC. MR identified corresponding associations for 31 of the 34 alcohol-related proteins with 22 neuropsychiatric endpoints, highlighting pleiotropic associations. *TESC*, *SLC6A5*, *HDGF*, and *CPS1* were not associated with other neuropsychiatric endpoints, suggesting potentially specific roles in alcohol consumption.
Conclusions Our findings highlight the power of integrating genetics, proteomics, and transcriptomics in elucidating the underlying biology of AUD and alcohol consumption behaviors, linking them with neuropsychiatric disorders and identifying novel drug targets that may aid the development of new therapeutics aimed at reducing problematic drinking.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4018 A network-based evaluation of sex-specific genetic differences in cross-phenotype associations from the UK Biobank

Authors:

V. Sriram[1], Y-Y. Ahn[2], D. Kim[1]; [1]Univ. of Pennsylvania, Philadelphia, PA, [2]Indiana Univ. Bloomington, Bloomington, IN

Across the human diseasome, genetic associations between complex diseases can lead to the onset of comorbidities and longer-term complications. Disease-disease networks (DDNs), graphs where nodes represent diseases and edges represent relationships between diseases, enable the unbiased investigation of genetic components in cross-phenotype associations. While previous studies have analyzed DDNs derived from individual biobanks to identify genetic variants linked to disease interactions, none of these publications have considered more than one DDN at a time. The comparison of multiple DDNs can reveal different patterns of genetic associations between diseases across populations. Additionally, studies over the past two decades have identified sex-specific genetic differences (GxS effects) in the onset of various phenotypes. Given that many diseases exhibit sex-specific differences in their genetic underpinnings, we aimed to determine if their interactions with each other also differ by sex. Thus, the objective of our study was to perform a biologically informed comparison of sex-specific DDNs to evaluate differences in genetics-based cross-phenotype associations between sexes. We built a male- and a female-specific DDN for 102 diseases using sex-specific phenome-wide association study (PheWAS) summary data from twhe UK Biobank (UKBB). We compared the two graphs using network topology, similarity of edge sets, clustering behavior, and node embeddings derived from graph representation learning with node2vec. Our analysis suggests that the diseasomes of males and females behave generally similarly to one another in terms of topology and key central diseases, with hypertensive, chronic respiratory, and thyroid-based diseases playing dominant roles in cross-phenotype associations for both networks. Conversely, autoimmune and inflammatory disorders including multiple sclerosis and osteoarthritis are centrally involved only in the female-specific DDN, while cardiometabolic diseases and skin cancer are more prominent only in the male-specific DDN. We also observed a Jaccard similarity of only 26.39%, indicating that the two networks display vastly different edge sets. These results indicate significant GxS effects in cross-phenotype associations when comparing sex-specific DDNs. Further work will involve an evaluation of clustering patterns in the two DDNs, an assessment of node embedding differences, and an interpretation of the specific genetic variants that contribute to contrasting disease connections across the networks.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4019 A new Approach to Identify Gene-Environment Interactions: Gene by cigarette smoking and alcohol consumption interactions for serum lipids in diverse populations

Authors:

X. Zhu[1], Y. Yang[1], N. Lorincz-Comi[1], G. Li[1], A. Bentley[2], P. De Vries[3], M. Brown[3], A. Morrison[3], C. Rotimi[2], W. Gauderman[4], the CHARGE Gene-lifestyle Interactions Working Group; [1]Case Western Reserve Univ., Cleveland, OH, [2]NIH, Bethesda, MD, [3]Univ. of Texas Hlth.Sci. Cntr Houston, Houston, TX, [4]Univ of Southern California, Los Angeles, CA

There is a long-standing disagreement to the magnitude of the contribution of gene-environment interaction (*G*×*E*) to phenotypic variations of complex traits. Part of the reason may be that very few interactions have been reported to date, presumably owing to the generally low statistical power to detect *G*×*E* by existing methods. To address this issue, the Gene-Lifestyle Interactions Working Group (GLI), as part of the Cohorts for Heart and Aging Research in Genetic Epidemiology Consortium (CHARGE), has been spearheading efforts to investigate *G*×*E* in large and diverse samples through meta-analysis. Here, we present a powerful approach to screen *G*×*E* interactions across the genome and the approach shares substantial similarity to the Mendelian randomization framework, which has been widely applied to infer causal relationships between exposures and disease outcomes. We identified and confirmed 5 loci (6 independent signals) that interacted with either cigarette smoking or alcohol consumption for serum lipids. We further estimated that the lower bound of the interaction and environmental mediated contribution with cigarette smoking or alcohol consumption ranges from 0.21% (LDL-C, regular drinking) to 1.66% (TG, regular drinking) of SNP heritability of serum lipids in Cross-Population data. Our study improves understanding of the genetic architecture and environmental contributions to serum lipids.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4020 A new genetic risk score model for incorporating Gene-Environment interactions

Authors:

E. Dor[1], N. Rappoport[2]; [1]Ben-Gurion Univ. of the Negev, Beer Sheva, Israel, [2]Ben-Gurion Univ. of the Negev, Beer-sheva, Israel

Gene-environment interactions (G×E), refer to how an individual's genetics interact with an environmental factor to influence their traits. These interactions can occur at various levels, from molecular to behavioral, and can have positive and negative effects. Understanding G×E is important because it can help inform the development of personalized interventions and treatments that consider an individual's unique genetic and environmental influences. However, unbiased methods to profile G×E genome-wide are nascent and poorly characterized. A polygenic risk score (PRS) is the overall risk score of an individual for a specific trait based on the individual's genetics. PRS is often computed as the sum of individual genetic loci effects estimated from genetic risk factors study. Existent G×E methods are not part of the PRS estimate. Instead, G×E is focused on detecting genetic loci associated with environmental factors. We developed a new PRS model (PRSG×E), which incorporates genetic data, environmental data, and G×E interactions. We used data of 500,000 subjects from the UK Biobank. We demonstrated the model for two phenotypes: Body mass index (BMI) and lung cancer, and three environmental factors: smoking status, alcohol consumption, and hypertension. We found that the predictive ability of the interaction-based model did not improve beyond that of the PRS model based on genetic variants analysis from GWAS alone. Our findings may suggest that a larger data set is required to uncover small effects of G×E interactions. However, our approach is practical in terms of computing resources and comparable to other PRS computing methods.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4021 A new method to quantify cell type-specific heritability in single cell gene expression data

Authors:

A. Dahl, M. Chen; Univ. of Chicago, Chicago, IL

Single-cell RNA-sequencing (scRNA-seq) has revolutionized our ability to explore cellular heterogeneity across cell types, microenvironments, and individuals. It is now feasible to collect scRNA-seq datasets spanning hundreds of individuals with paired genotype data, which enables studying genetic regulation of gene expression at the level of individual cells. Indeed, several recent papers have established that cell type-specific SNP effects on gene expression can be identified in scRNA-seq data and, moreover, that these effects can inform the causal genes and cell types underlying GWAS hits. However, prior studies have evaluated individual SNP effects on expression, which is less powerful than polygenic approaches which aggregate together many SNPs in a *cis* window near the gene. This is important because scRNA-seq datasets are small and because the causal genes and cell types are typically more biologically and clinically meaningful than the SNP effects. To improve the power of scRNA-seq to characterize genes and cell types driving GWAS effects, we have developed a polygenic model of cell type-specific heritability that applies to scRNA-seq data called GxCTMM. GxCTMM partitions the *cis*-heritability of a single gene's expression into cell type-specific and -shared components, which requires careful modelling of cell-level noise and the heritability of cell type proportion. Through simulations, we demonstrate that GxCTMM provides unbiased estimates of cell type-specific heritability, and in fact that modelling cell-level noise can remove bias in prior estimates of heritability using bulk RNA-seq. We apply GxCTMM to scRNA-seq data obtained from 1.27 million peripheral blood mononuclear cells (PBMCs) collected from 928 donors. Our analysis reveals generally low cis heritability per cell type (median~0.05 across genes), with almost all heritability being cell type-specific. This contrasts with results from bulk sequencing showing that heritability is mostly shared across tissues, clearly demonstrating the biological utility of studying cell types via scRNA-seq. GxCTMM is a powerful and robust tool for detecting and quantifying cell type-specific genetic effects in scRNA-seq data, which can better inform the causal genes and cell types driving GWAS loci.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4022 A novel approach for genetic association test identifies risk genes for dental caries.

Authors:

X. Xu[1], P. Wang[2], M. Li[3], X-Y. Lou[4], S. Xu[5], B. Wu[6], G. Gao[7], P. Yin[2], N. Liu[8]; [1]Indiana Univ. at Bloomington, Bloomington, IN, [2]Huazhong Univ. of Sci. and Technology, Wuhan, China, [3]INDIANA Univ. BLOOMINGTON, Bloomington, IN, [4]Univ. of Florida, Gainesville, FL, [5]Univ. of Hong Kong, Hong Kong, Hong Kong, [6]Univ. of Minnesota, Minneapolis, MN, [7]Univ. of Chicago, CHICAGO, IL, [8]Indiana Univ., Bloomington, IN

Dental caries, also known as tooth decay, is one of the most common chronic oral diseases. Studies suggested that dental caries are partially heritable. Besides, dental caries is influenced by shared genetic factors or close biological relationships which can introduce genetic dependencies. Although several genetic risk loci have been identified in previous genome-wide association studies (GWAS), association tests in these studies are limited to unrelated individuals. Challenges are still existed in detecting associations between genes and dental caries within family samples. Meanwhile, linear mixed model has been widely adopted in GWAS to account for the relatedness owing to its flexibility and effectiveness. In this study, we propose a novel gene-based approach, which integrats the advantages of linear mixed models, aiming at improving identification of genes associated with dental caries in family samples. Our approach utilizes the linkage disequilibrium matrix of SNPs to approximate the correlation between marginal Z scores from linear mixed model, assuming the internal block structure of kinship matrix. This approach enhances the power and accuracy of association signals by considering the intricate genetic mechanisms among samples with family structure. Further, it can also identify suspicious genes that may have been overlooked by prior GWAS in unrelated subjects. When the approach is applied to a dental caries dataset, few risk genes have been identified, which may provide deeper insight into the pathogenesis of dental caries.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4023 A Novel Fine-mapping Method for Multi-ancestry and Admixed Populations

Authors:

Z. Yang[1], I. Ionita-Laza[1], L. Liu[2]; [1]Columbia Univ., New York, NY, [2]Univ. of Pittsburgh, Pittsburgh, PA

One of the crucial steps in Genome-wide Association Studies (GWAS) analyses is conducting fine-mapping to prioritize putative causal variants. Fine-mapping can be achieved through a Bayesian statistical method that leverages linkage disequilibrium (LD) patterns between variants and nearby markers in risk regions identified by GWAS. While multi-ancestry (MA) GWAS have gained attention for discovering new susceptibility loci, existing MA fine-mapping methods assume universal causal variants across ancestries, potentially leading to reduced power and false positives. Additionally, current fine-mapping methods do not account for discrepancies among summary statistics and LD matrices from different reference panels, lacking robustness. Furthermore, no fine-mapping method has been proposed to tailor the unique situation of admixed individuals (e.g. African American), which offer a unique chance to investigate the effects of genetic variants across different ancestry backgrounds with control on environmental factors in contrast to outcomes from studies of cross-continental populations. In this study, we propose a novel MA fine-mapping method, "mCARMA", that utilizes a Bayesian adaptive prior (e.g. Horseshoe prior) to identify ancestry-specific or universal causal variants across MA. mCARMA also includes Bayesian hypothesis testing of MA to address discrepancies in summary statistics and LD matrices. Moreover, mCARMA is compatible with the summary statistics generated by GWAS methods like Tractor, which incorporates local ancestry information from admixed populations. We demonstrate that mCARMA offers increased power for identifying ancestry-specific causal variants, accommodates summary statistics and LD matrices from admixed populations, and exhibits robustness to outliers.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4024 † A novel meta-analysis of total mediation effect for high-dimensional omics mediators and cardiovascular traits in over 5800 TOPMed individuals

Authors:

Z. Xu, P. Wei; The Univ. of Texas MD Anderson Cancer Ctr., Houston, TX

Motivation: Meta-analysis is used to aggregate the effects of interest across multiple studies, while its methodology is largely underexplored in mediation analysis, particularly in estimating the total mediation effect of high-dimensional omics mediators. Large-scale genomic epidemiology consortia, such as the Trans-Omics for Precision Medicine (TOPMed) program, comprise multiple cohorts with diverse technologies to elucidate the genetic structure and biological processes underlying complex human traits. We aimed to estimate the total mediation effects of the transcriptome on age/sex-related phenotypic variation in cardiovascular traits in over 5800 individuals in the TOPMed Framingham Heart Study (FHS) and Multi-Ethnic Study of Atherosclerosis (MESA) using our novel meta-analysis method for total mediation effect accounting for heterogeneity in technology (microarray or RNAseq) and diverse ethnicity.
Methods: Built on our recently established asymptotic standard error of the R-squared (Rsq)-based mediation effect estimation for high-dimensional omics mediators, we have developed a novel fixed-effect and random-effect-based meta-analysis framework requiring only summary statistics and allowing inter-study heterogeneity due to differences in, e.g., high-throughput technologies and ethnicities. We applied the developed method to 8 studies involving 5864 unique individuals: FHS-Offspring/Gen3-microarray/RNAseq and MESA-RNAseq-White/Black/Hispanic/Asian to estimate the mediation effects of gene expression on age-related variation in systolic blood pressure (SBP) and sex-related variation in high-density lipoprotein (HDL) cholesterol.
Results: Extensive simulations showed that the developed meta-analysis method yielded satisfactory efficiency and coverage probability comparable to the joint likelihood-based method requiring individual-level data. Application to the FHS and MESA studies revealed that gene expression explained 18.5% (95% CI=(13.2%, 23.7%) by fixed-effect meta-analysis with I^2=1%) of age-related variation in SBP and 47.2% (95% CI=(28.8%, 65.6%) by random-effect meta-analysis with I^2=69.4%) of sex-related variation in HDL. The proposed Rsq-based mediation effect estimation and meta-analysis will be available in the R package “RsqMed”.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4025 A novel multi-omics mendelian randomization method for gene set enrichment and its application to psychiatric disorders

Authors:

H. Gedik[1], S-A. Bacanu[2]; [1]Virginia Commonwealth Univ., Richmond, VA, [2]Virginia Commonwealth Univ, Richmond, VA

Genome-wide association studies (GWAS) of psychiatric disorders (PD) yield numerous loci with significant signals, but often do not implicate specific genes. Because GWAS risk loci are enriched in expression/protein/methylation quantitative loci (e/p/mQTL, hereafter xQTL), transcriptome/proteome/methylome-wide association studies (T/P/MWAS, hereafter XWAS) that integrate xQTL and GWAS information, can link GWAS signals to effects on specific genes. To further increase detection power, gene signals are aggregated within relevant gene sets (GS) by performing gene set enrichment (GSE) analyses. Often GSE methods test for enrichment of "signal" genes in curated GS while overlooking their linkage disequilibrium (LD) structure, allowing for the possibility of increased false positive rates. Moreover, no GSE tool uses xQTL information to perform a mendelian randomization (MR) analysis. To make causal inference on association between PD and GS, we develop a novel MR GSE (MR-GSE) procedure. First, we generate a "synthetic" GWAS for each MSigDB GS by aggregating summary statistics for x-level (mRNA, protein or DNA methylation (DNAm) levels) from the largest xQTL studies available) of genes in a gene set. Second, we use synthetic GS GWAS as exposure in a generalized summary-data-based MR analysis of complex trait outcomes. We applied MR-GSE to GWAS of nine important PD. When applied to the underpowered opioid use disorder GWAS, none of the four analyses yielded any signals, which suggests a good control of false positive rates. For other PD, MR-GSE greatly increased the detection of GO terms signals (2,594) when compared to the commonly used (non-MR) GSE method (286). The uncovered signals mostly fall into synapse/neuronal, immune, cytoskeleton, RNA translation and chromatin accessibility, miRNA, apoptosis and "housekeeping" categories. Some of the findings might be easier to adapt for treatment, e.g., our analyses suggest modest positive effects for supplementation with certain vitamins and/or omega-3 for schizophrenia, bipolar and major depression disorder patients.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4026 A novel multi-view clustering approach for integrating genotypes and 3D facial images in the presence of confounders.

Authors:

Z. Li[1], S. F. Windels[2], M. L. Marazita[3], S. Walsh[4], S. M. Weinberg[5], M. D. Shriver[6], P. Claes[1], N. Malod-Dognin[2], N. Pržulj[2], K. Van Steen[7]; [1]KU Leuven, Leuven, Belgium, [2]Barcelona Supercomputing Ctr., Barcelona, Spain, [3]Univ Pittsburgh, Ctr. for Craniofacial and Dental Genetics, Pittsburgh, PA, [4]Univ. of Pittsburgh, Pittsburgh, PA, [5]Univ of Pittsburgh, Pittsburgh, PA, [6]Penn State Univ, University Park, PA, [7]Univ. of Liège / KU Leuven, Liège, Belgium

Human populations from distinct geographical locations possess unique genomic and facial patterns. By integrating these multi-view data, we can achieve a more comprehensive and complementary clustering of population structure. Among the various approaches for multi-view clustering, integrative nonnegative matrix tri-factorization (NMTF) has emerged as advantageous in learning low-rank representations of samples and features and straightforwardly interpreting them. When dealing with 3D image datasets, nonnegative Tucker decomposition (NTD) can be treated as a generalization of NMTF. Because confounders can also influence clustering within a population, it is essential to assess and deal with these confounding effects. Here we present a novel multi-view clustering method based on NMTF and NTD to find unconfounded subgroups of population structure from genotypes and facial images. The proposed method is evaluated on a real-life multi-view dataset to subgroup individuals with different population structures in a US cohort based on their genomics and 3D facial images. More specifically, the data consist of roughly 7M SNPs and 7K 3D facial landmarks collected from 4680 US individuals with European ancestry. Covariates, such as age, sex, weight, height, face size and camera system, are available and taken as confounders to population structure. The model generates three representation matrices separately for samples, SNPs and facial landmarks. We test whether the sample representation and its clustering are significantly associated with known covariates. For the SNP representation, K-Means clustering is initially performed and the clusters of SNPs show good enrichment of genes and Gene Ontology (GO) terms. Additionally, we apply hierarchical clustering to the facial representation to obtain a segmentation with improved cophenetic correlation score than a previously published segmentation of the same facial images. Out of all 48 vectors in the sample representation, 20 are not associated with confounders, from which we achieve the unconfounded clustering of population structure. The clusters of individuals are further characterized jointly by the most relevant GO terms and facial regions. In conclusion, this study introduces a novel multi-view clustering method that integrates genotypes and 3D images to derive unconfounded clustering. The evaluation via a real-life dataset demonstrates that the proposed model yields biologically meaningful representations for each data type and obtains unconfounded sample clustering for population structure with a comprehensive interpretation of the derived clusters.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4027 † A Phenome-Wide Systemic Approach to Large-Scale GWAS Meta-Analysis of More Than 1200 Traits in Over 30 Studies

Authors:

D. Jakubosky[1], J. Chen[1], T. Greene[1], B. Gonzalez[2], M. F. Ragsac[1], R. Avila[1], Z. Romer[1], G. McInnes[1], S. Bruse[2], B. Cajes[2], O. Gottesman[2]; [1]Empirico Inc, San Diego, CA, [2]Empirico Inc., San Diego, CA

In the past decade there has been a proliferation of increasingly large genome wide association studies (GWAS), most recently utilizing large-scale biobanks comprising genotyping and/or next generation sequencing data linked to rich phenotypic information, such as the UK Biobank, FinnGen and Biobank Japan. In order to address the lack of power and diversity in single studies, particularly for rare variants or diseases with a lower population prevalence, meta analyses have become an important tool for maximizing the utility of GWAS discovery. However, performing these analyses is difficult, requiring significant effort to curate studies and harmonize inputs, and most meta-analyses efforts consist of only one or a few traits. Here, we describe a phenome-wide systematic approach to large-scale meta-analysis, resulting in one of the largest meta-analyses performed to date. We combined primary GWAS results for thousands of binary and quantitative traits from the UK Biobank with publicly-available summary statistics from more than 30 studies, harmonizing their result formats and gathering detailed metadata about their associated cohorts, phenotypes and statistical methods. We then constructed a knowledge graph of cohorts and studies, and used OpenAI GPT3.5 and all-mpnet-base-v2 to estimate phenotype similarity, matching studies to a custom phenotype ontology. Using this knowledge graph, we algorithmically selected groups of study-trait pairs, requiring that (i) all studies in a group are independent with regard to samples and (ii) group composition is prioritized with regard to total case/sample count. We then performed inverse variance weighted and sample size weighted meta analyses on each group using a spark framework. We generated meta-analysis results for more than 1200 phenotypes with 2-8 studies per group, comprising over 200Million results spanning 12 million unique variants, and revealing a substantial number of previously sub-threshold associations that exceed genome-wide significance following meta-analysis. This study provides important insights into the strategies, methods and value of large-scale meta-analysis for maximizing the utility of GWAS from large and diverse cohorts for genetic discovery at scale. This research has been conducted using the UK Biobank Resource under Application Number 34229.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4028 A pipeline to curate GWAS summary statistics and identify independent traits

Authors:

C. Legros[1,2], A. Kim[1,2], S. Gazal[1,2]; [1]Dept. of Population and Publ. Hlth.Sci., Keck Sch. of Med., Univ. of Southern California, Los Angeles, CA, [2]Ctr. for Genetic Epidemiology, Keck Sch. of Med., Univ. of Southern California, Los Angeles, CA

The public release of genome-wide association studies (GWAS) summary statistics have enhanced discoveries in the genetic architecture of human diseases and complex traits, such as understanding how causal variants are shared across traits. However, the first encountered issues in downstream GWAS analyses are 1) the lack of standardization of summary statistics file formats and 2) the high number of summary statistics from traits that correlate with each other while sharing samples.
To address these issues, we have developed a pipeline designed to facilitate 1) the integration of GWAS summary statistics with different file formats, and 2) the selection of independent summary statistics for further genetic analyses. The pipeline takes as input unprocessed GWAS summary statistics file and handles various input formats of variant information (chromosome, position, allele, variant identification code) and significance of effect sizes (e.g., beta coefficients, p-values, Z scores, Odds Ratio values) and outputs uniformly formatted summary statistics. It includes common quality-control steps such as checking for allele flipping, non-biallelic SNPs and removing duplicated information. Finally, it leverages LD score regression to provide a genetic correlation matrix across traits and define a set of independent traits by excluding genetically correlated traits in overlapping samples.
We leveraged our pipeline to curate hundreds of informative (heritability Z score > 5) GWAS summary statistics of European ancestry from individual publications, the panUK Biobank and Global Biobank Meta-analysis Initiative (GBMI). We defined a set of 130 independent GWAS (mean N = 429K), more than doubling the largest set of independent GWAS previously reported. Our set prioritized 60 disease GWAS, and includes 16 psychiatric diseases and 14 immune related traits.
We anticipate that our pipeline and curated list of independent GWAS summary statistics will be widely used in the GWAS community.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4029 A polygenic risk score of postoperative nausea and vomiting improves risk prediction and shows genetic correlation with CNS conditions including vertigo and motion sickness

Authors:

W. Wang, C. German, Y. Liang, 23andMe Research Team, J. Tung, S. Shringarpure, M. Holmes; 23andMe, Sunnyvale, CA

Postoperative nausea and vomiting (PONV) is a frequent side effect following surgical procedures. Improvement of PONV risk identification may benefit postoperative care. Conventional risk factors, such as sex, smoking status and motion sickness, have been widely studied and utilized in PONV predisposition. However, the polygenicity of PONV and contribution of polygenic risk to PONV remains unclear. 23andMe's large-scale genotype and phenotypic database offers unparalleled opportunities for discovery and evaluation of the polygenic basis and risk of PONV that may have potential implications for clinical practice.

We studied 1.98M participants from 23andMe, Inc research cohort with 659K PONV cases and 1.32M controls across 5 populations (1.65M Europeans, 228K in Latines, 64.4K African Americans, 26.1K East Asians and 6.2K South Asians). We divided participants into 3 non-overlapping datasets: discovery dataset to perform genome-wide association study (GWAS), polygenic risk scores (PRS) training dataset to build PRS, and testing data to evaluate the PRS contribution to PONV risk. We compared the PONV prevalence across risk groups stratified by genetic risk based on PRS and commonly used conventional risk factors, including sex, smoking status and motion sickness.

In the discovery cohort, we identified 185 independent loci associated with PONV in European population and 7 loci in Latino population. The genetic-only AUC of PONV PRS was 0.606 in Europeans; 0.589 in Latines; 0.555 in African Americans; 0.585 in East Asians; 0.594 South Asians. We evaluated the increased prevalence of PONV among European participants in the hold-out testing data, using conventional risk factors and PRS. Women were at 1.26-fold (95% CI: [1.25%, 1.28%]) higher risk of experiencing PONV, adding non-smoking: 1.30 (95% CI: [1.29%, 1.31%]); adding motion sickness: 1.46 (95% CI: [1.44%, 1.48%]), adding high PRS: 1.77 (95% CI: [1.74%, 1.81%]). We found a high genetic correlation of PONV with codeine-induced vomiting (rg: 0.81, P-value: 2.6e-11); and moderate genetic correlation with recurrent vertigo (rg: 0.41, P-value: 1.3e-78), carsick (rg: 0.39, P-value: 5.8e-64), severity of morning sickness (rg: 0.38, P-value: 2.2e-56), dizziness (rg: 0.35, P-value: 5.0e-65), migraine (rg: 0.33, P-value: 1.9e-69) and altitude sickness (rg: -0.33, P-value: 1.9e-37).

Large-scale GWAS reveal the polygenicity of PONV that is predictive of PONV susceptibility. PRS provides additional information beyond conventionally used clinical risk factors at identifying individuals at elevated risk of treatment-emergent side effects.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4030 A Predictive Machine Learning Pipeline for Functional Genomics Screens

Authors:

A. Musolf[1], B. Guo[1], K. Shkura[2], T. Rajasekar[1], A. Vij[2], G. Bryman[1], K. Williams[1], C. Mann[1], A. Stiemke[1], D. Jha[1], N. Raghavan[1], J. Sul[1], S. Longerich[1], M. Kennedy[1], A. Tamburino[1], R. Mathew[1]; [1]Merck & Co., Inc., Rahway, NJ, [2]MSD, London, United Kingdom

Within recent years the application of machine learning (ML) in pharmaceutical industry data science has increased rapidly, where ML has been used for target identification, image analysis, and drug development. We have constructed an ensemble ML pipeline for functional screening prediction. As a general concept, the results of initial assays, such as functional genomic screens, are used as training data to predict which perturbations would pass a future screen. Successful, data-driven prioritization of target selection for screening ultimately would increase efficiency of target selection and screening cost. We have performed testing of our ML method on publicly available data. ML requires predictive features associated with the training set. In consultation with biologists, we leveraged over 70 sets of features, bucketed into 11 categories. Features include genes associated with genetic risk for the disease of interest as well as potential adverse events; protein-protein interactions; key pathway features from cells and tissues of interest as well as potential biological pathways to avoid; transcriptomic and protein expression from tissue of interest, and differential expression in healthy controls versus individuals with disease. Further feature engineering designed to increase informativity include the creation of graph embeddings and the one-hot encoding of location features. Many of the initial features were quantitative values that were binarized based on a predetermined threshold; we plan to use quantitative features in future analyses. Data is analyzed by an ensemble-based pipeline consisting of different learning machines. Each machine analyzes the data in parallel without input from the other machines. Current simulations have shown high power (>80% F statistics) when varying the number of causal features and noise levels. Power decreases as the data becomes more unbalanced; we are currently evaluating techniques for increasing power in unbalanced datasets. All machines output feature importance metrics to identify which features are most predictive; simulations show that these metrics have correctly identified the causal features with high power. This approach is modular and can be utilized for further prediction beyond assays and screens.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4031 A resampling-based approach to share reference panels.

Authors:

T. Cavinato[1], S. Rubinacci[2], A. Malaspinas[1], O. Delaneau[3]; [1]Univ. of Lausanne, Lausanne, Switzerland, [2]Brigham and Women's Hosp., Boston, MA, [3]Univ. of Lausanne, Lausanne-Dorigny, Switzerland

For many genome-wide association studies, imputing genotypes from a haplotype reference panel is a necessary step. Over the past 15 years, reference panels have become larger and more diverse, leading to improvements in imputation accuracy. However, the latest generation of reference panels is subject to restrictions on data sharing due to concerns about privacy, limiting their usefulness for genotype imputation. In this context, we propose RESHAPE, a method that employs a recombination Poisson process on a reference panel to simulate the genomes of hypothetical descendants after multiple generations. This data transformation helps to protect against re-identification threats and preserves important data attributes, such as linkage disequilibrium (LD) patterns and, to some degree, Identity-By-Descent (IBD) sharing, allowing for genotype imputation. Our experiments on gold standard datasets show that simulated descendants up to eight generations can serve as reference panels without significantly reducing genotype imputation accuracy. We suggest that this specific type of data anonymization could be used to generate synthetic reference panels available under less restrictive data sharing policies

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4032 A robust cis-Mendelian randomization method with application to drug target discovery

Authors:

Z. Lin, W. Pan; Univ. of Minnesota, Minneapolis, MN

Mendelian randomization (MR) uses genetic variants as instrumental variables (IVs) to investigate a causal relationship between two traits, an exposure and an outcome. Compared to conventional MR using only independent IVs selected from the whole genome, cis-MR focuses on a single genomic region using only cis-SNPs. For example, using cis-pQTLs for each circulating protein as an exposure for a disease opens an economical path for drug target discovery. Despite the significance of such applications, only few methods are robust to (horizontal) pleiotropy and linkage disequilibrium (LD) of cis-SNPs as IVs. In this work, we propose a cis-MR method based on constrained maximum likelihood, called cisMR-cML, which accounts for LD and (horizontal) pleiotropy in a general likelihood framework. It is robust to the violation of any of the three valid IV assumptions with strong theoretical support. We further clarify the severe but largely neglected consequence of the current practice of modeling marginal effects, instead of conditional effects, of SNPs in cis-MR analysis. Numerical studies demonstrated the advantage of our method over other existing methods. We applied our method in a drug-target analysis for coronary artery disease (CAD), including a proteome-wide application, in which three potential drug targets, PCSK9, COLEC11 and FGFR1, for CAD were identified.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4033 A scalable variational inference approach for increased mixed-model association power.

Authors:

H. Loya, G. Kalantzis, P. Palamara; Univ. of Oxford, Oxford, United Kingdom

The rapid growth of modern biobanks is creating new opportunities for large-scale genome-wide association studies (GWAS) and the analysis of complex traits. Available GWAS algorithms, however, rely on modeling simplifications that reduce statistical power to improve computational costs. We developed a method, called Quickdraws, that simultaneously achieves state-of-the-art statistical power and computational efficiency. Quickdraws improves association power by modeling polygenicity and relies on highly efficient stochastic variational inference and GPU-based operations for parallel trait analysis.

We compared Quickdraws with current GWAS algorithms including Regenie (Mbatchou et al, *Nat Genet*, 2021), FastGWA (Jiang et al, *Nat Genet*, 2019/21), Saige (Zhou et al, *Nat Genet*, 2018), and BOLT-LMM (Loh et al, *Nat Genet*, 2015/18) on simulated and real data. Quickdraws consistently provided higher statistical power while being robust to population stratification, relatedness, and case-control imbalance. In an analysis of 50 disease traits in 404k UK Biobank samples, Quickdraws provided the highest association power, detecting 7.8% more independent associations than Regenie, and 15.8% more than FastGWA-GLMM. Similarly, applied to 50 quantitative blood-related traits, Quickdraws matched the statistical power of BOLT-LMM, the gold standard for quantitative traits, while finding 4.7% more independent associations than Regenie and 31.2% more than FastGWA. The gain in statistical power was higher in less polygenic traits (e.g. 23.3% and 22.2% increase over Regenie for Celiac disease and Hypothyroidism). Quickdraw's higher association power also results in a higher number of replicated associations in other cohorts (e.g. 3.2% and 14.7% more replicated associations compared to Regenie and FastGWA in Biobank Japan).

We compared the computational cost of Quickdraws and other algorithms using 404k UK Biobank samples on the RAP platform. We used 458k model SNPs and tested 13.3 million imputed SNPs across 50 quantitative and 50 disease traits, tuning the hardware configuration for each method. Quickdraws was 147x cheaper than BOLT-LMM, the only method with similar association power in quantitative traits, and was comparable to more cost-effective methods that have lower association power (50 disease traits: Quickdraws=£452.1, Regenie=£506.9, FastGWA-GLMM=£98.1, Saige≈£2803; 50 quantitative traits: Quickdraws=£50.8, Regenie=£24.2, FastGWA=£37.3, BOLT-LMM≈£7467).

These results highlight the promise of using modern machine learning methodology to reduce the costs of GWAS analyses without sacrificing statistical power and robustness.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4034 A simulation study comparing cognitive decline measures in Alzheimer's disease

Authors:

R. Shi, E. Feingold, M. Kamboh, K. Fan; Univ. of Pittsburgh, Pittsburgh, PA

Background: Cognitive decline in Alzheimer's disease (AD) can be assessed over time using neuropsychological tests such as the Mini-Mental State Examination (MMSE) score. Various functions of longitudinal decline can then be used as phenotypes for genome-wide association studies (GWAS). One difficulty in choosing the most appropriate measure is that there is significant missing data. Left censoring occurs when the diagnosis time is uncertain, while right censoring happens when patients are lost to follow-up as the disease progresses to a severe stage. This study proposes several different MMSE-based cognitive decline measures and evaluates their power for GWAS analyses. **Methods:** We evaluated the power of six alternative cognitive decline measures calculated from the longitudinal MMSE scores across various simulated scenarios. We built different simulated cognitive decline patterns using parameters estimated from linear mixed-effect models of two real datasets. We employed masking strategies to mimic random missing, left-, and right-censoring patterns on the simulated full trajectories. For each measure, we conducted analyses with and without adjusting for baseline MMSE score as a covariate. **Results:** Most measures showed increased power after adjusting for baseline MMSE score, except for the annual cognitive decline rate and percentage MMSE score changes that already incorporate the baseline level in the calculations. Among all six measures, a simple measure of the slope of cognitive decline was more powerful under all missingness-induced sigmoid-like decline patterns, which was estimated to be over 75%, followed by the time-to-reach MMSE10. The time-to-reach MMSE10 is more powerful in selected scenarios in which the censoring rate is lower and with a shorter time to reach that point. **Discussion:** Overall, our results suggest that a simple measure of slope and the time-to-reach MMSE10 may be most powerful for capturing the genetic effects underlying AD-related cognitive decline, particularly in datasets with significant left and right censoring. An additional advantage of these methods is that they are computationally faster than repeated-measures approaches. Adjusting for baseline MMSE score is very important, as it helps to reduce confounding and improves the accuracy of decline rate estimates.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4035 A statistical approach for detecting trends in underpowered genome-wide association studies

Authors:

M. MacMillan, S. Demiriz, T. Furneaux, B. Ballew; Sequence Bio, St. John's, NL, Canada

Genome-wide association studies (GWAS) have accelerated our understanding of the genetic landscape of complex human disease. A well-powered GWAS, a function of sample size and effect size, is required for reliable signal; the field benchmark for association is genome-wide significant results of $p<5 \times 10\text{-}8$. Genomics groups who are embarking on a new area of study or are developing new in-house bioinformatics capabilities may find it challenging to achieve sufficient power when attempting to confirm that their methods are functioning as expected. When developing NGS methods, we often rely on comparisons to high confidence call sets via the Genome in a Bottle Consortium, but an analogous method to verify that GWAS workflows are capturing known signals at relatively small sample sizes does not yet exist. Here, we present a statistical method for benchmarking GWAS pipelines in underpowered studies. This approach aims to determine if, although not at genome-wide significance, our association findings are enriched for putative true positive signals at lower p-values than would be expected by chance. We developed our trend detection pipeline using a cohort of around 2,000 genotyped and QC'd participants from the Newfoundland and Labrador population. Based on the phenotypes for which we had the largest sample size, we performed association testing on our cohort for each of the selected traits to derive experimental p-values for all SNPs that passed QC. We used the European Bioinformatics Institute GWAS catalog, filtered to $p<5x10\text{-}8$, to establish a benchmark "true positive" dataset of variants that have been previously associated with our traits of interest. Receiver Operating Characteristic (ROC) curves were generated to evaluate the performance of our association workflow, and the GWAS catalog was mined for known signals in ~independent traits to serve as expected null controls. If the rate at which we are correctly identifying true positive and true negative signals is greater than the rate at which we are misclassifying signals, we'd expect a curve into the upper left quadrant of the plot. We compared different approaches to try to ensure expected behaviour of the null model, including implementing LD clumping for experimental SNPs, improving variant joining precision and controlling for allele frequency matching. Our performance curve showed a statistical trend toward significance for our test trait associations, boosting confidence in the methodology of our GWAS workflow, sample processing and data integrity. This technique provides a potential benchmarking approach for assessing the validity of association workflows in underpowered studies.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4036 A stringent filter for putative NMD escape variants improves rare variant association tests.

Authors:

A. Holleman, A. Deaton, R. Hoffing, L. Krohn, P. LoGerfo, P. Nioi, M. Plekan, S. Akle, L. Ward; Alnylam Pharmaceuticals, Cambridge, MA

Genetic variants like frameshift and stop-gain mutations typically result in gene loss of function (LOF) through nonsense-mediated decay (NMD). However, if these variants are located near the end of a gene's coding sequence, the resultant transcript may escape NMD, in which case the variant does not cause LOF. Such variants may nonetheless be classified as high confidence predicted LOFs (HC pLOFs) by commonly used annotation algorithms like LOFTEE, and they may reduce power for identifying associations if included in LOF burden analyses. Using UK Biobank (UKBB) whole-exome sequencing data for ~407,000 participants with European ancestry, we first generated sets of rare (MAF < 1%), LOFTEE-designated HC pLOFs for all genes. These "standard" variant sets included 531,919 HC pLOFs. We then generated "curated" pLOF variant sets by taking the standard variant sets and removing variants located within the last exon and the last 25% of coding sequence for each transcript in which it was designated by LOFTEE as a HC pLOF. This removed 37,054 (7%) of the LOFTEE HC pLOFs. In total there were 11,300 genes with a reduced number of HC pLOFs in the curated versus standard variant sets. We used UKBB Olink proteomics data to examine whether the curated variant sets may be enriched for true LOFs. For 219 genes, we tested the burden of curated and standard set variants for association with levels of the gene's encoded protein. The curated set yielded a stronger estimated effect on decreased serum protein levels ($\beta_{curated} < \beta_{standard} < 0$) for 147 genes (67%; $p_{binom}=4.5x10^{-7}$), and for 5 of these 147 genes the 95% CIs for $\beta_{curated}$ and $\beta_{standard}$ did not overlap. Among 17 genes for which the curated set had ≤ 50% of the carrier count of the standard set, we observed $\beta_{curated} < \beta_{standard} < 0$ for 14 genes (82%; $p_{binom}=0.01$), and for 2 of these 14 genes the 95% CIs for $\beta_{curated}$ and $\beta_{standard}$ did not overlap. These results are consistent with the curated variant sets being enriched for true LOF variants. We then tested the 11,300 gene-based variant sets mentioned above for association with 607 UKBB quantitative traits (QTs), with the with the aim of finding associations not detected using the standard variant sets. Analyses of the curated variant sets identified 69 associations at $p < 1.25x10^{-6}$ not identified by the standard sets. These included novel associations between *PLIN1* and six cholesterol- and fat-distribution-related QTs, strengthening evidence for previously known associations between *PLIN1* LOFs and decreased visceral adiposity. This work helps to demonstrate that additional curation of pLOFs beyond the standard output of LOF-prediction algorithms may aid in the discovery of novel therapeutic targets.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4038 A Systematic Exploration of Phenotypic and Genetic Relationships Between Physical Activity and Brain Health

Authors:

Y. Guo[1], B. Zhao[2], P. Paschou[1]; [1]Purdue Univ., West Lafayette, IN, [2]Univ. of Pennsylvania, Dept. of Statistics and Data Sci., Philadelphia, PA

Physical activity, an essential lifestyle aspect, is associated with various health outcomes including mental health and brain functions. Using multiple data types from the UK Biobank and a variety of tools, we aimed to examine the phenotypic and genetic relationships between physical activity and brain health. We identified brain regions and networks that are strongly associated with the variation in daily physical activity volume and circadian rhythm. We also expect to quantify putative causality between them with the help of genetic instruments. This study is designed to advance our knowledge of the genetic architecture and biological networks involved in the beneficial effects of physically active and balanced lifestyle behaviors on brain health in a systematic way.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4039 A test of genotyping platform bias for multiethnic case/control association studies merging external controls

Authors:

T. Tu[1], R. Gbadegesin[2], A. Adeyemo[3], A. Ochoa[1]; [1]Computational Biology and Bioinformatics Program, Duke Ctr. for Statistical Genetics and Genomics, and Dept. of Biostatistics and Bioinformatics, Duke Univ., Durham, NC, [2]Dept. of Pediatrics, Div. of Nephrology, Duke Univ. Med. Ctr., Durham, NC, [3]NIH, Bethesda, MD

Every genotyping platform, including array, sequencing, and imputation, is known to have biased errors that vary between platforms. Merging data from different genotype platforms is a common step prior to conducting GWAS, especially in case-control studies where control samples from various sources are often combined to increase the sample size while reducing costs by reusing existing data. There is an increased risk of type I error when these platform-differential biases are correlated with a variable of interest, confounding the analysis.
We propose a framework for detecting platform-differential biases in case-control data from two genotyped datasets with different platforms and labeled ancestry. The two datasets must have control samples with one or more overlapping homogeneous ancestries. However, no cases are required, allowing analysis where one of the datasets is entirely control samples. Using control samples of the same ancestry only, we assume that allele frequencies are the same between datasets, and identify biased SNPs if allele frequencies are found to be significantly different. Our working model assumes Binomial distribution of genotypes and evaluates significance using a likelihood ratio test, but alternative models are also proposed. This method is extended to a joint ancestry test by treating each ancestry as independent data and considering the joint null hypothesis that all allele frequencies of each ancestry between the two datasets are equal. In a real data scenario where a mixture of cases and controls genotyped on an array are combined with 1000 Genomes sequencing data, which are all treated as controls, this approach successfully identifies for exclusion numerous false positive SNP associations without removing true signals. Additionally, this method can be applied to align reverse complement allele pairs between datasets, which are often misaligned due to strand ambiguity. The test is based on allele counts and can be applied to summary statistics, which we demonstrate on a replication study that uses gnomAD as the control group. Overall, this approach can allow us to safely increase study sample size by efficiently quantifying and controlling platform-differential SNP biases when controls of the same ancestries are available across platforms.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4040 A Transformer-Based Architecture for Cell Type-Level Analysis of Alzheimer's Disease Using Single-Cell RNA Sequencing Data

Authors:

X. Liu, J. Gu, R. R. Butler III, F. M. Longo, Z. He; Stanford Univ., Palo Alto, CA

Alzheimer's disease (AD) is a neurodegenerative disorder characterized by cognitive and behavioral impairments significantly impacting daily activities. Understanding the pathogenesis of AD is crucial for developing effective prevention and treatment strategies. Recent advances in single-cell RNA sequencing (scRNA-seq) technology have provided unprecedented insights into the molecular mechanisms underlying AD.
This study presents an end-to-end architecture based on the Transformer for analyzing scRNA-seq data for AD, focusing on cell type-level analysis. We train our models on fourteen different cell types with cell state classification as the objective to validate model fitting. The proposed model takes all the cells in a particular cell type as input and uses the Transformer-based architecture to effectively capture both linear and nonlinear patterns within the data. Notably, our method incorporates an essential feature extraction step that identifies and ranks genes based on their relevance to predicted cell states independently for each cell type. Additionally, our model constructs cell-type-specific gene-gene interaction networks and identifies gene modules that provide valuable insights into AD status. Results demonstrate superior performance compared to existing methods in identifying known AD-related genes. Moreover, the generated cell-type-specific gene interaction networks capture a higher representation of unique AD-related genes, surpassing traditional marginal association methods with minimal overlap observed between cell types. Furthermore, functional enrichments of the cell-type-specific AD modules reveal significant enrichment of biologically relevant pathways associated with AD, highlighting distinct functional characteristics specific to different cell types. Importantly, our model achieves accurate classification of AD versus control status for most cells across all cell types with AUC up to 0.97, underscoring the broad impact of AD across multiple cell types and genes. These findings suggest the potential utility of several cell types for disease diagnosis in the future.
In summary, our proposed method provides a robust framework for defining disease states and offers valuable insights into exploring causal genes in the context of AD. Beyond the scope of AD, the gene-feature-aware classification approach aware classification has considerable potential to impact various complex diseases—through diagnosis/stratification, the discovery of gene targets, and implication of cell types—hopefully leading to more personalized disease treatment and improved study of disease heterogeneity.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4041 A Two-stage Linear Mixed Model (TS-LMM) for Summary-data-based Multivariable Mendelian Randomization

Authors:

M. Ding; Univ. of North Carolina at Chapel Hill, Chapel Hill, NC

Multivariable Mendelian randomization (MVMR) methods provide a strategy for applying genome-wide summary statistics to assess simultaneous causal effects of multiple risk factors on a disease outcome. In contrast to univariate MR methods that assumes no horizonal pleiotropy (genetic variants only associate with one risk factor), MVMR allows for genetic variants associate with multiple risk factors and models pleiotropy by including summary statistics with risk factors as multiple variables into the regression model. Here, we propose a two-stage linear mixed model (TS-LMM) for MVMR that accounts for variance of summary statistics not only in outcome, but also in all of the risk factors. In stage I, we apply linear mixed model to treat variance in summary statistics of disease as fixed-/random-effects, while accounting for covariance between genetic variants due to linkage disequilibrium (LD). Particularly, we use an iteratively re-weighted least squares algorithm to obtain estimates for the random-effects. In stage II, we account for variance in summary statistics of multiple risk factors simultaneously by applying measurement error correction methods that take into consideration LD between genetic variants and correlation between summary statistics of risk factors. We compared our MVMR approach to other approaches in a simulation study. When most of the instrumental variables (IVs) were strong, our model generated the highest coverage of true causal associations, the highest power of detecting significant causal associations, and the lowest false positive rate of identifying null causal effect for a range of scenarios that varied correlation (weak, strong) between summary statistics of risk factors and LD among genetic variants (weak LD [$2 \leq 0.1$], moderate LD [$0.1 < 2 \leq 0.5$]). When the proportion of strong IVs was reduced, our model showed performances comparable to MVMR-Egger and MVMR-IVW. The more accurate inference of our model in the presence of correlation among risk factors supports potential wide application to -omics data that are commonly multi-dimensional and correlated, as shown in application to determinants of longevity, where our method nominated a specific significant lipoprotein subfraction for causal association from a panel of 10 lipoprotein cholesterol measures. The robustness of our model to correlation structure suggests that in practice we can allow moderate LD in selection of IVs, thereby potentially leveraging genome-wide summary data in a more effective manner. Our model is implemented in ‘TS_LMM’ macro in R (https://github.com/mingding-hsph/TS_LMM).

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4042 A unified framework for phenome-wide genetic effect variance and correlation estimation utilizing GWAS summary statistics with applications to UK Biobank

Authors:

G. Chen, T-Q. Thach, P. Sham; The Univ. of Hong Kong, Hong Kong, China

Whilst pleiotropic SNPs are ubiquitous in the genome, we have also observed widespread correlations in the marginal effect estimates (Z-scores) of associated SNPs across numerous GWAS studies. This finding underscores the critical need to delve deeper into the complex architecture of the joint genetic effect matrix, encompassing multiple SNPs across hundreds of phenotypes. We're primarily interested in phenome-wide **genetic-effect variance and correlations**—relatively uncharted facets of genetic architecture, which quantify the functional relevance of a SNP and the functional commonalities between two SNPs, respectively. These perspectives parallel the heritability of individual traits and the genetic correlation between phenotypes, but are examined across the phenome instead. However, interpreting GWAS-wide variance and covariance components of summary statistics is complicated. These can inflate due to genuine genetic-effect relationships or be skewed by confounding factors, such as linkage disequilibrium (LD), shared population stratification, and GWAS non-independence. To distinguish true signals from potential bias using current methodologies, we've devised a unified statistical framework assuming **a point matrix normal distribution** of joint genetic effects. This framework enables two methods—**LRCP and LRCQ**—to respectively estimate genetic-effect correlations and variance. It leverages likelihood and regression-based approaches to analyze relationships in summary statistics and examines the connection between Z-score cross products and LD cross components.

When we applied LRCP to UK Biobank data, it revealed prevalent genetic-effect correlations within pleiotropic SNPs, presenting a fresh avenue to identify functionally-linked SNPs, gene clusters, and pathways. In contrast, LRCQ provided an innovative assessment of the functional significance of SNPs across different phenotypic categories in UK Biobank such as cardiometabolic and neuropsychiatric diseases. The detection of many "omni-SNPs" through LRCQ alludes to the possibility of identifying associated core genes and regulatory pathways. Meanwhile, "null SNPs"—those with a zero variance estimate and thus no discernable functional implications on any traits—may be disregarded in further analyses, reducing much analytical burden. Combined evidence from diverse functional annotation data further substantiates some of our findings. The combined use of LRCP and LRCQ allows for a nuanced interpretation of large-scale GWAS data, steering us towards a comprehensive comprehension of phenome-wide genetic architecture and its functional consequences.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4043 A unified model for cell-type resolution genomics from heterogeneous omics data

Authors:

Z. Chen[1], E. Rahmani[2], E. Halperin[2,1,3,4]; [1]Dept. of Computer Sci., Samueli Sch. of Engineering, Univ. of California Los Angeles, Los Angeles, CA, USA, Los Angeles, CA, [2]Dept. of Computational Med., David Geffen Sch. of Med., Univ. of California Los Angeles, Los Angeles, CA, USA, Los Angeles, CA, [3]Dept. of Anesthesiology and Perioperative Med., David Geffen Sch. of Med., Univ. of Califonia, Los Angeles, CA, USA, Los Angeles, CA, [4]Dept. of Human Genetics, David Geffen Sch. of Med., Univ. of Califonia, Los Angeles, CA, USA, Los Angeles, CA

INTRODUCTION: The vast majority of population-scale genomic datasets collected to date have been measured as "bulk" samples from heterogeneous tissues. The critical need for cell-type level analysis to enable discovery therefore calls for computational methods that can disentangle the convolution of cell-type signals in these population-scale bulk datasets. Such deconvolution methods exist, yet they neglect the fact that genomics is often different yet coordinated between cell types, which can provide additional information for deconvolution. Furthermore, existing methods are either heuristic or sensitive to violations of parametric assumptions.
OBJECTIVE: We aimed to develop a general framework for heterogeneous bulk omics that enables deconvolution of unknown cell-type level genomic profiles (samples by features by cell types) from bulk genomic data (samples by features).
METHODS: We present a non-parametric model, which, unlike previous work, explicitly models covariance across cell types, and an inference procedure that is theoretically justified for all heterogeneous omics. We benchmarked our method with CIBERSORTx and TCA, state-of-the-art (SOTA) methods for deconvolution of gene expression and DNA methylation.
RESULTS: We first benchmarked all methods in deconvolving in-silico mixtures of cell-type profiles we sampled from PBMC (n=118 donors) and lung (n=90) scRNA-seq data using a range of sample sizes and number of cell types. Our method consistently achieves substantial gain in the correlation between estimated and the true underlying cell-type level profiles by up to 37% and 35% over SOTA in PBMC and lung, respectively. We further confirmed the transferability of our model to other genomic modalities by demonstrating a 69% improvement in deconvolving bulk DNA methylomes for which corresponding cell-type methylation levels were experimentally available for validation. We next asked whether deconvolution can recapitulate differential expression (DE) in follicular lymphoma B-cells with CREBBP mutation that was previously found via experimental cell sorting. Strikingly, not only did our method outperform SOTA in identification of known DE genes, but it was also the only method that improved upon a straightforward analysis of the bulk data (p=8e-5; paired wilcox-test). Finally, we show that accounting for cell-type covariance establishes superior replication rates in identification of sex- and age-associated cell-type level differential methylation across four large whole-blood methylation datasets (average F1=0.94, 0.84 for sex and age, respectively, across pairs of datasets; ≤0.89 and ≤0.81 for SOTA).

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4044 Accelerated Failure Time and Parametric Shared Frailty Model with the Application of Infant Mortality EDHS 2016 Data, 2023.

Authors:

A. Aragaw[1], E. Muluneh[1], A. Azene[2]; [1]Bahir Dar Univ., Bahir Dar, Ethiopia, [2]Dept. of Epidemiology and Biostatistics, Coll. of Med. and Hlth.Sci., Bahir Dar Univ, Bahir Dar, Ethiopia

Background: Infant mortality is a major public health issue in Ethiopia. The main objective of this study was to compare accelerated failure time with parametric shared frailty model and identify predictors of infant mortality in Ethiopia using EDHS 2016 data.
Methods: The study used data from the 2016 Ethiopian Demographic and Health Survey. Survival analysis techniques such as the Kaplan-Meier survival estimates, accelerated failure time model and parametric shared frailty models were employed to compare and analyze the data for survival time of infants.
Results: The clustering effect was significant and the Weibull-Gamma shared frailty model was best model for predicting survival time of infant in Ethiopia. Mothers educational level prolonged the survival time of infants by a factor of (AF =1.403 and AF =2.094) for primary and secondary and above education as compared to uneducated mothers. The acceleration factor for infants who were breastfeed was 6.307 times greater than those who were not breastfeed (reference) (AF=6.307, 95%CI= 4.976, 7.995), this indicates mothers who breastfeed their infants was prolonged the survival time of infants by the factor of 6.307 than the reference group.
Conclusion: Weibull-gamma shared frailty model was best for survival time of infants in Ethiopia. Considering the region and cluster variation, intervention should be given for who had not educated mother and fulfill the facilities for regions to achieve the target for Sustainable Development Goals (SDGs) by 2030, and one of the important components is to reduce the infant mortality rates. This study also implies that health professionals should give an emphasis on improve the awareness to increase birth spacing and encourage breastfeeding's to reduce infant mortality in the country.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4045 Accounting for cardiac cell composition in GTEx with single-nucleus reference profiles yields novel genetic-trait associations.

Authors:

B. Gural, C. Rau; Univ. of North Carolina at Chapel Hill, Chapel Hill, NC

Heart failure (HF) is a heterogeneous condition defined by an inability of the heart to meet the circulatory needs of the body. HF is the result of a maladaptive response to chronic stressors during which the heart undergoes structural and cellular remodeling. For example, cardiomyocytes undergo apoptosis and necrosis after chronic overstimulation while fibroblasts proliferate to fill gaps left by the loss of larger cells. Although changes to the cellular makeup of the heart are known to mediate HF pathology, estimates of basal cardiac cellular composition vary widely. Moreover, despite our understanding of HF as a complex condition that is influenced by environmental and genetic factors, most studies on heart composition only examine small groups of individuals or inbred models. Recent efforts have used single-nucleus RNAseq (snRNAseq) to tally nuclear proportions as a proxy for cellular proportions. However, cellular dissociation biases for certain cell types and skews the resulting estimates. Furthermore, snRNAseq studies are expensive, making them infeasible and underpowered for population-scale analyses. To avoid these issues, we developed a method to estimate cellular composition of bulk RNAseq using cell-type-specific references defined in snRNAseq. Current deconvolution tools are designed for single-cell RNAseq inputs. Our method applies existing deconvolution tools, preprocesses the inputs to control for technical and biological differences when using snRNAseq, incorporates validation tests with appropriate compositional statistics, and allows for the inclusion of additional reference datasets. We apply our method to the Genome-Tissue-Expression atlas (GTEx), a multi-tissue database with bulk RNAseq from hundreds of subjects and snRNAseq from a small subset of those subjects. We estimated left ventricular cellular composition 386 individuals, revealing considerable heterogeneity. We found significant correlations between subject phenotypes and discrete cell populations. Further, we have used these estimates both as covariates in eQTL analysis and as a trait of their own in a cell-type QTL (ctQTL). To improve the rigor of our results, we have also conducted colocalization of these separate signals. This analysis demonstrates our ability to leverage existing, public datasets to accurately map the understudied variability of cardiac cell type variation across human populations.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4046 Accounting for distant relatedness with genealogical data in a multivariate analysis of eye disease and cognitive phenotypes

Authors:

M. Rakesh, A. Montazeri, K. Burkett, E. Freeman, M-H. Roy-gagnon; Univ. of Ottawa, Ottawa, ON, Canada

Background: Retinal diseases and cognitive phenotypes may share biological mechanisms. To explore this relationship, we investigate 36 candidate single nucleotide polymorphisms (SNPs) previously identified to be associated with either glaucoma, age-related macular degeneration (AMD), Alzheimer's, or cognitive decline in a cross-sectional study of 312 participants from the French Canadian founder population recruited in the Ophthalmology Clinics of the Maisonneuve-Rosemont Hospital, Montreal, Canada.
Methods: We tested for association between each candidate SNP and the multivariate phenotype of glaucoma (or AMD) and 6 continuous cognitive traits including the Verbal Fluency Test animal and letter versions, the Verbal Digit Span forward and backward versions, and the logical memory story with immediate and delayed recall. We used a multivariate analysis of variance approach and a reverse regression approach, where the SNP is set as the independent variable and the multivariate phenotype as the dependent variable, allowing for covariate adjustment. No genome-wide genotype data are available on these individuals. However, genealogical data from the BALSAC database are available for a subset of 227 individuals. In order to assess whether cryptic relatedness may cause spurious associations within our sample, we used these genealogical data to perform gene dropping simulations (1000 replicates of 2200 genome-wide SNPs 50kb apart) under the null hypothesis of no association with the R package GENLIB.
Results: Overall, our simulations did not indicate an inflated type 1 error (estimated over the 2200 SNPs) in the majority of the simulation replicates. However, in the specific replicates where type 1 error was elevated (n=38), mixed models with the genealogical kinship matrix as the covariance structure controlled the type I error rate. Our results replicate some of the previously identified associations with the selected candidate SNPs. In the multivariate analysis, we found a significant association ($p < 0.001$) with SNP rs11200638 in gene *HTRA1* involving the AMD and the logical memory story phenotypes.
Conclusion: We showed that, in the absence of whole-genome genotype data, using BALSAC genealogical data can adjust for cryptic relatedness in the French-Canadian founder population, which allowed us to find evidence of a multivariate association involving AMD and cognitive phenotypes.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4047 Accurate Estimation of Genetic Admixture Proportions Via Haplotype Modeling

Authors:

S. Ko, E. Sobel, H. Zhou, K. Lange; Univ. of California, Los Angeles, Los Angeles, CA

Estimating admixture is essential for creating personal genealogies, studying human history, and conducting genome-wide association studies (GWAS). There are many computer programs available for determining the admixture proportions of individuals. Some employ Bayesian inference on the binomial model (such as STRUCTURE, fastSTRUCTURE, and TeraStructure), while others involve direct optimization of the modeled log-likelihood (like ADMIXTURE and OpenADMIXTURE). Yet, there are also programs that decompose the genotype matrix algebraically (such as ALStructure and SCOPE).

Over time, each of the approaches mentioned above has been scaled up to accommodate one million individuals and one million single nucleotide polymorphisms (SNPs). Although these methods often perform linkage disequilibrium (LD) filtering before analysis, unfortunately, this process loses some information and still retains some SNPs in LD. In this work, we explicitly incorporate haplotypes into ancestry estimation. Our latest program, HaploADMIXTURE, aims to improve ancestry accuracy by grouping nearby SNP pairs and jointly estimating their haplotype frequencies with admixture coefficients. This strategy takes advantage of the rich information present in haplotypes, resulting in improved admixture estimation and better identification of population clusters in real-world datasets.

Our simulations demonstrate superior inference on large datasets. Real data analyses also show better performance in clustering self-identified population labels compared to OpenADMIXTURE, SCOPE, and PCA. The number of populations can be determined using cross-validation or the Akaike information criterion. Both criteria accurately select the number of populations in simulated data and identify reasonable structures in curated real datasets.

By leveraging high-performance computing techniques such as GPUs and multithreading, we are able to run haplotype-based ancestry inference on several large contemporary datasets. For instance, it takes less than two hours to analyze 1000 Genome Project data with 1718 individuals and 1.8 million SNPs using 10 threads. Only SCOPE and OpenADMIXTURE yield shorter computing times, ignoring, of course, the extra information present in haplotypes.

Additionally, we consider unsupervised ancestry informative marker (AIM) selection through the sparse K-means with feature ranking (SKFR) method implemented in OpenADMIXTURE. Inferred ancestries demonstrate increased stability after AIM selection. In short, attempting to extract extra information from uninformative SNPs is actually counterproductive.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4048 Adapting the AD8 Dementia Screening Interview and the Clinical Dementia Rating (CDR) scale for use in Tongan

Authors:

B. Vaitohi[1], J. P. Tavana[1], R. J. Hansen[1], C. S. L. Pakalani[1], T. D. Packard[1], A. T. Perez[1], S. Polaapau[1], S. F. Funaki[1], F. Matthes[1], J. T. Tschanz[2], J. Kauwe[3], P. Ridge[1]; [1]Brigham Young Univ., Provo, UT, [2]Utah State Univ., Logan, UT, [3]Brigham Young Univ, Provo, UT

Native Hawaiians and Pacific Islanders (NHPI) are the least represented racial group in most Alzheimer's disease (AD) datasets (e.g., in NACC, the National AD Coordinating Center, there are only 36 total NHPI samples out of 50,000 total). Moreover, English language-based neuropsychological evaluations are ineffective for many NHPIs whose first language is not English.

The Clinical Dementia Rating (CDR) scale and the AD8 are robust tools for assessing dementia. The CDR evaluates individuals in six categories: memory, orientation, judgment and problem-solving, community affairs, home and hobbies, and personal care. The AD8, derived from the CDR, is a concise screening instrument comprising eight questions designed to quickly determine whether or not an individual has even very mild dementia. However, the tests are only effective for people who speak English with native proficiency. Our objective in this study is to validate both CDR and AD8 in the Tongan language.

We modified and translated the CDR and AD8 into Tongan. We evaluated the accuracy of both the translations and assessments. A bilingual psychologist and expert translator back-translated the Tongan versions of the AD8 and CDR. The back-translated AD8 and CDR were nearly identical to the original versions, demonstrating that our translations to Tongan are accurate.

Next, we validate the Tongan version of the AD8 and CDR tools. We recruited 50 mono- and bilingual Tongans. NACC-certified bilingual examiners conduct the interviews. We are using the NACC neuropsychological test battery to validate our results. We will present our validation results at ASHG; our initial results indicate our Tongan tests are effective.

NHPIs account for 0.08% of NACC participants, despite being the second fastest-growing racial group in the USA, second only to Asian Americans. However, accurately diagnosing AD in NHPIs is imperative to building diverse cohorts and reducing the medical disparities in NHPIs generally and Tongans specifically. Our Tongan partners are eager to participate in ways to improve medical outcomes in their communities.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4049 Adjusting genetic confounders lead to detection of causal molecular traits underlying complex phenotypes

Authors:

S. Qian[1], W. Crouse[2], S. Zhao[3], K. Luo[1], M. Stephens[1], X. He[1]; [1]Univ. of Chicago, Chicago, IL, [2]Univ. of Chicago, Apex, NC, [3]Dartmouth Coll., Hanover, NH

While GWAS have identified many loci associated with complex traits, causal genes in these loci often remain unknown. A number of methods have been developed to leverage expression QTL data to nominate candidate genes, including Transcriptome-wide association studies (TWAS), colocalization analysis, and Mendelian Randomization (MR) methods that use cis-eQTLs as instrumental variables. All these methods, however, suffer from various issues, particularly false positive findings. For example, TWAS may find associations in non-causal genes when their eQTLs are shared with nearby causal genes or are in linkage disequilibrium (LD) with nearby causal variants.
The fundamental problem of existing methods is that, when assessing the role of one gene on the phenotype, nearby variants and nearby genetic components of expression can be correlated with the eQTL(s) of the test gene while also affecting the phenotype directly, thus acting as "genetic confounders". This motivates our approach, causal TWAS (cTWAS), which jointly models the effects of all nearby gene expression and nearby variants, which controls for potential genetic confounders. To make the model identifiable, we use a Bayesian variable selection strategy that learns parameters from genome-wide data.
In simulation, cTWAS produced calibrated false positive (FP) rates while existing methods all suffered from high FP rates. We applied cTWAS to GWAS summary statistics of several commonly studied traits, such as low density lipoprotein (LDL) cholesterol and Systolic blood pressure using eQTL data from GTEx. Genes identified by cTWAS are enriched with known risk genes, as well as relevant gene sets, and include novel genes and pathways, e.g. activin receptor signaling in LDL.
We have also extended cTWAS to perform joint analysis of multiple types of molecular QTLs and multi-tissue eQTLs. This extension greatly improves the power of detecting putative causal genes. By using both eQTL and splicing QTL (sQTL) data from liver in the analysis of LDL, we double the number of genes found, compared with using eQTL alone. In the case of multi-tissue eQTL analysis, by adding adipose as a secondary tissue to liver, the number of candidate genes of LDL increased by 50%, compared to using liver eQTL alone. These results thus demonstrate the great potential of gene discovery using the extended cTWAS.
In conclusion, cTWAS solves a fundamental challenge of integrative QTL-GWAS analysis, greatly reducing false positives. It also provides a framework for incorporating multi-tissue and multi-molecular traits data. cTWAS is available in easy-to-use R package.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4050 Aggregated best subset selection using summary statistics for polygenic risk prediction.

Authors:

T. Chen[1], H. Zhang[2], X. Lin[1], R. Mazumder[3]; [1]Harvard T.H. Chan Sch. of Publ. Hlth., Boston, MA, [2]Natl. Cancer Inst., Cambridge, MA, [3]Massachusetts Inst. of Technology, Cambridge, MA

Background: Polygenic risk scores (PRS) enhance population risk stratification and advance personalized medicine, with wide ranging applications from cancer to psychiatric disorders. However, existing methods face a tradeoff between predictive power and computational efficiency.
Method: We introduce ALL-Sum, an innovative approach that combines rapid L0Learn-based optimization, known for selecting sparse models, with an aggregation step to improve PRS prediction across varying penalty parameters. By combining estimates from multiple parameterizations, we can efficiently construct a highly accurate PRS.
Results: In extensive large-scale simulations with diverse polygenicity and genome-wide association studies (GWAS) sample sizes, ALL-Sum consistently outperforms popular alternative PRS methods, with up to 10% improvement in out-of-sample prediction accuracy. Our method further shows tremendous efficiency gains in runtime and memory usage compared to Bayesian computation. We further compare all methods on 27 published GWAS summary statistics for 11 complex traits from 9 reputable data sources, such as the Breast Cancer Association Consortium and Global Lipids Genetics Consortium, evaluated using individual-level data from the UK Biobank. In our analysis of published GWAS data, ALL-Sum typically achieves the highest or second-highest prediction accuracy, with particularly noticeable performance gains using GWAS with larger sample sizes.
Conclusion: With better prediction, faster computation, and easier implementation, ALL-Sum stands as a promising clinic tool and a foundation for future developments in PRS methodology. We provide ALL-Sum as a user-friendly command-line software with pre-computed reference data for streamlined user-end analysis.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4051 Allelic series-based analyses of rare variants are enhanced by sequence models.

Authors:

K. Fletez-Brant, T. Bhangale; Genentech, Inc, South San Francisco, CA

Genome-wide association studies on common variants have yielded thousands of associations providing novel insights into biological processes driving common diseases. By contrast, despite large-scale sequencing-based datasets enabling gene-level rare variant (RV) aggregation based analyses, RV studies have not been as successful. Two major challenges in RV analyses are (i) determining which variants in a gene actually affect the phenotype i.e. large proportion of variants do not directly impact phenotypes and (ii) determining the effect size of causal variants. Recent advances can potentially assist in addressing these: (i). applications of deep-learning models for variant effect prediction [e.g. EVE, ESM1b] substantially outperform traditional approaches, and may help identify relevant rare variants in a gene (ii). application of allelic-series models to variants in a gene can outperform traditional approaches for RV-aggregation analyses. We propose a Bayesian latent variable model that (i) assigns a posterior inclusion probability (PIP) to each variant and (ii) estimates an average effect of all variants tested. We leverage variant effect predictions from a recent protein sequence model, ESM1b, as prior information on variant PIP. Our test of average effect is implicitly a test on the presence of an allelic series, where the variant effect on the phenotype is a function of the prediction from ESM1b. We operationalize this model by extending the method of Logsdon et al, and deploy on all rare (MAF < 1%) protein coding variants genome-wide in cohorts of Parkinson's disease (PD) patients from both 23andMe and the UK Biobank. We recover the known relationship between rs76763715 (a rare coding variant in GBA; PIP = 91%) and PD, as well as the known lack of relationship to rs75548401 (also in GBA; PIP = 39%). We find the average variant effect in GBA to be an odds ratio (OR) of 13.9, with 4 variants with PIP > 0.6. Additionally, we find an association with POU3F3 (OR = 12.4), through a deletion at chr2:104856391 (PIP = 84%); POU3F3 has previously been described to have a role in PD pathogenesis via its function in autophagic-lysosomal system. Our approach accommodates multiple sources of prior information, and can be extended to include gene level features such as constraint scores (e.g. pLI) that affect the power to detect gene associations or variant level features such as MAF or protein structure based spatial information from Alphafold2 predictions. Through the integration of machine learning predictions into rare variant analysis, this work meaningfully extends our ability to interrogate rare coding variation to both effect size estimation and variant inclusion.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4052 Alzheimer's disease and herpes simplex virus type 1 - gene-viral interactions

Authors:

K. Lopatko Lindman[1], B. Weidung[2], J. Olsson[1], S. Eriksson[1], G. Hallmans[1], F. Elgh[1], H. Lövheim[1]; [1]Umeå Univ., Umeå, Sweden, [2]Uppsala Univ., Uppsala, Sweden

Background: Alzheimer’s disease (AD) risk is influenced by a combination of genetic and environmental factors. Herpes simplex virus type 1 (HSV-1), a virus capable of infecting the central nervous system, represent an environmental factor that has been implicated in certain forms of AD and especially in individuals carrying the APOE4 allele [1] Here, we aimed to explore the interactions between genes and HSV-1 in the context of AD. To achieve this, we created a polygenic risk score (PRS) for AD and evaluated its association with HSV-1 in relation to AD risk. Methods: PRS quantifies the aggregate risk of genetic variants. This approach can reveal interactions with environmental exposures such as HSV-1, by linking the environmental exposure to genetic predisposition of disease severity or susceptibility. The AD PRS was calculated as the sum of effect alleles weighted by their log odds, using summary statistics from a genome-wide association study of AD [2]. In the case-control sample, the participants underwent serological testing for anti-HSV-1 IgG. Subjects were classified into genetic AD risk tertiles: low (lowest score, 1/3), intermediate (middle score, 1/3) and high (highest score, 1/3) based on their PRS. All analyses were adjusted for population structure by inclusion of the first five principal components.Results: A total of 326 cases and 1022 sex- and age-matched controls passed quality control. The mean age at the time of the event was 71 years, and 74 % of the participants were female. On average, the blood samples were drawn 9.7 years before the AD diagnosis date. The odds ratio (OR) for AD was 1.8 per 1 standard deviation increase in the PRS (95 % confidence interval [CI] 1.54 - 2.01, $p < 0.001$). We found that anti-HSV-1 IgG was associated with an increased risk of AD (OR 2.3 per HSV-1 seropositivity, 95 % CI 1.47 - 3.52, $p < 0.001$). This association was stronger in individuals having high and intermediate genetic AD risk (OR 2.8, 95 % 1.15-6.73, $p = 0.023$ and OR 3.2, 95 % CI 1.04-9.70, $p = 0.042$), compared to those with low genetic AD risk (OR 1.7, 95 % CI 0.28-10.75, $p = 0.55$). No interaction was observed between the PRS and anti-HSV-1 IgG seropositivity ($p = 0.30$). Conclusion: The association between HSV-1 infection and AD appears to be greater in individuals with intermediate and high genetic AD risk, indicating that HSV-1 infection might play a role in those with a genetic predisposition to AD.1. Linard, et al., Interaction between APOE4 and herpes simplex virus type 1 in Alzheimer's disease. Alzheimers Dement, 2020. 16(1): p. 200-208.2. Bellenguez, et al., New insights into the genetic etiology of Alzheimer's disease and related dementias. Nat Genet, 2022. 54(4): p. 412-436.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4053 An accurate machine learning model trained on clinically curated variants to predict the variant pathogenicity score

Authors:

R. Gupta; MedGenome Labs Ltd., Bangalore, India

Whole genome and exome sequencing data generates several thousands of variants including many rare variants. Identifying and prioritizing the disease-causing variants for diagnosis of rare disorders in clinical labs is a manual and time-consuming process. Ranking the rare variants based on pathogenicity will help in speeding up the report generation and improve the accuracy and consistency of clinical reporting. Models trained on variants obtained from public databases such as ClinVar have not been so successful because of the presence of noisy data. Recent studies have shown that the model built on variants identified from clinical labs in patients would perform better.
On a similar concept we have developed a machine learning model trained on manually curated and reviewed disease-causing variants from a cohort of 100,000 clinical samples. This covers more than 4,000 different symptoms and several diseases. We have also devised a unique approach to select a list of benign variants which are rare in nature and share similar variant type and genes with that of pathogenic variants to avoid feature bias and overfitting. We studied more than 200 variant features derived from sequence, population database frequencies, in-silico predictions, disease databases and other features. After performing the feature selection and ranking, we identified 33 features that suited best for scoring of the variants.
We built a random forest model and trained on 44,978 variants (9,009 disease-causing, 35,969 benign) covering 10,000 genes. Our model obtained F1 and sensitivity value of 0.98 on 15,000 unseen variants that covers 6,000 genes. We further evaluated our model on 166 cases which were diagnosed and resolved through manual interpretation in our laboratory. These cases were obtained from different diseases including metabolic (28%), nervous system diseases (22%), hereditary cancer (16%), development disorders (11%). Our model ranked the manually selected variant in Top20 for 89% of cases and after applying the patient phenotype filtering our model ranked the prioritized variants in Top20 for 95% of cases. We also compared the VaRTK model on a recent study published variants and could classify 98% of the pathogenic variant correctly. We believe that the VaRTK model will help the clinical laboratories in prioritizing variants quickly and accurately.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4054 An association study on adolescent cases of BNT162b2 vaccine-induced myocarditis in Hong Kong.

Authors:

L. She, H-W. Tsang, S. Tsao, X. Yang, G. Chua, S. Chan, W. Yang, M. Kwan, P. Ip; The Univ. of Hong Kong, Hong Kong, Hong Kong

Vaccine-induced myocarditis is a rare complication of BNT162b2 vaccine and is more commonly reported in male adolescents in Hong Kong [1]. Although the clinical presentation of mRNA vaccine-induced myocarditis was well documented [2-5], its underlying mechanism and whether genetics played a role remained unclear. In this study, we performed an association analysis on a cohort of 43 Hong Kong Chinese diagnosed with vaccine-induced acute myocarditis after administration of BNT162b2 mRNA vaccine (Fosun Pharma-BioNTech; BNT162b2 vaccine) from July 2021 to June 2022 and 481 controls. Whole-genome sequencing data was acquired for all individuals and a clustering approach (ClusterAnalyzer) was developed in-house to extract genomic regions with significant association signals that are also supported by multiple SNPs in linkage disequilibrium, to more or less overcome the false positive issue caused by the small sample size. We provided summary statistics of SNPs in 2,156 selected regions based on the strength of the association signals and analysis by ClusterAnalyzer. We illustrated in 3 regions (chr4:81,892,037-81,965,272; chr4:86,019,317-86,219,034; chr11:40473809-40585162) that true signals are enriched with reduced noise in the selected regions, as variants associated with COVID severity were found in each of the 3 regions (rs56038305, rs2575679 and rs1001952 respectively). Our results provide useful information on elucidating the potential genetic contribution on mRNA vaccine-induced myocarditis and valuable data for future replications.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4055 An atlas of gene regulatory variants mediating the association between brain morphology and obesity.

Authors:

D. Fusco[1], L. Troiani[1], M. Noè[1], F. Pizzagalli[1], D. Marnetto[2], P. Provero[1]; [1]Università di Torino, Torino, Italy, [2]Univ. of Turin, Torino, Italy

Given the support for a role of the central nervous system in susceptibility to obesity, several studies examined the correlation between obesity, often quantified as body mass index (BMI) and brain structural features. While there is evidence that increased BMI is associated with reduced gray matter volume, the related neurobiological mechanisms have yet to be characterized. In addition it is not clear what are the specific brain regions whose morphology is associated with BMI, whether such associations subtend causal relationships, and which are the genes mediating their genetic relationships at the molecular level.

We analyzed genome wide association studies (GWAS) data on structural brain magnetic resonance imaging (sMRI) from 33 thousands UK BioBank samples. By using LD score regression we estimated the genetic correlation with BMI of 437 sMRI-derived traits, including multiple cortical measurements for Desikan-Killiany parcellations and subcortical volumes and intensities.

We then performed three-way colocalization analysis for BMI- and sMRI-associated loci, adding as a third trait the expression in brain tissues of the genes located near the relevant locus, based on the GTEx dataset.

Overall, the genetic correlation analysis showed significant, typically negative correlations between brain sMRI phenotypes (such as those related to cortical area and volume, and subcortical volume) and BMI, in agreement with previous studies, suggesting that genetic factors associated with an increased BMI are also related to reduced volume of brain structures.

Moreover, we identified several loci for which the SNP putatively causal for both sMRI traits and BMI is also affecting the expression of a gene in brain tissues, thus providing a potential molecular mechanism that explains the common genetic basis of the two traits and might suggest new preventive and therapeutic strategies.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4056 An augmented reference panel with 22,134 haplotypes enhances GWAS analysis and the rare variants detecting power of hip bone mass density.

Authors:

M. Yang, J. ZHONG, H. Zheng; Westlake Univ., HangZhou, China

The Westlake BioBank for Chinese (WBBC) cohort has established a population-specific reference panel and an online imputation server (https://wbbc.westlake.edu.cn/), promising significant improvements in imputation performance for the Chinese population. In an effort to enhance genome-wide association study (GWAS) analyses and the detection of rare variants, we integrated the WGS data from multiple sources, including SG10K project (13.7×, 4,563 samples), GenomeAsia (GAsP) project (36×, 1,031 samples), WBBC pilot (13.9×, 4,480 samples), and 1000 Genomes Project (1KGP) phase 3 (30×, 993 East and South Asian from 3,202 samples) to create a combined panel, boasting the highest diversity of populations within South and East Asia. The no-singleton version of this combined panel includes 22,134 haplotypes and 80,367,720 variants. For imputation in Human Genome Diversity Project (HGDP) populations, the combined reference panel consistently exhibited the highest average non-reference (NR) allele concordance rate for East Asian populations compared with 1KGP panel. When imputing the Chinese population data (5,679 samples), the combined reference panel successfully imputed a total of 7,976,019 well-imputed sites ($R^2 > 0.8$), surpassing the SG10K panel, which ranks second with 6,725,695 sites. Across nine MAF bins for rare and low-frequency variants, the combined reference panel demonstrates the highest number of well-imputed variants compared with SG10K, WBBC, 1KG and GAsP panel. Moreover, within 12,923 Asian GWAS catalog association sites, the combined panel exhibits the highest number of well-imputed variants, amounting to 4,424. Its main advantage over the other four panels lies in the imputation of low-frequency and rare variants, encompassing a total of 356 low-frequency variants and 133 rare variants. Finally, we applied two hip bone mineral density (BMD) cohorts from the WBBC to the augmented reference panel to conduct a discovery (n = 2,322) and replication (n = 3,037) association study. In downstream analysis, including meta-analysis and fine-mapping, we identified three BMD-related genome wide significant (GWS [$P < 5 \times 10^{-8}$]) signals, which were annotated to one novel gene *SNTG1* (rs60600379, MAF = 0.009: $P_{discovery} = 1.15 \times 10^{-3}$, $P_{replication} = 2.26 \times 10^{-6}$, $P_{meta} = 4.77 \times 10^{-8}$) and two known genes (*FMN2* and *SOX4*). In summary, we generated a reference panel for the South and East Asian populations, significantly improving the imputation performance for rare and low-frequency variants, thereby enhancing the utility and precision of GWAS analysis.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4057 An ensemble machine learning method for phenotypic refinement from binary symptom data.

Authors:

R. Danning, X. Lin; Harvard T.H. Chan Sch Pub Hlth, Boston, MA

The clinical heterogeneity of complex neuropsychiatric disorders presents a significant challenge for gene discovery. In the absence of consistent biomarkers, many neuropsychiatric conditions are diagnosed based on broad symptomatic criteria, leading to noisy phenotype definitions. The ability to refine these phenotypes is thus critical to the unmasking of genetic associations. Autism spectrum disorder is one such potentially noisy phenotype; despite evidence of heritability, its genetic etiology remains largely unknown. Given the diversity of behavioral presentations of autism, it is possible that the autism phenotype comprises several latent subtypes, each of which may have a distinct genetic profile. However, statistical methods for phenotypic subtype detection in autism and other complex neuropsychiatric conditions are hindered by the prevalence of binary symptom data representing the presence or absence of a specific behavior. Existing methods for clustering binary data have various drawbacks, including the necessity of pre-specifying the number of subtypes, bias in the presence of extraneous symptoms, and the assumption that there is no correlation of responses within an individual. We have developed an ensemble machine learning method for detecting subtypes from binary symptom data that does not require prespecifying the number of subtypes, is robust to the inclusion of extraneous symptoms, and is likewise robust to the violation of conditional independence of symptoms within individuals. The method ensembles latent class analysis, a model-based clustering method; principal components analysis, a mathematical dimensionality reduction algorithm; and UMAP, a cutting-edge model-free dimensionality reduction algorithm. In simulation studies, this method outperforms gold-standard techniques across a variety of scenarios, including in the presence of local symptom correlation structure within individuals, as well as when extraneous symptoms that are unrelated to the phenotype are included in analysis. We apply this method to the parent-report Social Communication Questionnaire data from the Simons Foundation's SPARK study of autistic individuals and find three distinct subtypes of communication skills and social functioning. We propose methods for synthesizing subtypes across the various diagnostic dimensions of complex disorders, address the potential impact of phenotype refinement on gene discovery for autism, and discuss the broader implications for psychiatric genetics and precision medicine.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4058 An individualized Bayesian inference framework for detecting genomic variantsandtheir interactions for complex diseases.

Authors:

J. Liu[1], M. Rahman[1], C. Cai[2], N. Bo[2], D. McNamara[2], Y. Ding[2], G. F. Cooper[2], X. Lu[2]; [1]Missouri Univ. of Sci. and Technology, Rolla, MO, [2]Univ. of Pittsburgh, Pittsburgh, PA

Population-based genome-wide association studies (GWAS) have been successful in discovering genomic variants associated with diseases, but they may not capture all factors that contribute to disease risk of individual subjects. In addition, it is challenging for GWAS to detect with sufficient power the association of rare or low-frequency genomic variants without a very large sample. As a complement to GWAS, we propose an individualized Bayesian inference (IBI) algorithm that can estimate genomic variants underlying complex traits such as hypertension at the level of an individual (e.g., a patient) and their "patients-like-me" subgroup. By modeling at the individual and their subgroup levels, IBI aims to identify genomic variants that provide a plausible explanation of the phenotype observed in each individual. We applied IBI to the whole-genome genotyping and blood pressure data from the Framingham Heart Study (FHS). Our results support IBI as a promising approach for complementing GWAS, especially in detecting low-frequency genomic variants as well as learning personalized genomic variants of traits and diseases to help advance personalized medicine. To tackle the challenge of understanding the interactions among variants or genes in complex diseases, we further extended IBI to develop an IBI Decision Tree (IBI-DT) learning algorithm, which allows for the detection of variant-variant or gene-gene interactions that act disjunctively or conjunctively to influence a trait of interest. IBI-DT builds subpopulations with distinct genomic heterogeneity using a tree and derives statistics across these subpopulations to reveal the influence of single variants/genes or variant/gene interactions. We applied IBI-DT to the cancer domain by utilizing The Cancer Genome Atlas data across 16 cancer types. The results demonstrate that IBI-DT not only outperforms expression quantitative trait loci (eQTL) analysis in identifying well-known cancer drivers responsible for differential gene expression in cancer patients but also uncovers biologically significant interactions among these cancer drivers. Our results provide support that IBI-DT has the potential to help identify signaling networks underlying cancer and other diseases and could be a valuable tool in advancing precision medicine.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4059 An integrative framework mapping the functional effects of rare noncoding variation in autism spectrum disorder leveraging high-resolution enhancer annotation

Authors:

Y. Zhang[1], J. Tian[2], A. Sun[1], Y. Jin[1], S. R. Shah[1], M. I. Paramo[1], S. Dong[3], L. Yao[1], V. Fangal[1], S. Sanders[3], K. Roeder[2], B. Devlin[4], H. Yu[1]; [1]Cornell Univ., Ithaca, NY, [2]Carnegie Mellon Univ., Pittsburgh, PA, [3]Univ. of California San Francisco, San Francisco, CA, [4]Univ. of Pittsburgh, Pittsburgh, PA

Damaging *de novo* coding variants are found in 10-20% of individuals diagnosed with autism spectrum disorder (ASD), and these variants are assumed to confer substantial liability. Inherited rare coding variation also plays a role, although how much is difficult to determine. It is reasonable to speculate that rare noncoding variants, both *de novo* and inherited, also play some role, although identifying their contribution is challenging due to insufficient functional annotation of the noncoding genome. In this study, we analyzed rare noncoding variants derived from whole-genome sequencing data of a cohort comprising 17,882 individuals, alongside rare coding variations sourced from a cohort of 63,237 individuals. We employed 20,736 high-resolution enhancer elements defined by gene-distal transcription events in neuronal progenitor cells to prioritize functional rare noncoding variants. Subsequently, we linked the prioritized noncoding variants to the target genes and developed a gene-based statistical model to integrate coding and noncoding signals. This model offers insight into the relative effect sizes of rare variants in coding and noncoding regions. In theory it should have better power to identify genes associated with ASD relative to using rare coding variants only, and our analyses support this conjecture. Next, we used an integrative framework that combines gene-specific test statistics, rare coding variation affecting protein-protein interactions (PPIs), as revealed by the presence of damaging *de novo* missense variants on the 3D structural interfaces, and a PPI network to delve further into ASD etiology. This holistic approach aids in enhancing the discovery of ASD genes as well as identifying relevant dysregulated cellular subnetworks. Overall, our study explores the potential of utilizing high-resolution enhancer annotation strategies to prioritize functional rare noncoding variations, and strives for effectively integrating this information with multiple complementary data sources to gain a more comprehensive understanding of ASD.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4060 Analysis of vitamin D receptor-binding variants in susceptibility to pediatric-onset multiple sclerosis.

Authors:

D. Yilmaz[1,2], C. Adams[1,2], M. Horton[3], J. Graves[4,5], H. Quach[1,2], D. Quach[1,2], G. Aen[6], T. Lotze[7], S. Mar[8], J. Ness[9], Y. Wheeler[9], M. Gorman[10], B. Weinstock-Guttman[11], A. Waldman[12], T. Schreiner[13], M. Rodriguez[14], J-M. Tillema[14], T. Chitnis[15], L. Krupp[16], J. Rose[17], C. Casper[18], M. Rensel[19], J. Hart[20], E. Waubant[5], L. Barcellos[1,2], The Network of Pediatric Multiple Sclerosis Centers; [1]Univ. of California, Berkeley, Berkeley, CA, [2]Ctr. for Computational Biology, Div. of Computing, Data Sci. and Society, Univ. of California, Berkeley, Berkeley, CA, [3]NIH, Bethesda, MD, [4]Dept. of NeuroSci.s, Sch. of Med., Univ. of California, San Diego, San Diego, CA, [5]Dept. of Neurology, Univ. of California, San Francisco, San Francisco, CA, [6]Pediatric MS Ctr., Loma Linda Univ. Children's Hosp., Loma Linda, CA, [7]Texas Children's Hosp., Houston, TX, [8]Pediatric-onset Demyelinating Diseases and Autoimmune Encephalitis Ctr., St. Louis Children's Hosp., Washington Univ. Sch. of Med., St. Louis, MO, [9]UAB Ctr. for Pediatric Onset Demyelinating Disease, Univ. of Alabama, Birmingham, Birmingham, AL, [10]Dept. of Neurology, Boston Children's Hosp., Boston, MA, [11]Pediatric Multiple Sclerosis Ctr., Jacobs Neurological Inst., SUNY Buffalo, Buffalo, NY, [12]Div. of Neurology, Children's Hosp. of Philadelphia, Philadelphia, PA, [13]Dept.s of Pediatrics and Neurology, Univ. of Colorado, Aurora, Aurora, CO, [14]Mayo Clinic's Pediatric Multiple Sclerosis Ctr., Rochester, MN, [15]Partners Pediatric Multiple Sclerosis Ctr., Massachusetts Gen. Hosp. for Children, Boston, MA, [16]Lourie Ctr. for Pediatric Multiple Sclerosis, Stony Brook Children's Hosp., Stony Brook, NY, [17]Dept. of Neurology, Univ. of Utah Sch. of Med., Salt Lake City, UT, [18]Dept. of Pediatrics, Univ. of Utah Sch. of Med., Salt Lake City, UT, [19]Mellen Ctr., Cleveland Clinic, Cleveland, OH, [20]Regional Pediatric MS Ctr., Neurology, Univ. of California, San Francisco, San Francisco, CA

The genetic basis of multiple sclerosis (MS) in adults has been extensively studied and more than 230 susceptibility variants have been identified to date through GWAS; less is known about genetic contributions to MS occurring in children (<18 years, 'pediatric-onset' MS) which comprises ~5% of patients. MS prevalence is higher in regions farther from the equator, supporting the hypothesis that vitamin D exposure has a protective effect on MS risk. Mendelian randomization studies have identified causal associations between lower serum vitamin D and increased risk of MS. Vitamin D is important for many biological processes. After being ingested or absorbed, serum vitamin D is first converted to 25-hydroxyvitamin D [25(OH)D], its more stable form, and then to 1,25-dihydroxy vitamin D [1,25(OH)2D]. Previous studies have established that 25(OH)D signals through the nuclear vitamin D receptor (VDR), a ligand-regulated transcription factor that modulates vitamin D-regulated gene expression. Directly testing for associations between VDR binding and disease phenotypes in large-scale human studies poses many challenges. SNPs associated with genetic variation in VDR binding affinity (VDR binding variants or 'VDR-BVs') have been recently identified using ChIP-exo data from calcitriol-stimulated lymphoblastoid cell lines. We recently studied these VDR-BVs as genetic instrumental variables in adult-onset MS and identified strong evidence for association between several VDR-BVs and MS. The objective of this study was to study adult-onset MS VDR-BVs candidates for a role in pediatric-onset MS. We utilized a multiethnic pediatric cohort comprised of 1,303 pedMS cases and 984 controls from the U.S. Network of Pediatric MS Centers. Genotyping was performed using Illumina Infinium 660K OmniExpress and OmniExpressExome BeadChip arrays and imputed against reference haplotypes from the 1000 Genomes Project Phase 3 using IMPUTE4. After quality control measures, VDR-BVs were excluded if minor allele frequency (MAF) <5%. This resulted in 17 candidate VDR-BVs for analysis. Associations between VDR-BVs and pediatric-onset MS were tested using logistic regression in *PLINK* and adjusted for the first six genetic principal components. In the full multiethnic cohort, three VDR-BVs were associated with pediatric-onset MS at $p<0.05$: rs10995246 (OR: 0.86, 95% CI: 0.75-0.99), rs11729497 (OR: 1.15, 95% CI: 1.01-1.30), and rs2531804 (OR: 1.25, 95% CI: 1.10-1.42). rs2531804 is located on chromosome 6 and is upstream of *TEC*. These findings provide further evidence that VDR-BVs associated with adult-onset MS are also implicated in pediatric-onset MS.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4061 *APOE4*-stratified GWAS of multiple cognitive domains in non-Hispanic white and non-Hispanic black older adults

Authors:

A. Contreras[1], S. Walters[2], J. M. Eissman[1], A. N. Smith[2], S. Mukherjee[3], M. Lee[3], S-E. Choi[3], P. Scollard[3], E. Trittschuh[3], J. B. Mez[4], W. S. Bush[5], B. W. Kunkle[6], A. C. Naj[7], K. A. Gifford[2], M. Bilgel[8], A. B. Kuzma[7], The Alzheimer's Disease Neuroimaging Initiative, Alzheimer's Disease Genetics Consortium, The Alzheimer's Disease Sequencing Project, M. L. Cuccaro[9], C. Cruchaga[10], M. A. Pericak-Vance[9], L. A. Farrer[4], L-S. Wang[7], G. D. Schellenberg[7], R. P. Mayeux[11], J. L. Haines[5], A. L. Jefferson[2], W. A. Kukull[3], C. Keene[3], A. J. Saykin[12], P. M. Thompson[13], E. R. Martin[9], M. Albert[14], S. C. Johnson[15], C. D. Engelman[16], L. Ferrucci[17], D. A. Bennett[18], L. L. Barnes[18], J. A. Schneider[18], S. M. Resnick[19], R. A. Sperling[20], P. K. Crane[3], L. Dumitrescu[2], T. J. Hohman[2]; [1]Vanderbilt Univ., Nashville, TN, [2]Vanderbilt Univ. Med. Ctr., Nashville, TN, [3]Univ. of Washington, Seattle, WA, [4]Boston Univ., Boston, MA, [5]Case Western Reserve Univ., Cleveland, OH, [6]Univ. of Miami, Miami, FL, [7]Univ. of Pennsylvania, Philadelphia, PA, [8]NIH, Bethesda, MD, [9]Univ. of Miami Miller Sch. of Med., Miami, FL, [10]Washington Univ., Saint Louis, MO, [11]Columbia Univ., New York, NY, [12]Indiana Univ. Sch. of Med., Indianapolis, IN, [13]Univ. of Southern California, Los Angeles, CA, [14]Johns Hopkins Univ. Sch. of Med., Baltimore, MD, [15]Univ. of Wisconsin Sch. of Med. and Publ. Hlth., Madison, WI, [16]Univ. of Wisconsin, Madison, WI, [17]Johns Hopkins Univ. Bloomberg Sch. of Publ. Hlth., Baltimore, MD, [18]Rush Univ., Chicago, IL, [19]Lab. of Behavioral NeuroSci. Natl. Inst. on Aging NIH, Baltimore, MD, [20]Harvard Med. Sch., Boston, MA

Apolipoprotein E4 (*APOE*-ε4) is the most influential common genetic risk factor for late-onset Alzheimer's disease (AD). However, the impact of *APOE*-ε4 carrier status on the genetic architecture of cognitive decline remains unclear. Here, we sought to uncover genetic associations with cognitive decline via stratified genome wide-association analyses (GWAS), thus exploring differences in *APOE*-ε4 carriers and non-carriers, as well as across two racial/ethnic groups, non-Hispanic Whites (NHW) and non-Hispanic Blacks (NHB). We leveraged a harmonized cognitive dataset from 10 aging and AD cohorts, encompassing a large multi-ancestry population (Ntotal= 36,483, NHW_ε4pos = 12,047, NHW_ε4neg = 20,253, NHB_ε4pos = 1,810, NHB_ ε4neg = 2,373). The mean age at baseline was 73 years, with 58% females and 65% cognitively normal. Adjusting for age, sex, and the first 5 principal components of genetic ancestry, GWAS were performed for each neuropsychological domain (memory, executive function, and language) at baseline and for longitudinal decline. Each GWAS was conducted in a cohort-specific manner within each ancestry group by *APOE*-ε4 status, followed by fixed-effect meta-analysis. We saw 14 novel associations among E4 carriers including variants near *GRIN3A* and *NME7* genes among NHW and variants near the *WDPCP* gene among NHB participants. In contrast, we saw 10 associations among E4 non-carriers including variants near *LOC101927668* gene in NHW and *GALNT7* genes among NHB. Interestingly, eQTL evidence for the association on chromosome 9 among NHW E4 carriers/non-carriers implicates *GRIN3A* gene. Sensitivity analyses, excluding comorbidities and subsetting individuals aged 60 and above, were performed to confirm our findings. These preliminary findings have potential implications for precision medicine approaches targeting cognitive impairment and AD prevention. Future work will extend these preliminary findings to incorporate cross-ancestry analysis, gene-based tests, and genetic correlation.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4062 † Application of network-based heterogeneity clustering for investigation of genotype-phenotype correlations in BioMe BioBank

Authors:

M. Kars, Y. Wu, C. Bayrak, B. Gelb, Y. Itan; Icahn Sch. of Med. at Mount Sinai, New York, NY

The genetic basis for many human diseases often exhibits heterogeneity, resulting in different genes in the same or related biological pathways being responsible for the same or similar phenotypes. Conventional gene burden methods, which are used to identify genetic signals in case-control studies, lack the power to detect these signals in the presence of genetic heterogeneity and small cohorts. To address this issue, we have previously developed a computational method called network-based heterogeneity clustering (NHC) that can detect physiological homogeneity within genetically heterogeneous cohorts with small sample sizes, and demonstrated that our method can effectively converge genes that are biologically related on a protein-protein interaction network and accurately identify gene clusters with potentially deleterious rare variants. Here, we utilized NHC to analyze three disease cohorts from the Mount Sinai BioMe BioBank, where previous gene burden approaches were ineffective in identifying candidate genes. The cohorts comprised 50 individuals with lactation disorders (agalactia or hypogalactia) and 500 controls, 71 patients with severe COVID-19 and 390 controls with mild-moderate disease, and 431 children with food allergy and 2,155 controls, which was also analyzed in a population-specific manner using data from individuals of Ashkenazi Jewish (AJ, n = 826) and non-AJ European (n = 1,760) ancestries. NHC analysis in the lactation disorders cohort identified gene clusters involved in intra-golgi vesicle transport, mammary stem cell differentiation and proliferation, and trans-differentiation of white adipocytes in the mammary gland during lactation. NHC also uncovered two gene clusters potentially related to increased risk for severe COVID-19, including genes that preserve endothelial barrier function and genes involved in snRNA processing, which have previously been implicated in susceptibility to HSV-1 infection. Furthermore, NHC facilitated the identification of candidate gene clusters in the food allergy cohort with population-specific signatures, including genes proposed to be associated with allergic asthma and food allergies, such as *DYNC1H1*, toll-like receptor encoding genes, genes involved in IL-17 signaling and cell adhesion, and genes encoding mitochondrial ribosomal proteins. These findings indicate that NHC is a useful approach for uncovering candidate genes in disease groups with genetic heterogeneity, especially in smaller population-specific subgroups, outperforming traditional case-control studies for such cohorts.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4063 Applying local genetic covariance of functional annotations to increase the power of multi-trait association test to identify pleiotropic loci

Authors:

W. Choi[1], C. Lee[2], B. Han[3]; [1]Seoul Natl. Univ. Graduate Sch., SEOUL, Korea, Republic of, [2]Columbia Univ. Med. Ctr., Fort Lee, NJ, [3]Seoul Natl. Univ. Coll. of Med., Seoul, Korea, Republic of

The identification of pleiotropic loci is important because the interpretation of them can provide us with valuable information to understand the common etiology of diseases or complex traits. Our previous method, PLEIO, effectively identifies pleiotropic loci by accounting for genetic architecture information, such as genetic covariance and heritability. However, PLEIO utilizes global genetic architecture information. If the genetic covariance differs by genomic region, accounting for the region-specific genetic covariance may improve the results. It is well known that some sub-genomic regions (a subset of SNPs belonging to an annotation group) may exhibit different genetic covariance compared to the global covariance in terms of direction and magnitude. If this is the case, applying the sub-genomic regional genetic covariance to PLEIO can increase the power. To verify how much the power can increase theoretically, we first simulated extreme situations where sub-genomic regions have significantly different genetic covariances among themselves. In these simulations, as expected, we observed a clear improvement in the performance of Stratified-PLEIO (PLEIO applied to the SNPs in the annotation region with the annotation-specific genetic covariance). For example, given 3 disjoint regions with largely disparate local correlations, S-PLEIO identified 423 significant signals out of 3,000 causal markers while PLEIO identified no signals at all. Next, we wanted to examine whether S-PLEIO will outperform PLEIO in real datasets, similar to the simulations. We used the real summary data of the 18 cardiovascular disease phenotypes and 33 annotations from the baseline-LD model(v2.2, 1000 Genome phase3) of S-LDSC. For simplicity, we focused on each annotation separately. For each annotation, we partitioned the whole genome (global region) into the annotation region and non-annotation region (the complement of the annotation region). We compared the genetic covariances of the annotation and non-annotation regions and chose the top 3 annotations that showed the most significant difference. For each of the top 3 annotations, we ran PLEIO and S-PLEIO and compared the results of the two methods, limiting our scope to the SNPs belonging to the annotation. Contrary to the expectation, S-PLEIO did not outperform PLEIO even for these top 3 annotations. The majority of the association p-values obtained with S-PLEIO were less significant than those obtained with PLEIO across all 3 annotations. Therefore, we suspect that the difference in true genetic covariance among annotations may not be large enough to confer higher performance of S-PLEIO for association mapping.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4064 Artificial Intelligence and Geographic Analysis of Clinical Genetic Data in California's Central Valley

Authors:

J. Woods; Valley Children's Hosp., Madera, CA

Introduction
Clinical genetic data is often unstructured and rarely applied outside of the context of an individual patient encounter. Therefore, there is a need for the development of data science methodologies to ascertain and analyze this real-world data. Comprehensive utilization of clinical genomic data can be leveraged to drive clinical decision-making and precision medicine at minimal cost to healthcare organizations.
Methods
Germline variant allele data were obtained over a five-year period from all genetic testing sent to reference laboratories from Valley Children's Hospital, which has a catchment area of 1.4 million patients. Data were obtained from integrated testing results in the electronic health record as well as directly from testing laboratories, as available. The variants were automatically reinterpreted utilizing the Franklin© Artificial Intelligence (AI) to flag cases for reinterpretation/reconsideration to drive appropriate follow-up. Genetic disorder counts were inferred from the data by sorting variants by gene, inheritance pattern, pathogenicity interpretation, sex, and zygosity. Utilizing PowerBI©, the variant data were matched to patient demographic data in the Electronic Health Record (EHR) to generate a data table and map variants as a choropleth at the resolution of Zip Code Tabulation Area (ZCTA).
Results
Three-thousand-sixty-five variants were identified and 98% were matched to patients in the EHR. The Franklin© AI changed the interpretation for 27% of variants. The top half of variant reinterpretations were Variant of Uncertain Significance (VUS) > Likely Benign (114 or 21.7%), VUS > Likely Pathogenic (110 or 20.9%) and Pathogenic > Likely Pathogenic (98 or 18.6%). A total of 723 genetic disorders were identified with 176 structural variant disorders, 543 monogenic disorders, three methylation disorders and one repeat expansion disorder. Mapping of variants demonstrated geographic hot-spots for pathogenic genetic variation such as *PEX6*-associated Zellweger Spectrum Disorder. Seven patients were identified with Bardet-Biedl syndrome and seven patients with Rett syndrome who were all amenable to newly FDA-approved therapeutics. Outreach programs were developed for underserved communities, tailored to genetic disease most prevalent locally, both to educate the patient and provider population.
Discussion
Utilizing common software, we developed a database and Exploratory Data Analysis (EDA) methodology enabling us to systematically reinterpret variants, estimate variant prevalence, identify patients amenable to new treatments, and localize geographies enriched for pathogenic variants.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4065 Assembling unified disease trajectories from sparse longitudinal molecular data.

Authors:

R. Thompson, G. E. Hoffman, D. M. Del Valle, The Mount Sinai COVID-19 Biobank Team, S. Kim-schulze, R. Sebra, M. Merad, S. Gnjatic, S. E. Eric, A. W. Charney, N. D. Beckmann; Icahn Sch. of Med. at Mount Sinai, New York, NY

Modeling the natural history of disease from onset to resolution is an essential prerequisite to understanding disease and optimizing interventions at actionable points in the disease course. Densely sampled longitudinal cohorts are ideal for this but rarely (if ever) achievable in practice given limited access to relevant tissues, long timespans of chronic diseases, high costs, and ethical issues. Cross-sectional cohorts, which are not constrained by longitudinal sampling burdens, lack temporal information. A growing number of research cohorts are hybrid in nature, with sparse longitudinal, high-dimensional data being generated. The short, disjointed sampling periods across the individual disease courses in these cohorts are not directly comparable due to the lack of a common time axis, limiting the ability of existing methods to accurately model natural disease histories. With more sparse longitudinal data sets being generated all the time, there is thus an urgent need for new methods to analyze these data. We devised a basic statistical framework to test the ability of a method to reproduce the correct ordering of samples within each individual in a cohort and used this framework to implement a proof of concept to show the feasibility and accuracy of this approach in a large cohort of patients hospitalized with COVID-19 with sparse longitudinal whole blood molecular profiling data. In the RNA-seq data, we assembled a common time axis (analogous to assembly of a contig from individual DNA sequence reads) using an elastic net and showed that this time axis correctly ordered 73.6% of within-patient sample pairs in 5-fold nested cross-validation, significantly more than the 50% expected by chance. Predicting time of COVID-19 symptom onset in a holdout set of patients using this RNA-seq time axis yielded a significant correlation with the patient-reported day of symptom onset ($p = 1.0E-4$). We also assembled a common time axis from Olink data for the same blood samples with a cross-validated correctly ordered fraction of 77.2%. These time axes from two different data modalities were strongly positively correlated ($R^2 = 0.37$, $p=8.9E-92$). Furthermore, the RNA-seq and Olink time axes both showed strong differential expression in the RNA-seq data, with 13241 and 7593 differentially expressed genes (DEGs) respectively (adjusted $p < 0.05$), even when controlling for time since first sample (339 DEGs). Our work demonstrates that reconstructing disease trajectories is indeed feasible in sparse longitudinal data sets and provides a straightforward and widely applicable method for assembling a cross-validated common time axis for any disease or biological context.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4066 † Assessing Risk in Complex Human Disease Using Polyexposure Scores in the Multiethnic Personalized Environmental Gene Study Cohort

Authors:

D. Lloyd[1], F. Akhtari[1], J. House[1], S. Schurman[2], C. Schmitt[1], J. Hall[3], A. Motsinger-Reif[4]; [1]NIEHS, Durham, NC, [2]Natl. Inst on Aging, NIH, Baltimore, MD, [3]niehs, Durham, NC, [4]NIH, Durham, NC

Common complex diseases such as asthma, diabetes and hypertension present enormous burdens on healthcare systems, and more importantly, on patients. Polygenic scores (PGSs) have emerged as reliable predictors of genetic risk of complex diseases, with emerging possibilities for early disease screening and prophylactic recommendations. However, etiologies of complex diseases and individual susceptibility involve both genetic and environmental components. Comparison of the predictive power of these two components of disease risk is important in understanding and recommending personalized prevention and treatment options. Prior research has shown that development of polyexposure scores (PXSs) that consider a wide array of environmental exposures have higher predictive power for type II diabetes. In the current study, we compare the predictive power of PXS and PGS for common complex traits in the Personalized Environment and Genes Study (PEGS). PEGS is a diverse North Carolina based cohort, representing a multiethnic population with an extensive collection of health, genetic and exposure data. We selected a subset of disease phenotypes to study including asthma, migraines, and hypertension. The environmental exposure data in PEGS includes results from three large surveys that ask about an extensive array of occupational, hobby, and lifestyle exposures as well as geospatial exposure estimates such as distance from environmental bads, the Social Vulnerability Index or the Environmental Justice Index from residential history. We developed PXS for each phenotype to compare to published PGSs. We used LASSO regression to select a weighted sum of exposures for each disease phenotype. For asthma, a PXS built using survey exposome data had an AUC of 0.71, while published PGSs AUC results ranged from 0.50 to 0.70 on multi-ancestral populations. For hypertension, our PXS had an AUC of 0.79, compared with published PGSs AUCs ranging from 0.57 to 0.73. For migraines, our PXS exhibited an AUC of 0.73 versus published a PGS AUC of 0.60. These results highlight the potential PXSs have to complement PGSs, allowing for increased disease screening without the increased costs associated with genomic sequencing. Future work in our group will combine both PGSs and PXSs to allow for identification of high-risk individuals for preventative early treatment and recommended lifestyle changes.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4068 Bayes factors for linear mixed models in genetic association tests

Authors:

Y. Guan[1], D. Levy[2]; [1]Natl. Heart, Lung, and Blood Inst., Framingham, MA, [2]Framingham Heart Study, NHLBI, Framingham, MA

Using linear mixed models in genetic association tests is now a standard practice to effectively control for population stratification and (cryptic) relatedness. Several software packages are available to compute p-values, but none can compute Bayes factors. Bayes factors have many advantages over p-values as evidence for association. For example, p-values are not directly comparable unless they have the same power, while Bayes factors are comparable because the alternative hypothesis is incorporated in the Bayes factor calculation; Bayes factors penalize SNPs of low minor allele frequencies, due to their relatively small information content, to reduce false positives. In addition, we show that Bayes factors can be computed efficiently for the linear mixed model. We first recasted the linear mixed model as Bayesian linear regression with a special prior, where the random effect is replaced by eigenvectors of the genetic relatedness matrix as nuisance covariates, whose priors for effect sizes are proportional to their respective eigenvalues (by a factor η). We then used the standard normal-inverse-gamma conjugate priors on other parameters, which allows us to compute Bayes factors in a closed form. Instead of specifying prior on η, which determines the (relative) size of the random effect, we seek its MAP (maximum a posterior) estimate (one estimate for each SNP) and plug in Bayes factor calculations. We devised a novel iterative procedure to efficiently estimate η. For a given η, we demonstrate that Bayes factors can be evaluated as efficiently as the p-values for a given η. The cohorts in the Framingham Heart Study (FHS), funded by the National Heart, Lung, and Blood Institute (NHLBI), include many independent three generational pedigrees, nuclear families, trios, duos, and singletons. This is an ideal setting in which to apply linear mixed models to control for relatedness among the participants. We analyze samples with whole genome sequencing data and multi-omics phenotypes through NHLBI's TOPMed program and demonstrate the advantage of the having both Bayes factors and p-values in genetic association tests.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4069 Bayesian genome-wide TWAS with reference transcriptomic data of brain and blood tissues identified 93 risk genes for Alzheimer's disease dementia

Authors:

S. Guo, J. Yang; Emory Univ., Atlanta, GA

Background: Transcriptome-wide association study (TWAS) is a powerful tool for identifying novel genes associated with complex diseases whose genetic effects are potentially mediated through transcriptome. TWAS leverages reference genetic and transcriptomic data to estimate genetic effect on expression quantitative traits of target genes sizes (i.e., a broad sense of expression quantitative trait loci, eQTL, effect sizes). Estimated effect sizes will then be taken as variant weights in burden gene-based association test statistics to map risk genes of complex phenotypes using GWAS data. However, existing TWAS of Alzheimer's disease (AD) dementia only considered *cis*-eQTL while neglecting potential *trans*-eQTL. To overcome this limitation, we applied the Bayesian Genome-wide TWAS (BGW-TWAS) method to leverage both *cis*- and *trans*- eQTL of brain and blood tissues to improve mapping risk genes for AD dementia. **Methods:** We first applied BGW-TWAS to the Genotype-Tissue Expression (GTEx) V8 dataset to estimate *cis*- and *trans*- eQTL effect sizes of prefrontal cortex, cortex, and whole blood tissues. Second, estimated eQTL effect sizes were integrated with the summary data of the most recent GWAS of AD dementia to calculate BGW-TWAS (i.e., gene-based association test) p-values of AD dementia per tissue type. Third, we used the omnibus aggregated Cauchy association test (ACAT-O) to combine TWAS *p*-values across three tissues to obtain omnibus TWAS *p*-values per gene. **Results:** We identified 37 genes in prefrontal cortex, 55 in cortex, and 51 in whole blood that were significantly associated with AD dementia. By combining BGW-TWAS *p*-values across these three tissues, we obtained 93 significant risk genes including 29 genes primarily due to *trans*-eQTL and 50 novel genes. With these 93 significant risk genes, we detected 5 functional clusters comprised of both known AD risk genes and novel genes by protein-protein interaction network analysis, and 7 enriched phenotypes by phenotype enrichment analysis. **Conclusion:** In this study, we applied BGW-TWAS and ACAT-O methods to integrate both *cis*- and *trans*- eQTL data of brain and blood tissues with GWAS summary data to identify risk genes of AD dementia. Both known and novel risk genes provide novel insights into the genomic etiology of AD dementia.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4070 BayesKAT: A powerful Bayesian framework to detect combined phenotype-associations of genetic variant groups.

Authors:

S. Das Adhikari, Y. Cui, J. Wang; Michigan State Univ., East Lansing, MI

Although GWAS approaches can help to identify individual SNPs that are significantly associated with specific phenotypes, many complex diseases are polygenic and are controlled by multiple genetic variants that are usually non-linearly dependent. These genetic variants are marginally less effective and remain undetected in GWAS analysis. Kernel Based tests (KBT), which evaluate the joint association effects from a group of genetic variants, are therefore critical for complex disease analysis. However, choosing different kernel functions in KBT can significantly influence the power and accuracy, and automatically selecting the optimal kernel is statistically challenging. A few existing methods suffer from inflated type 1 errors, limited scalability, inferior power, or issues of ambiguous conclusions. Here, we present our Bayesian framework, BayesKAT, which overcomes these kernel specification issues by selecting the optimum kernel adaptively from the dataset while testing genetic associations simultaneously. Furthermore, BayesKAT implements a scalable computational strategy to boost its applicability, especially for high-dimensional cases where other methods become less effective. Based on a series of performance comparisons using both simulated and real large-scale genetics data, BayesKAT outperforms the available methods in detecting complex associations and controlling type I errors simultaneously, leading to new insights into the genetic basis of complex diseases.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4071 Block LASSO for the genome: speeding up the training of polygenic scores in the age of whole genomes and costly computation

Authors:

T. Raben[1], L. Lello[2,1], E. Widen[2,1], S. Hsu[1,2]; [1]Michigan State Univ., East Lansing, MI, [2]Genomic Prediction, Inc., North Brunswick Township, NJ

We develop the "block" LASSO (*blockLASSO*) to generate polygenic scores (PGS). PGS generated with the traditional LASSO use the entire autosome/genome. The L1 penalty in the regression enforces sparsity by minimizing the inclusion of highly correlated SNPs. LASSO based PGS methods have consistently performed amongst the best methods for generating single-ancestry PGS with results validated out of sample, within families, and in longitudinal studies.

However, LASSO based methods are historically, computationally intensive. Genetic databases have typically used array sequencing and recruited upwards of a million participants. Genetic matrices are typically stored as memory mapped instances, but loading a million SNPs for a million participants can take ~8TB of memory. Running a LASSO algorithm requires holding in memory at least two matrices this size. This requirement is so large that even most large high performance computing clusters cannot perform these calculations. To circumvent this issue, most analyses use subsets: e.g., taking a representative sample of participants and filtering SNPs via pruning and thresholding. High end LASSO training uses ~500GB of memory (e.g., ~400k samples and ~50k SNPs) and takes 12-24 hours to compute.

Fortunately, the correlation structure of the human genome is largely block diagonal, that is, SNPs are generally only correlated with other SNPs on the same chromosome (and conservatively within a few million base pairs). We take advantage of this and perform LASSO based regression on subsets of sequencing data and stitch them back together. The main challenge is developing a scheme to find relative normalizations between segments (e.g., the BASIL algorithm by Qian et. al.). We show that simple regression methods can find relative weights that stitch together sub-LASSOs which then match or outperform traditional LASSO PGS as judged by AUC (area under receiver operator curve) and correlation. For example, an asthma PGS, for SNPs alone, leads to an AUC ~0.615 which is a 7% increase from traditional LASSO. We show that the example job above (~400k participants and 50k SNPs) can run in <10 minutes and use ~15 GB of memory using *blockLASSO*. Expanding ten-fold to ~500k SNPs takes 24-48 hours and uses ~150 GB of memory.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4072 Blood pressure and kidney-function traits as component causes of diabetic retinopathy: A two-sample Mendelian randomization study using the Million Veteran Program

Authors:

T. Basnet[1], N. Sarma Vaitinadin[1], J. Breeyear[1], S. Mitchell[2], J. Hellwege[1], B. Charest[3], A. Khakharia[4], C. Halladay[5], O. Wilson[2], P. Reaven[6], J. Meigs[7], M. Rhee[8], Y. Sun[9], M. Lynch[10], C. Nealon[11], M. Breyer[1], C. Dupont[12], S. Iyengar[13], M. Brantley[14], N. Peachey[15], Y. Sun[16], P. Wilson[17], A. Hung[1], L. Phillips[8], T. Edwards[18], A. Giri[1]; [1]Vanderbilt Univ. Med. Ctr., Nashville, TN, [2]VA Tennessee Valley Hlth.care System (626), Nashville, TN, [3]Massachusetts Veterans Epidemiology Res. and Information Ctr. (MAVERIC), VA Boston Hlth.care System, Boston, MA, [4]Dept. of Med. and Geriatrics, Emory Univ. Sch. of Med., Atlanta, GA, [5]Providence VA Med. Ctr., Providence, RI, [6]Coll. of Med., Univ. of Arizona, Phoenix, AZ, [7]Program in Med. and Population Genetics, Broad Inst., Cambridge, MA, [8]Div. of Endocrinology, Metabolism, and Lipids, Dept. of Med., Emory Univ. Sch. of Med., Atlanta, GA, [9]Dept. of Ophthalmology, Stanford Univ. Sch. of Med., Palo Alto, CA, [10]Atlanta VA Med. Ctr., Decatur, GA, [11]Eye Clinic, VA Northeast Ohio Hlth.care System, Cleaveland, OH, [12]2VA Tennessee Valley Hlth.care System (626), Nashville, TN, [13]Case Western Reserve Univ., Cleveland, OH, [14]Dept. of Med., Harvard Med. Sch., Boston, MA, [15]Cole Eye Inst., Cleveland Clinic, Cleveland, OH, [16]Dept. of Epidemiology, Emory Univ. Rollins Sch. of Publ. Hlth., Atlanta, GA, [17]Atlanta VA Med. Ctr., Atlanta, GA, [18]Vanderbilt Univ. Med. Ctr., Brentwood, TN

Diabetic retinopathy (DR) is a highly prevalent microvascular complication of diabetes mellitus and is associated with vision loss. High blood pressure and impaired kidney function often co-occur with DR. Prior epidemiological studies identified hypertension and impaired kidney function as strong risk factors for DR, but it is unclear if these are predictors or have an actual causal role. We used two-sample Mendelian randomization (MR) to assess if BP traits (systolic blood pressure (SBP), diastolic blood pressure (DBP), pulse pressure (PP)) or kidney function traits (estimated glomerular filtration rate (eGFR), urine albumin-creatinine ratio (UACR), blood urea nitrogen (BUN)) are causally associated with risk of DR in European-ancestry individuals. We used statistically significant (P-value $<5x10^{-8}$) uncorrelated (R^2 <0.01) single nucleotide variants (SNVs) from published GWAS as genetic instruments for blood pressure traits (N-SNVs: 213 for SBP, 270 for DBP, and 225 for PP) and kidney-function traits (N-SNVs: 310 for eGFR, 57 for UACR, and 61 for BUN) as exposures, individually. Summary statistics for the same SNVs for DR were generated with logistic regression model in 38,944 DR cases and 69,416 controls from the Million Veteran Program while adjusting for age, sex, duration of diabetes, HbA1c, and ten genetic principal components. We used inverse variance weighted (IVW) random-effects analysis as the primary analysis, followed by MR-Egger, MR-mode, MR-weighted median, and multivariable MR as sensitivity analyses. As a positive control, we performed MR analyses with type 2 diabetes mellitus (T2DM) as exposure for DR. Genetically predicted blood pressure traits were not associated with a higher risk of DR (SBP: OR, 1.01; 95% CI, 1.00-1.01; *P* = 0.089; DBP: OR, 1.00; 95% CI, 0.99-1.01; *P* = 0.47; PP: OR, 1.00; 95% CI, 1.00-1.01; *P* = 0.359) in IVW analysis. Similarly, there was no significant association for genetically determined eGFR (OR, 1.26; 95% CI, 0.76-2.07; *P* = 0.37), UACR (OR, 1.09; 95% CI, 0.90-1.32; *P* = 0.379), or BUN (OR, 0.97; 95% CI, 0.61-1.53; *P* = 0.89) with increased risk of DR. These findings were supported by sensitivity analysis. Moreover, BP and kidney function traits were not associated with proliferative DR. In contrast, T2DM exposure as a positive control, on the other hand, was significantly associated with a higher risk of DR (OR, 1.21; 95% CI, 1.16-1.27; *P* < 0.001). Evidence guided by instrumental genetic variables shows T2DM is causally associated with DR. However, contrary to existing epidemiologic studies, the evidence does not support a role for elevated blood pressure or impaired kidney function as causal factors for DR.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4073 Body mass index stratification improves polygenic prediction of type 2 diabetes in trans-biobank analysis.

Authors:

T. Ojima[1,2], S. Namba[1], K. Suzuki[3,1,4], K. Yamamoto[5,1], K. Sonehara[6,1], A. Narita[2], Y. Kamatani[7], G. Tamiya[2,8], Y. Masayuki[2], T. Yamauchi[4], T. Kadowaki[4,9], T. Kadowaki[4], Y. Okada[10]; [1]Osaka Univ. Graduate Sch. of Med., Osaka, Japan, [2]Tohoku Univ., Sendai, Japan, [3]the Univ. of Manchester, Sale, United Kingdom, [4]Dept. of Diabetes and Metabolic Diseases, Graduate Sch. of Med., the Univ. of Tokyo, Tokyo, Japan, [5]OSAKA UNIV. GRAD. SCHOOL OF MED., Suita, Japan, [6]Graduate Sch. of Med., the Univ. of Tokyo, Bunkyo-ku, Japan, [7]The Univ. of Tokyo, Minato-ku, Japan, [8]Ctr. for Advanced Intelligence Project, RIKEN, Tokyo, Japan, [9]Toranomon Hosp., Tokyo, Japan, [10]Osaka Univ. / The Univ. Tokyo / RIKEN, Osaka, Japan

Type 2 diabetes (T2D) is a heterogeneous subtype composition with different body mass index (BMI) sensitivity. These subtype compositions are an essential factor of exacerbating polygenic prediction across populations. We performed stratified random sampling based on BMI to isolate highly heritable subtypes and to maintain case-control ratios for BMI-related diseases, namely T2D and cardiac artery disease (CAD). Thus, using >97,000 individuals (N_{T2D} = 27,642, N_{CAD} = 22,650) from each of BioBank Japan (BBJ) and UK Biobank (UKBB), we obtained datasets stratified into two or three groups and BMI-unstratified datasets. Comparing two BMI-stratified groups, T2D heritability in the low BMI group was 84.5% and 64.7% greater than for high BMI in BBJ and UKBB, respectively. Polygenic predictions of T2D toward low BMI targets showed 22.2% increased pseudo-R^2 for BBJ and 23.5% for UKBB over predictions for unstratified targets. Using a pathway-specific clustering model, we showed that the improved T2D predictions in low BMI groups were due to biological contributions from multiple pathogenic pathways. By combining BMI stratification and a method integrating cross-populational effects, T2D predictions showed a 37.4% and 47.0% improvement for BBJ and UKBB over unstratified matched-populational predictions, respectively. The improved T2D predictions at low BMI were confirmed when comparing three BMI groups and evaluating odds ratios and net reclassification improvement. The low BMI T2D cases showed higher rates of neuropathy and retinopathy, suggesting that the appropriate medication for low BMI cases could prevent complications. We also replicated the results of improved T2D prediction in the low BMI group with two independent cohorts, Tohoku Medical Megabank Project (N_{T2D} = 6,000, $N_{control}$ = 20,000) and the second cohort of BBJ (N_{T2D} = 11,236, $N_{control}$ = 21,860). In contrast, CAD predictions showed no improvements, suggesting that T2D subtyping improved the predictions. Our results suggest the utility of target stratification based on existing traits for polygenic prediction of heterogeneous diseases.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4074 Boosting SNP discoverability: How the selection of traits affects the power of multi-trait GWAS

Authors:

Y. Suzuki[1], H. Menager[1], B. Brancotte[1], R. Vernet[2], C. Nerin[1], C. Boetto[1], R. Torchet[1], P. Lechat[1], L. Troubat[1], M. Cho[3], E. Bouzigon[2], H. Aschard[1], H. Julienne[1]; [1]Inst. Pasteur, Paris, France, [2]INSERM, Paris, France, [3]Brigham and Women's Hosp., Duxbury, MA

Modern human datasets provide access to hundreds of phenotypes, including quantitative traits and diseases and a range of high throughput molecular data. This plethora of data has fostered the development and application of multi-trait genome-wide association studies (GWAS). Extensive works have optimised the power of those approaches and investigated their robustness and computational efficiency. Conversely, little has been done regarding the strategy to select traits to be analysed jointly, leaving investigators with overwhelming combinatorial possibilities. For example, from a set of ten traits, there are over 1,000 possible sets of 2 to 10 traits that can be analysed jointly. Here, we investigated to what extent genetic architecture can inform the ability of multi-trait approaches to identify new associations missed by univariate GWASs. We used JASS, an efficient Python package, to perform a standard joint analysis of ~20K randomly selected sets of 2 to 12 traits, pulled from a curated database of 72 quantitative traits and common human diseases GWASs. For each set, we derived the gain of the multi-trait test over univariate tests and measured single GWAS features (e.g. heritability, sample size, etc) and GWAS set features (genetic correlation, sample overlap, etc) associated with that gain. We identified four features driving the power gain of the multi-trait test relative to univariate GWAS: mean genetic effect size (P=2e-57), the number of traits analysed jointly (P=3e-20), the genetic covariance across the traits (P=7e-23), and the unaccounted heritability (the difference between common variants heritability and heritability explained by GWAS hits) (P=5e-47). Altogether these parameters carry a substantial predictive power with a predicted versus observed gain measured as squared Pearson's r=0.19 (r=0.4). We also identified multiple scenarios where the proposed data-driven selection approach outperformed clinically based selection (e.g. the power gain in the joint analysis of two sleep-related traits, chronotype and mean least active 5-hour period after previous midnight (L5) was 0.28, while the power gain with L5 and waist-hip ratio was 0.77). Overall, this demonstrates that trait GWAS characteristics can be used to guide the selection of relevant sets of traits for multi-trait analyses in order to maximise the identification of variants with low discoverability in univariate GWAS. (Funding: ANR-20-CE36-0009)

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4075 Breast cancer family history in participants with BRCA variants: A study in*All of Us*Research Program.

Authors:

L. Sulieman[1], C. Chen[1], R. Cronin[2]; [1]Vanderbilt Univ. Med. Ctr., Nashville, TN, [2]Ohio State Univ. Med. Ctr., Columbus, OH

BRCA pathogenic variants are associated with an increased risk for breast cancer (BC). Most studies that investigated BRCA mutations and family history (FH) had small sample sizes or lacked diverse underrepresented populations. Moreover, the scope of most studies was limited by cohort selection that exclusively focused on estimating the prevalence of BRCA mutations in cohorts exhibiting breast cancer family history (BCFH). To close the knowledge gaps, we estimated the prevalence of BCFH in those with BRCA mutation and the prevalence of BRCA mutation in those with BCFH in the *All of Us* Research Program, a large longitudinal cohort from historically underrepresented populations. We used the latest data release (CDR-7) from *All of Us* to identify female participants with BRCA1 and BRCA2 pathogenic variants and completed FH surveys. We identified a control cohort of those that completed FH surveys without BRCA mutations. Using the control population as a reference, we applied the proportion test to check if the BCFH proportions in BRCA1 and BRCA2 cohorts are significantly different from the control population. Among 248,135 female participants, 118,008 completed FH, 145,563 had genome data, and 85,256 completed the FH survey and had genome data. Of those with FH survey and genome data, 16073 (18.9%) reported BC in their family, 411 (2.6%) had BRCA1 and 1054 (6.6%) had BRCA2. In group of multiple family members with BC, 11.2% had BRCA variants. 1130 (77%) BRCA carriers did not have BCFH. In BRCA1 carriers, the prevalence of BCFH was 29.9%. Among BRCA2 carriers, the prevalence of BCFH was 20.1%. The number of female BRCA1 and BRCA2 carriers who reported multiple BCFH were 94 (22.9%) and 153 (14.5%) respectively. Among BCFH members, 134 BRCA1 and 184 BRCA2 carriers reported Grandparent history with BC, which was the most common family member with BC history. In the control cohort that had 83798 participants, the prevalence of BCFH was 18.8% and 12.9% reported multiple family members with BC. BCFH was significantly more prevalent in participants with BRCA variants than in the control cohort (P-val < .00001 for BRCA1/BRCA2). The prevalence of BRCA variants in participants with FH was lower than 7%, which was like a study that calculated BRCA prevalence in cohort with FH, but different than another study found that BRCA mutation prevalence was 21% in 195 patients with BCFH. BCFH was significantly more prevalent (p-val =0.02) in females with BRCA1 variants. Our study was the first large scale that calculated BRCA mutation prevalence in relation to BCFH. The results show that BRCA mutations increase the likelihood of having BCFH, especially for participants with multiple family members with BC.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4076 † BTS: efficient Bayesian framework for functional evaluation of GWAS results against large-scale, heterogeneous functional genomics data collections

Authors:

P. Kuksa[1], M. Ionita[1], K. Clark[2], L. Carter[1], O. Valladares[1], Y. Leung[1], L-S. Wang[1]; [1]Univ. of Pennsylvania, Philadelphia, PA, [2]Univ. of Pennsylvania, Baltimore, MD

Summary statistics from genome-wide association studies (GWASs) are often used in downstream analyses such as identifying tissue/cell context or causal variants/genes underlying GWAS signals, e.g., via fine-mapping or colocalization analyses. The growing number of loci identified by GWASs and available functional genomic (FG) annotation provide a unique opportunity to perform large-scale, systematic functional analyses of GWAS results. It is critical to have scalable algorithms that can integrate thousands of genome-wide FG annotations to identify putatively causal variants, affected tissue/cell types, regulatory mechanisms.

To this end, we propose BTS, a novel, highly efficient algorithm for joint estimation of potentially causal variants, loci, and their relevant functional context. BTS uses large-scale FILER FG database (Kuksa et al 2022) and performs systematic evaluation and prioritization of GWAS variants/loci across >1,000 tissues/cell types and >40,000 enhancer, open chromatin, transcription factor binding, epigenetic marks and other FG tracks. BTS scales well with the number of input annotations and GWAS loci, achieving its efficiency by 1) pre-computing GWAS/LD Bayes factors (BFs) for use across annotations, 2) improved evaluation of annotation-dependent priors, and 3) combining priors/precomputed BFs into annotation-specific models. Compared to state-of-the-art functional fine-mapping method (fastPaintor), BTS estimates a genome-wide Bayesian model for each input cell type/tissue-specific annotation *>2,000x* faster. BTS on average only needs 43s on a 387-loci GWAS and EpiMap/ENCODE enhancer/DNase-seq tracks as opposed to 90,401s using fastPaintor.

We demonstrated BTS utility and performance in discovering and prioritizing relevant tissue context, loci and variants by analyzing GWAS summary statistics on inflammatory bowel disease (IBD), rheumatoid arthritis (RA), systemic lupus erythematosus (SLE), coronary artery disease (CAD). BTS identified and analyzed 387 IBD, 331 RA, 262 SLE, and 385 CAD loci. BTS prioritized tissues/cell types, genes and variants such as ETS2, DAB2 in digestive/blood/immune cell types in IBD; LPL, PLG, SLC22A in blood vessel/lung in CAD, and these findings align with known/putative biological mechanisms.

To conclude, BTS is a joint GWAS variant and functional context mapping framework with greatly improved efficiency and is valuable for researchers to explore tissue and cell-type context in the gene-phenotype association architecture. BTS is available online.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4077 Calibrated prediction intervals for polygenic scores across contexts in diverse populations

Authors:

K. Hou[1], Z. Xu[2], Y. Ding[1], B. Pasaniuc[1]; [1]UCLA, Los Angeles, CA, [2]Univ. of California, San Diego, Los Angeles, CA

Polygenic scores (PGS) have emerged as the tool of choice for genomic prediction of complex traits in a wide range of fields from agriculture to personalized medicine. A critical barrier for their adoption in humans is their context-specific accuracy—variable performance across contexts (e.g., genetic ancestries and/or social determinants of health) raising equity concerns of their applications.

We propose that context-specific prediction accuracy can be captured using prediction intervals that are allowed to vary across contexts; trait prediction intervals denote the range within which the true trait value is likely to be at a pre-specified confidence level (e.g., 90%). Therefore, well-calibrated prediction intervals can properly inform context-specific uncertainties associated with the point predictions to be leveraged in subsequent applications.

We analyzed 72 PGS in UK Biobank (UKBB) and All of Us (AoU) data to find widespread context-specific accuracies; many contexts, including age, sex, and income, impact PGS accuracies with similar magnitude as genetic ancestries. For example, PGS accuracy differed by 50% for individuals across “education years” contexts averaging across 11 PGSs when applied to AoU. In addition, patterns of context-specific PGS accuracy were similar for single-ancestry as well as multi-ancestry PGSs.

We introduce CalPred, a statistical method that uses calibration data to estimate context-specific prediction intervals as a function of each individual's context information. CalPred yields well-calibrated context-specific intervals in simulations and real data analysis of diverse populations from UKBB and AoU. Using CalPred, we show that trait prediction intervals need to be increased by up to 40% to maintain proper calibration across contexts. Overall, our results highlight the need to account for context-specific PGS accuracies in genomic prediction of complex traits.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4078 CARTaGENE reference panel improves genotype imputation in Québec

Authors:

M. Sayahian Dehkordi[1,2,3], R. Laflamme[4,5], K. Sin Lo[5], S. Gravel[1,3], G. Lettre[4,5], C. Bhérer[1,2,3], D. Taliun[1,2,3]; [1]Dept. of Human Genetics, Faculty of Med. and Hlth.Sci., McGill Univ., Montreal, QC, Canada, [2]Canadian Excellence Res. Chair in Genomic Med., McGill Univ., Montreal, QC, Canada, [3]Victor Phillip Dahdaleh Inst. of Genomic Med. at McGill Univ., Montreal, QC, Canada, [4]Université de Montréal, Montreal, QC, Canada, [5]Montreal Heart Inst., Montreal, QC, Canada

CARTaGENE (CaG) is a population-based biobank and a prospective health study containing genotyping data from 29,330 individuals recruited in the province of Québec, Canada. 87% (25,681) of these individuals are of European ancestry, mainly from the French-Canadian founder population, which formed by immigration of ~8,500 mostly French settlers in the 17th century. In this project, we built the first public Québec reference panel for genotype imputation using 2,173 whole genomes (at ~30X coverage) from CaG and showed that it outperforms TOPMed, the largest publicly available reference panel, when imputing rare population-specific genetic variants.

We conducted imputation experiments comparing the performance of the CaG panel (N=2,073), the TOPMed panel (N=97,256), and meta-imputation using three evaluation sets of 100 individuals each.

The CaG panel covered more true alternate alleles (AA) per individual than TOPMed, 99.81% vs 87.25%. On average, TOPMed imputed 99% of AA present in this panel concordantly with true AA, while CaG concordantly imputed only 96.07% of its AA. But in absolute numbers, the CaG panel concordantly imputed on average ~300,000 more AA per individual than TOPMed. The meta-imputation improved the coverage of true AA (99.89%) while maintaining high concordance (96.19%). Meta-imputation had lower concordance than TOPMed due to the inclusion of panel-specific variants from the CaG panel.

With the CaG panel, we concordantly imputed 478,704 (94% with R2>0.6) and 364,259 (95.5% with R2>0.6) rare variants (AF < 0.01) which were absent from or discordantly imputed in TOPMed, respectively. 50.8% of rare variants absent in TOPMed were >5-fold enriched compared to gnomAD non-Finnish Europeans. 39.2% of rare variants discordantly imputed by TOPMed were >5-fold enriched. Meta-imputation recovered all the rare variants that were absent from TOPMed, as well as 357,936 (98.2%) variants that were discordantly imputed by TOPMed.

The CaG panel concordantly imputed all known French-Canadian founder mutations implicated in Mendelian diseases and present in the evaluation set (N=12, e.g. Leigh Syndrome). Three of them were absent from TOPMed, and one was discordantly imputed by TOPMed. Additionally, 4%(8,394 out of 218,337) of variants reported in the GWAS Catalog, were concordantly imputed by CaG in all individuals but were discordantly imputed by TOPMed in at least one individual in the evaluation set or absent from TOPMed.

Our findings highlight the benefits of local reference panels for the imputation of rare population-specific variants.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4079 Casebase analysis for two-phase failure time studies

Authors:

D. Soave[1], K. Kwan[2], C. Greenwood[3], N. Cheng[4], P. Awadalla[4]; [1]Wilfrid Laurier Univ., Waterloo, ON, Canada, [2]McGill Univ., Montreal, QC, Canada, [3]Jewish Gen. Hosp., Montreal, QC, Canada, [4]Ontario Inst. for Cancer Res., Toronto, ON, Canada

Two-phase study designs are ideal for executing targeted sub-studies based on large prospective cohorts when the outcome of interest is a rare event in the full cohort, and additional covariates are expensive or difficult to measure. However, when analyzing such studies, individuals must be properly weighted to account for biased case/control sampling. To address this issue, we propose a weighted casebase framework using logistic regression to obtain unbiased estimates of relative and absolute risks. The benefit of this approach is that it provides smooth-in-time absolute risk functions. To obtain standard errors for the parameter estimates, bootstrapping can be used. We conducted a simulation study to examine the performance of the proposed methodology, and we then used the modelling procedure to predict breast cancer using cell-free DNA methylation profiles derived from blood samples from the Ontario Health Study (OHS).

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4080 Catch me if you can: Genome-wide fine mapping with knockoff e-values

Authors:

P. Gablenz, C. Sabatti; Stanford Univ., Stanford, CA

A common goal in GWAS is to identify DNA variants that carry distinct information on a trait of interest. However, linkage disequilibrium between nearby variants makes it challenging to distinguish which of the many correlated SNPs most directly influence the phenotype. A common solution is then to identify sets of variants that cover the truly important ones. Depending on the signal strengths, local fine mapping strategies allow individual variant contributions to be resolved with more or less precision. However, assuring false discovery rate (FDR) control on the reported findings with these multi-step approaches is often impossible. We design a procedure that, operating on the entire genome, maximizes the number of independent discoveries and their precision, while controlling the FDR. Analysis of data from the UK Biobank illustrates the power of our adaptive approach.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4081 Causal effect of recurrent mild malaria on dyslipidemia in African Ancestry Individuals: A Mendelian randomization study.

Authors:

M. Traore[1,2], H. Sangare[2], O. Diabate[2], C. Cissé[2], O. Nashiru[3], J. Li[4], J. Shaffer[4], M. Wélé[2], S. Doumbia[2], T. Chikowore[5], O. Soremekun[6], S. Fatumo[1]; [1]MRC/UVRI & LSHTM Uganda Res. Unit, Entebbe, Uganda, [2]African Ctr. of Excellence in Bioinformatics of Bamako, Bamako, Mali, [3]H3Africa Bioinformatics Network (H3ABioNet) Node, Ctr. for Genomics Res. and Innovation, NABDA/FMST, Abuja, Nigeria, Abuja, Nigeria, [4]Sch. of Publ. Hlth.and Tropical Med., Tulane Univ., USA, New Orleans, LA, [5]Univ. of Witwatersrand, Johannesburg, South Africa, [6]MRC/UVRI $ LSHTM, Entebbe, Uganda

Dyslipidemia is becoming prevalent in Africa, where malaria is endemic. Observational studies have documented the long-term effect of malaria on lipid metabolism; however, these studies are vulnerable to confounding factors. Therefore, we used Mendelian randomization (MR- a method robust to confounders and reverse causation) to determine the causal effect of recurrent mild malaria (RMM) on lipid traits. We performed two-sample Mendelian randomization Genome Wide association study (GWAS) summary statistics for RMM conducted in Benin, (N=775) and lipid traits from African ancestry individual in Million Veteran Program (N= 57,332). We found a linear association between RMM and levels of low-density lipoprotein cholesterol (LDL-C) (Beta = -0.025, 95% CI, -0.042 to -0.007 p-value=0.005) and total cholesterol (Beta = -0.019, 95% CI, -0.035 to -0.002, p-value= 0.028). No significant association was obtained with High-density lipoprotein cholesterol (HDL-C) and levels of triglycerides. The finding of this study supports a causal relationship between RMM and levels of LDL-C and total cholesterol. We believe that larger studies on the link between malaria and dyslipidemia in Africa will help to manage the burden of both diseases better.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4082 cellSTAAR: Incorporating single-cell-sequencing-based functional data to boost power in rare variant association testing of non-coding regions

Authors:

E. Van Buren[1], Y. Zhang[2], X. Li[3], Z. Li[4], H. Zhou[5], G. Peloso[6], J. Rotter[7], P. Natarajan[8], X. Lin[9], NHLBI Trans-Omics for Precision Medicine (TOPMed)Consortium, TOPMed Lipids Working Group; [1]Dept. of Biostatistics, Harvard T.H. Chan Sch. of Publ. Hlth., CAMBRIDGE, MA, [2]Dana-Farber Cancer Inst., BOSTON, MA, [3]Dept. of Biostatistics, Harvard T.H. Chan Sch. of Publ. Hlth., Cambridge, MA, [4]Indiana Univ. Sch. of Med., Indianapolis, IN, [5]Harvard T.H. Chan Sch. of Publ. Hlth., BOSTON, MA, [6]Boston Univ., Boston, MA, [7]Lundquist Inst., Harbor-UCLA Med Ctr, Torrance, CA, [8]Massachusetts Gen. Hosp., Boston, MA, [9]Harvard T.H. Chan Sch Pub Hlth, Boston, MA

Introduction: Whole genome sequencing (WGS) studies have cumulatively identified hundreds of millions of rare variants, the majority of which are in non-coding regions and of unknown function. Given this large number of genetic variants, existing methods for gene-centric Rare Variant Association Tests (RVATs) in WGS studies have identified relatively few associations between candidate Cis-Regulatory Elements (cCREs) and complex human diseases. Because the regulatory landscape of many cCREs varies across cell types, it is of substantial interest to incorporate single-cell sequencing data into RVATs to capture the functional variability that exists across cell types in the non-coding genome and boost statistical power in the process. **Methods**: We propose cellSTAAR to address two opportunities to improve existing gene-centric RVAT methods as applied to genetic variants in cCREs. First, cellSTAAR integrates single-cell ATAC-seq data to capture variability in chromatin accessibility across cell types via the construction of cell-type-specific variant sets and the upweighting of relevant variants using cell-type-specific functional annotations. Second, cellSTAAR links cCREs to their target genes using an omnibus framework that aggregates results from a variety of linking approaches, each of which uses differing kinds of genomic data and computational approaches, to reflect the uncertainty in element-gene linking. We applied cellSTAAR on Freeze 8 (N = 60,000) of the NHLBI Trans-Omics for Precision Medicine (TOPMed) consortium data to three quantitative lipids traits: LDL, HDL, and TG. **Results**: In at least one cell type, genome-wide significant promoter and enhancer associations were found in several known lipids loci, including *APOE*, *APOA1*, and *CETP*. Unlike existing methods, cellSTAAR reveals variability in the significance of these loci across a variety of cell types and uncertainty in the target gene for significant enhancers. For example, out of 19 cell types analyzed, the significant enhancer near *APOE* was found in only 6 cell types. Included in these 6 are 5 cell types known *a priori* to be highly relevant to lipids: hepatocytes, fetal hepatoblasts, adipocytes, liver endothelial cells, and enterocytes from the small intestine. Although the associated enhancer is contained with the *APOE* gene, 3D-based evidence from SCREEN suggests possible regulation of nearby genes *APOC2* and *APOC4.* This uncertainty in target gene is not reflected in existing RVAT methods. Using a weakened genome-wide significance threshold, the most discoveries using cellSTAAR are found in cell types that are the most relevant to lipids such as those mentioned above.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4083 Characterising polygenic risk for glaucoma in genetically admixed individuals

Authors:

J. Ong[1], W. He[2], Y. Wu[2], X. Han[3], P. Gharahkhani[2], S. MacGregor[2]; [1]QIMR Berghofer Med. Res. Inst., Herston, Australia, [2]QIMR Berghofer Med. Res. Inst., Brisbane, Australia, [3]Massachusetts Inst. of Technology, Boston, MA

Polygenic risk scoring is a common technique to characterise the aggregation of genetic liability on human complex diseases. In the context of glaucoma genetics, we have shown how profiling one's polygenic risk score (PRS) on glaucoma can achieve reasonable risk stratification, where patients at the top-most glaucoma PRS decile on average was diagnosed with glaucoma ten years earlier than the bottom-most group. However, to date there are limited resources to investigate how our findings translate to patients of admixed ancestry - especially those with African ancestries since the baseline glaucoma prevalence among Africans is much higher.

We first derived theoretical distributions of glaucoma PRSs for individuals with varying degrees of admixture and compare them with real-life genetic data. Using 1000Genomes reference datasets, we simulate bi-ancestry haplotypes. Haplotypes from two local ancestries are randomly mixed to specified admixture fractions [e.g. 50% European (EUR), 50% African (AFR)] across five generations. This process is repeated to generate genotypes at different admixture fractions and pairings of five super-populations. We then map the distribution of European-derived glaucoma PRSs onto the admixed population, finding strong similarity to PRS distributions obtained empirically from the UK Biobank cohort.

During the meeting, we will demonstrate simple heuristics on how to estimate glaucoma genetic risk in UK Biobank individuals with (inferred) AFR-EUR admixed ancestries. We compared two approaches: jointly modeling PRSs in local ancestries with PRS-CSx and adjusting for genetic distance by regressing top ancestral PCs. PRS-CSx slightly improved predictions for moderately admixed groups but performed worse for homogeneous Europeans (reduction in log(OR) from 0.70 to 0.35 per unit SD change). Predicting South Asian and East Asian admixtures using the EUR-AFR joint model resulted in lower overall prediction, indicating a potential shift in genetic architecture among Africans. Our study provides insights for delivering PRS findings to admixed individuals and emphasizes the need for larger GWASs in understudied ancestries for better prediction.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4084 Characterization of mitochondrial phenotypes across 28 tissues among four genetically diverse mouse strains as a foundation for QTL mapping

Authors:

M. Lancaster, D. J. Acri, A. Fischer, N. Olchawa, S. Dosunmu, C. Elliott, P. Gillespie, J. Kim, B. Graham; Indiana Univ. Sch. of Med., Indianapolis, IN

Mitochondrial diseases are a common form of multisystem disease (combined prevalence of 1 in 4,300) with limited treatment options. They are typically marked by defects in oxidative phosphorylation (OXPHOS), leading to a failure of the mitochondria to produce enough cellular ATP to meet energy demands. In addition to primary mitochondrial disease, mitochondrial dysfunction is also implicated in chronic and complex diseases, such as cardiovascular and neurodegenerative diseases, cancer, and diabetes. Mitochondrial disorders are also notable for wide clinical, genetic, and tissue-specific heterogeneity. The genetic architecture underlying tissue specific mitochondrial phenotypic variability remains incompletely understood. The first goal of this project, therefore, was to characterize mitochondrial phenotypic variation across 28 tissues within four genetically diverse inbred mouse strains: C57Bl/6J, WSB/EiJ, CAST/EiJ, PWK/PhJ. Specifically, the activity of citrate synthase (CS), the first enzyme within the tricarboxylic acid cycle, as well as the copy number of the mitochondrial genome (mtDNA) was measured in all tissues. CS activity is commonly used as a biomarker of intact mitochondria, with perturbations of mitochondrial content observed in disease states commonly associated with alterations in CS activity. Additionally, mtDNA is located within the mitochondrial matrix and encodes for 37 genes essential for mitochondrial function, including 13 protein coding genes encoding subunits within the electron transport chain (ETC). MtDNA copy number varies broadly over different cell types; however, little is known about tissue-specific mtDNA content regulation. The characterization demonstrates distinct patterns of variation between strains across tissues. Of the 28 tissues analyzed, 16 had significant differences in mean CS activity between strains, and 19 had one or more strains with significant differences in mean mtDNA copy number, suggesting genetic contribution in regulating these traits. The chosen strains are four of the eight founders of the Diversity Outbred (DO) mouse model and capture significant variability that can be utilized for genetic mapping. Using three tissues identified as highly variable in our original founder experiment, we performed quantitative trait locus (QTL) mapping and identified a known cis-coding variant for CS. These results provide validation for a larger future study, and in its completion, this project will aid in the identification of genetic modifiers of tissue-specific CS activity and mtDNA content and provide candidate targets for mitochondrial dysfunction and disease therapies.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4085 Characterization of non-coding variants associated with transcription factor binding through ATAC-seq footprinting in liver.

Authors:

M. F. Dudek[1], B. M. Wenz[1], S. F. Grant[2,1], C. D. Brown[1]; [1]Univ. of Pennsylvania, Philadelphia, PA, [2]Children's Hosp. of Philadelphia, Philadelphia, PA

Chronic liver disease (CLD) represents a massive healthcare burden worldwide, most commonly due to non-alcoholic fatty liver disease (NAFLD). A better understanding of CLD genomic etiology can inform the development of novel treatments. The polygenicity of CLD and related traits has been confirmed by genome-wide association studies (GWAS), which have revealed multiple associated loci. Most of these loci reside in non-coding regions with unknown effects on gene regulation, limiting their utility in elucidating the biological mechanisms of liver disease. Further studies have detected quantitative trait loci (QTLs) with a genetic association for a molecular trait such as gene expression or chromatin accessibility. However, identifying the causal variant within trait-associated loci remains challenging due to linkage disequilibrium (LD), which limits the genomic resolution of both GWAS and QTL studies.

Non-coding GWAS variants are enriched at the binding sites of transcription factors (TFs), suggesting a link between disease association and the disruption of cis-regulatory sequence. TF binding can be detected in ATAC-seq experiments, where bound TFs block the DNA-cleaving transposase Tn5, leaving a pattern of relatively depleted Tn5 cutsites known as a “footprint”. Here we used computational tools to scan ATAC-seq reads from 189 human liver samples, both separated and combined, and quantified TF binding likelihood as *footprint scores* at variants genome-wide.

We observe that when reads across samples were combined, footprints were significantly enriched at liver eQTLs, caQTLs, disease-associated variants, and promoter-interacting regions from Capture-C. Analyzing samples separately according to genotype, we detected in excess of 9,000 variants significantly associated with footprint-inferred TF binding at 5% FDR, known as footprint QTLs (fpQTLs). In particular, we observed fpQTLs at the obesity-associated *FTO* locus (rs8050136), the NAFLD-associated *IFT172* locus (rs2303370), and the triglyceride-associated *CELSR2* locus (rs7528419). Given that variant footprint scores are not in LD, fpQTLs can aid GWAS fine-mapping by precisely locating TF activity within broad disease-associated loci which typically includes several candidate variants. fpQTLs can therefore be leveraged to study the role of TF binding site disruption in disease, and provide functional interpretations for non-coding variants.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4086 Characterizing proteomic and transcriptomic features of missense variants in amyotrophic lateral sclerosis genes

Authors:

A. Dilliott[1,2], S. Kwon[3], G. A. Rouleau[1,2], S. Iqbal[3,4], S. M. K. Farhan[1,2]; [1]Montreal Neurological Inst.-Hosp., Montreal, QC, Canada, [2]McGill Univ., Montreal, QC, Canada, [3]Broad Inst. of MIT and Harvard, Cambridge, MA, [4]Massachusetts Gen. Hosp., Boston, MA

Within recent years, there has been a growing number of genes associated with amyotrophic lateral sclerosis (ALS), resulting in an increasing number of novel variants, particularly missense variants, many of which are of unknown clinical significance. Here, we leveraged the sequencing efforts of the ALS Knowledge Portal (3,864 individuals with ALS and 7,839 controls) and Project MinE ALS Sequencing Consortium (4,366 individuals with ALS and 1,832 controls) to perform proteomic and transcriptomic characterization of missense variants in 24 ALS-associated genes.

The two sequencing datasets were interrogated for rare missense variants in the 24 genes, and variants were annotated with protein sequence features, including Uniprot functional site annotations and PhosphoSitePlus post-translational modification (PTM) site annotations; protein structural features from AlphaFold predicted monomeric 3D structures; and transcriptomic expression levels from Genotype-Tissue Expression (GTEx). We applied missense variant enrichment and gene-burden testing following binning of variation based on the selected proteomic and transcriptomic features to identify those most relevant to pathogenicity in ALS-associated genes.

Using predicted human protein structures from AlphaFold, we determined that missense variants carried by individuals with ALS were significantly enriched in β-sheets and α-helices, as well as in core, buried, or moderately buried regions. We also identified that hydrophobic amino acid residues, compositionally biased protein regions, and protein-protein interaction regions are significantly enriched in missense variants carried by individuals with ALS. Assessment of expression level based on transcriptomics revealed significant enrichment of variants of high and medium expression across all tissues and within the brain. We further explored enriched features of interest using burden analyses and identified individual genes were indeed driving certain enrichment signals. We performed a case study of *SOD1* to demonstrate proof of concept of how enriched features align with the consequences demonstrated by experimentally solved structures of *SOD1* variants and how the features may aid in defining variant pathogenicity.

Our work demonstrates that proteomic and transcriptomic features are important indicators of missense variant pathogenicity in ALS and are distinct from features associated with neurodevelopmental disorders. Incorporating gene-specific features into variant classification may offer the ability to better identify true pathogenic missense variants in ALS clinical cohorts.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4087 Characterizing substructure via mixture modeling in large-scale summary statistics

Authors:

H. Wolff[1], A. Price[1], N. Scribner Trout[2], K. Crooks[1], N. Rafaels[1], N. Pozdeyev[1], S. Tifour[2], K. Marker[1], C. Gignoux[1], A. Hendricks[2,1]; [1]Univ. of Colorado, Anschutz Med. Campus, Aurora, CO, [2]Univ. of Colorado, Denver, CO

Genetic summary data are both broadly accessible and highly useful. Nevertheless, collapsing individual-level data into groups masks intra- and inter-sample heterogeneity (e.g., population structure), biasing results especially in admixed or understudied populations. We address this gap with the development of *Summix v2* (available on GitHub/Bioconductor) that estimates, adjusts for, and even harnesses substructure in summary level data. The suite of methods includes local-ancestry estimation and a statistical test to identify regions of selection; fine-scale ancestry estimation and visualization; allele frequency harmonization for secondary data analysis; and estimation of genetic risk of disease enabling adjustment for differences in ascertainment and sampling.

We evaluated over a comprehensive set of simulation scenarios and application in publicly-available data, including gnomAD v3.1.2 and Genome Wide Association Study summary statistics. In simulations, *Summix v2* produced precise and accurate local-ancestry proportion estimates (≥ 99%) of continental ancestry similarity (African, East Asian, European, Indigenous American, South Asian) using as few as 250 genetic variants. Our test of selection (i.e., ancestry deviance) controlled type I error and had 92% power to detect 400kb regions with 5% difference in local vs global ancestry. For finer-scale ancestry, minimum accuracy of proportion estimates is related to reference group pairwise F_{st}: 89% for F_{st} =0.005; 97% for F_{st} =0.009. Median accuracy over reference groups is always high, >99%. In application to gnomAD v3.1.2, we replicated signatures of selection at the *HLA* locus in the gnomAD African/African-American, (FDR=6.5x10-11) and Latinx (FDR=6.68x10-9) groups and 1p33 in the Latinx group containing *CYP4A11*, a candidate gene in pharmacogenetics (FDR=9.63x10-6). After a Bonferroni correction, we identified 8, 6, and 2 regions containing putative signatures of selection in African/African-American, Latinx, and both, respectively, identifying genes relevant to innate immune response and multiple cancers. Finally, to highlight the ability of *Summix v2* to identify other types of substructure, we estimated the genetic risk of prostate cancer in the Colorado Center for Personalized Medicine Biobank from 146 previously-associated variants, reproducing the observed proportions of cases for those ≥60 years and identifying yet to be realized risk for those 40-60 years and ≤40 years. Through multiple functionalities, *Summix v2* enables a breadth of substructure estimation and improves data harmonization ultimately increasing the robust use of publicly available summary data.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4088 Circulating metabolite levels vary by genetic ancestry in individuals of African ancestry.

Authors:

O. Agasaro[1,2], H. Xu[3], X. Sheng[4], H. Fuller[1], D. Albanes[5], D. V. Conti[4], C. A. Haiman[4], C. Chiang[4], K. Ye[3,6], B. Darst[1,7]; [1]Publ. Hlth.Sci., Fred Hutchinson Cancer Ctr., Seattle, WA, [2]Inst. of Publ. Hlth.Genetics, Univ. of Washington, Seattle, WA, [3]Dept. of Genetics, Univ. of Georgia, Athens, GA, [4]Ctr. for Genetic Epidemiology, Dept. of Population and Publ. Hlth.Sci., Univ. of Southern California, Los Angeles, CA, [5]Natl. Cancer Inst., NIH, Bethesda, MD, [6]Inst. of Bioinformatics, Univ. of Georgia, Athens, GA, [7]Dept. of Epidemiology, Univ. of Washington, Seattle, WA

BACKGROUND: The majority of metabolomic epidemiology investigations have been limited to individuals from European ancestry populations and as such, little is known about how metabolites vary by diverse genetic ancestries. We evaluated whether circulating metabolites vary by genetic ancestry in individuals of African ancestry. METHODS: Participants included 706 men of African ancestry from the Multiethnic Cohort (MEC) with serum metabolomic data measured using mass spectrometry and 6,750 men and women of African ancestry from the UK Biobank (UKB) with plasma metabolomic data measured using nuclear magnetic resonance. Genetic ancestry proportions were calculated using ADMIXTURE with unsupervised K=4 and 15,115 common independent SNPs in 4,765 individuals from West African, South African, East African, and European ancestry populations. The learned population structure was then used to project ancestry proportions in MEC and UKB participants. Linear regression models were used to evaluate the association between each metabolite and genetic ancestry proportions, adjusting for age at blood draw, batch (MEC only), and sex (UKB only). In MEC, enrichment analyses were performed to determine if any pathways were over-represented by metabolites associated with genetic ancestry. To investigate the potential implications of metabolites associated with genetic ancestry, identified metabolites were also tested for association with body mass index (BMI) and for mediation of the association between genetic ancestry and BMI. RESULTS: Of the 1,098 metabolites tested in MEC, 81 were associated with genetic ancestry, including 53 associated with West African ancestry, 9 with South African ancestry, 5 with East African ancestry, and 76 with European ancestry. These metabolites were enriched for diacylglycerols and plasmalogen phospholipids, with metabolites associated with West African ancestry uniquely enriched for xenobiotics involved in xanthine metabolism. Seven of these metabolites, including three phospholipids, were associated with BMI and mediated the relationship between genetic ancestry and BMI. In the UKB, of the 249 metabolites tested, 44 were associated with genetic ancestry, including 37 associated with West African ancestry, 1 with East African ancestry, and 30 with European ancestry, with the strongest association suggesting that West African ancestry was positively associated with total omega-3 fatty acids and particularly, docosahexaenoic acid (DHA). CONCLUSION: These findings suggest that metabolite levels vary by genetic ancestry and that such metabolites could have important implications in health and health disparities.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4089 Classification of gene functional properties using evolutionary constraint.

Authors:

Y. Duan[1], T. Hayeck[2], A. Allen[1]; [1]Duke Univ., Durham, NC, [2]Children's Hosp. of Philadelphia, Bryn Mawr, PA

The identification of genes subject to purifying selection has important applications to disease genetics. The rapid increase in the number of human genomes available has resulted in a large increase in the power of population genetics analyses to detect such genes. In a linear regression-based model (RVIS, 2013), the deviation of common functional variants per gene from genome-wide average was taken as a measure of purifying selection on gene. This approach does not explicitly model the mutation rate per site, which can be different due to the local genetic context, but conditions on the total number of variants observed in the gene as a proxy for this rate. In this work, we explicitly model sequence context dependent mutation rates and directly compare the standing variants of genes to neutral expectations given the mutation rate. In addition, selection on different classes of mutation types depends on the functional mechanism of the gene. Some genes tolerate loss-of-function (LOF) mutations but not missense mutations, e.g., genes that are not dosage sensitive but are affected by protein structural changes.
We developed a likelihood ratio test (LRT) based on a multinomial model involving five classes (non-functional, codon-optimality-change, benign missense, pathogenic missense, LOF) to test the deviation of observed composition of variant types from expectation for each gene. The p-value of the LRT was taken as the measure of purifying selection on the gene. By comparing constraint across variant classes, we categorized genes into different mechanistic functional groups. The results were validated by AUC analyses of several functional and disease relevant gene sets. We found our LRT outperform RVIS in all gene sets. For example, a stillbirth gene set AUCs improve from 0.77 to 0.90 and a ClinGen haploinsufficient gene set AUCs improve 0.78 to 0.87. Also, genes belonging to distinct super gene families exhibit varying patterns of enrichment for constraint in different variant types. Important for diagnostic applications, we find that genes that are more constrained in missense variants than LOF variants, are more likely to contain more ClinVar pathogenic missense variants than LOF variants. Differences in missense and LOF constraint are correlated with the differences in the pathogenic missense and LOF variant counts in ClinVar (spearman's r = -0.45, p-value < 2.2e-16). Finally, we show eGenes from GTEx have greater intolerance to LOF variants compared to non-eGenes, suggesting the potential for identifying causal regulatory variants.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4090 Cluster based approach for deciphering complexity in individuals with neurodevelopmental differences

Authors:

M. Leclercq; CHU de quebec, Québec, QC, Canada

This study aims to assess the use of two different clustering approaches, agglomerative and divisive, in order to group individuals with global developmental delay (GDD) for future basket trials and identify disease related genes. Basket trials have shown success in oncology but have not yet been used in GDD due to the heterogeneity of individuals with neurodevelopmental disorders. The study used the largest cohort of individuals with GDD, the Deciphering Developmental Disorders (DDD), and extracted genotypic and phenotypic information from 6,588 individuals. K-means clustering (divisive) and hierarchical agglomerative clustering (HAC) were used to identify subgroups of individuals. Gene network and molecular function information about the clusters were then extracted. The HAC based on phenotypes identified 16 clusters in individuals with GDD, each presenting one dominant phenotype and other minor phenotypes. The most common phenotypes were delayed speech, absent speech, and seizure. Each phenotypic cluster had several sub-clusters of more closely related genes with diverse molecular functions. K-means clustering also segregated individuals with these phenotypes, but the genetic pathways identified were different from the ones identified by HAC. This study highlights the potential of using in a complementary manner both divisive and agglomerative clustering to group individuals with GDD for basket trials. The results suggest that phenotypic clusters should be subdivided into molecular clusters for increased likelihood of successful treatment and that a combination of both clustering approaches may be necessary for comprehensive treatment development.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4091 Clustering Algorithms for Rare Variant Association Studies: Generating effective therapeutic hypotheses for Complex Traits

Authors:

M. Rivas; Maze Therapeutics, Newburyport, MA

Rare variants play a crucial role in the understanding of complex human diseases and traits. With advancements in sequencing technologies, there has been an increasing interest in studying the association between rare genetic variants and phenotypic outcomes. However, the identification and interpretation of rare variants present unique challenges due to their low frequency and the need for specialized statistical methods. Here, we present an approach we call the Multiple Rare variants and Phenotypes Mixture Model (MRPMM), which clusters rare variants into groups based on their effects on the multivariate phenotype and makes statistical inferences about the properties of the underlying mixture of genetic effects. Using summary statistic data drawn from Genebass in UK Biobank across 30 renal, liver, and sex hormone biomarkers from 394,841 individuals, we demonstrate that our mixture model can identify clusters of variants responsible for significantly disparate effects across a multivariate phenotype. We estimate (1) the proportion of non-null variants, (2) whether variants with the same predicted consequence in one gene behave similarly, and (3) whether different annotations differ in the magnitude of their effects. We find examples of three clusters of rare protein truncating variants in *UMOD* consistent with protection against chronic kidney disease, i.e. lowers creatinine levels, and a second cluster of rare protein truncating variants that tag an ADTKD inframe deletion haplotype that increases creatinine levels. We detect three clusters of effects for protein altering variants in *ANGPTL8* and triglyceride levels showcasing a bidirectional effect profile. However, challenges remain in the application of clustering algorithms for rare variant association studies. These include the selection of appropriate genetic variants due to the reduction of power to detect signals when including neutral alleles when analyzing protein altering variants. Here, we will present approaches for addressing some of those challenges.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4092 COLOC-Boost: A new gradient boosting informed colocalization algorithm improves the identification of functional disease causal variants

Authors:

X. Cao[1], H. Sun[1], R. Mazumder[2], K. Dey[3], G. Wang[1]; [1]Columbia Univ., New York, NY, [2]Massachusetts Inst. of Technology, Cambridge, MA, [3]Sloan Kettering Inst., Mem. Sloan Kettering Cancer Ctr., New York, NY

Background
1. Expression quantitative trait locus (eQTL) studies have had considerable success in understanding the regulatory function of disease variants; recent efforts expand QTL studies to molecular phenotypes beyond expression, like histone acetylation, methylation, protein abundance. 2. Colocalization methods like COLOC, eCAVIAR (Wallance 2021, Hormozdiari et al. 2016) have been widely used to identify variants and loci with shared causal variant underlying a molecular phenotype and trait. 3. The expansive types and nature of the QTL studies demand a colocalization method that can accurately pinpoint the functional disease causal variant from tagging variants, and that can be easily extended to multiple phenotypes.
Method
1. We propose a new colocalization method, COLOC-Boost, that uses coupled gradient boosting algorithm to automatically distinguish colocalized sets of variants from un-colocalized variants with superior precision and recall compared to existing approaches.2. Though inherently a locus-level colocalization method, COLOC-Boost assigns a probabilistic grade to each variant in a colocalized set, that can be used to perform variant-level colocalization.3. The conceptual gradient boosting framework of COLOC-Boost can be easily extended to multiple phenotypes and even to a single phenotype fine-mapping.
Results
1. We performed a systematic comparison of COLOC-Boost against SuSiE-COLOC for a broad range of simulation settings where the causal variants are in varying levels of linkage disequilibrium (LD) from 0.1 to 0.9 and the phenotype heritability varies from 0.2 to 0.7. 2. For locus-level colocalization, COLOC-Boost exhibited 1.08x-1.83x significant higher power than locus-level version of SuSiE-COLOC at high levels of LD (LD>0.5; p-value<0.001) and similar power at low LD levels, while had similar false discovery rate (FDR). 3. Locus-level colocalization methods (COLOC-Boost and locus-level version of SuSiE-COLOC) reduced FDR to 2.9%-58.1% of SuSiE-COLOC, and improved 1.27x-6.08x power across all settings (p-value<0.001). 4. The fine-mapping equivalent of COLOC-Boost, term FINE-Boost, had comparable performance to SuSiE at low to moderate levels of LD, but slightly outperformed the latter for high LD (LD>0.9) scenarios.5. Currently, COLOC-Boost is being applied to integrate an emerging multi-omics QTL and AD GWAS resource developed by the FunGen-xQTL Project, yielding promising preliminary evidences of propagation of xQTL on the etiology of complex diseases.6. Overall, COLOC-Boost provides a novel framework to identify colocalized disease-critical functional signals for varying number of phenotypes.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4093 Common Genetic Variants are Associated with Plasma and Skin Carotenoid Metabolism in Ethnically Diverse US Populations

Authors:

Y. Han[1], S. Mwesigwa[2], M. N. Laska[3], S. B. Jilcott Pitts[4], N. E. Moran[5], N. A. Hanchard[1]; [1]Natl. Human Genome Res. Inst., NIH, Bethesda, MD, [2]Dept. of Med. Microbiol., Coll. of Hlth.Sci., Makerere Univ., Kampala, Uganda, [3]Div. of Epidemiology & Community Hlth., Sch. of Publ. Hlth., Univ. of Minnesota, Minneapolis, MN, [4]Dept. of Publ. Hlth., East Carolina Univ., Greenville, NC, [5]USDA/ARS Children's Nutrition Res. Ctr., Dept. of Pediatrics, Baylor Coll. of Med., Houston, TX

Carotenoids, natural pigments found in plants and lower organisms, possess antioxidant and anti-inflammatory properties with potential benefits against oxidative disorders, arteriosclerosis, obesity, and certain cancers. Blood and skin carotenoid levels are reliable biomarkers for assessing carotenoid dietary intake. However, the impact of genetic variability on carotenoid metabolism in diverse populations, and the generalizability of findings from European populations regarding the genetic control of carotenoid metabolism in other populations remains understudied. To address this, we investigated the association between genetic variation and plasma carotenoid species and total skin carotenoid concentrations in a diverse cohort of 207 US adults. Genome-wide genotyping used the H3Africa microarray, complemented by targeted sequencing of 35 genes reportedly important in carotenoid metabolism. Principal Component Analysis effectively grouped individuals self-reporting as 'white' and 'African American' with 'European' and 'African' ancestry, respectively. Self-reported 'Asian' clustered with both Indian/South Asians and East Asians, while Hispanic clustered with admixed Mexican and Latino groups from 1000 Genomes. With 7,467,403 imputed SNPs (MAF $\geq$ 0.5%) from the Michigan Imputation Server ($r^2 \geq 0.3$), we conducted multivariate logistic regression analysis including covariants (age, sex, BMI, race-ethnicity and carotenoids intake), identified a novel genome-wide significant association between *ATF6* variants (lead SNPs *rs11579627*) and plasma total carotenoids (*p*=2.5x10e-8; *Beta*=0.34). Two common missense variants in *PKD1L2* (*rs4889261 and rs7194871*) from targeted sequencing showed significant associations with plasma beta-carotene levels (*p*=9.2 x10e-5, adjusted *P*=0.04, *Beta*=-0.3; *p* = 1 x10e-4, adjusted *P*=0.04, *Beta*=-1.3, respectively). Pathway analysis revealed a link between suggestive SNPs (p < 5x10e-6) and lipid metabolic pathways, highlighting the lipid-soluble nature of carotenoids. We also found that allele frequency differences primarily drove variation in skin carotenoid metabolism between ancestry populations. Our research enhances the understanding of genetic factors in carotenoid metabolism, revealing a novel role for lipid metabolism-related variants. It emphasized the need for diverse genetic ancestries in genomic studies and considering genetic and environmental variables for carotenoid biomarkers. Overall, our study contributes to advancing precision nutrition and precision health.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4094 Comparative analysis of fine-mapping methods at loci with multiple association signals

Authors:

R. Walters[1], J. Sidorenko[2], K. Lin[1], A. Pozarickij[1], K. Young[3], S. Berndt[4], Y. Lu[5], S. Vedantam[6], L. Yengo[2], C. Spracklen[7]; [1]Univ. of Oxford, Oxford, United Kingdom, [2]Inst. for Molecular BioSci., Brisbane, Australia, [3]Univ North Carolina, Chapel Hill, NC, [4]Natl. Cancer Inst., Rockville, MD, [5]Vanderbilt Univ. Med. Ctr., Nashville, TN, [6]Boston Childrens and Broad Inst., Boston, MA, [7]Univ. of Massachusetts, Amherst, MA

Genome-wide association studies (GWAS) have proved highly effective at identifying loci associated with diseases and complex traits. A critical step in translating these findings into precision medicine is pinpointing the underlying causal variant(s). Fine-mapping to single variant resolution remains challenging, especially at loci with >1 distinct association signal, and multiple statistical methods have been developed for identifying and characterizing distinct signals from a combination of summary statistics and a linkage disequilibrium (LD) reference dataset.

LD between variants at a locus is a key factor limiting identification of causal variants, and the performance of fine-mapping methods is recognised to be optimal when the LD reference closely matches the GWAS population. However, achieving a closely matched LD panel is challenging for association signals derived from large multi-cohort (and potentially multi-ancestry) meta-analyses, and it is unclear how sensitive these approaches are to the extent of mismatch or size of the LD panel.

To investigate the sensitivity of commonly-used fine-mapping methods to variation in the size of the provided LD panel, we performed a GWAS of height in 456K European-ancestry individuals in the UK Biobank and selected 7 loci of varying complexity: 4 ‘simple’ loci with 1-2 signals within 1 Mbp (at *TANC1, RBMS1, FAM178B, UGGT1*); and 3 ‘complex’ loci with 6-8 signals (at *ZFAT, LTBP1, DIS3L2*). We then compared the results from analyses of these loci using Stepwise Exact Conditional Analysis (SECA), GCTA-cojo, FINEMAP, and SuSiE, with varying LD references of unrelated UK Biobank individuals, one of 350K, and triplicate randomly-selected subsets of 80K, 40K, and 10K.

Compared with SECA, GCTA-cojo produced near-identical results at all 7 loci when using the 350K LD panel and was the method most robust to reductions in LD panel size; nevertheless, it gave more disparate results for complex loci as panel size decreased. By contrast, even with the 350K panel, lead variants and credible sets for the association signals identified by FINEMAP and SuSiE differed substantially, not only from SECA but also from each other, with only limited agreement even in the number of identified signals; for instance, even at the ‘simple’ *FAM178B* locus, SECA and GCTA-cojo identified 2 signals, but FINEMAP and SuSiE identified 5 and 9 signals, respectively. The extent of these differences increased with decreasing panel size. The substantial disagreement between these approaches, as well as their high sensitivity to the size of even a well-matched LD panel, emphasises the need for continued development of fine-mapping methods for complex loci.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4095 Comparing imputation quality in two distinct Latin American populations: Influence of super population exclusion

Authors:

J. French-Kwawu[1], V. Borda Pua[2], R. Laboulaye[2], T. Peixoto Leal[3], I. Fernandez Mata[4], T. O'Connor[2]; [1]Univ. of Maryland Baltimore, Baltimore, MD, [2]Univ. of Maryland Sch. of Med., Baltimore, MD, [3]Cleveland Clinic, Cleveland, OH, [4]Cleveland Clinic Fndn., Cleveland, OH

GWAS are a powerful tool for identifying potential associated between variants and phenotypes. However, GWAS rely on imputation and many existing imputation reference panels are mainly on individuals of European ancestry, making it difficult to reach the same levels of imputation quality in non-European populations. Specifically for Latin Americans (LatAms), historical events have led to sex-biased admixture generating differences in ancestry composition between the autosomes and sex chromosomes (chrs). It is also important to include chr X as sex is a risk factor for many phenotypes. We aim to analyze how the composition of the imputation reference panel affects imputation quality in LatAm populations. Two of the many ways to explore this is by excluding one of three super populations (African, European, or predominantly Native American) in the reference panel, and including LatAms with increasing proportions of Indigenous American (IA) ancestry. To address this, we created imputation panels for both chrs X and 7, since chr 7 is most similar in size to chr X.

We created 4 panels for each chr with 12,000 individuals from the Genetics of Latin American Diversity (GLAD) Project and the Trans-omics for Precision Medicine (TOPMed) program, each containing: 1) Europeans and LatAms (EL), 2) Europeans and Africans (EA), 3) Africans and LatAms (AL), and 4) individuals from all 3 populations (ALL). To compare the effects of increasing IA ancestry, we ordered the 20,079 LatAm individuals by IA ancestry proportion and selected the first 4,000 individuals and joined them with 4,000 African and European individuals. We repeated this step, sliding the window of LatAm individuals by 2,000, until 10 panels were created.

We included two target populations: the Latin American Research Consortium on the Genetics of Parkinson's Disease (LARGE-PD, predominantly South American) and the Columbia University Study of Caribbean Hispanics with Familial and Sporadic Late-Onset Alzheimer's Disease (AD). We limited genotyped SNPs to those with an $R^2 \geq 0.8$, and compared empirical R^2 values, which estimate the correlation between true genotypes and imputed dosages for only genotyped SNPs, between cohorts and chrs.

In the AD cohort, we observed that EL outperformed AL on chr 7, but performed similarly on chr X. In LARGE-PD, EL outperformed AL on chr 7, but AL outperformed EL on chr X. The differences between results in the cohorts can relate to different admixture dynamics between Caribbeans and South Americans, suggesting that population structure and sex-biased admixture may affect the quality of imputation. We expect to find differential effects with increasing the IA ancestry proportions.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4096 Comparison of ancestry calibration methods for colorectal cancer polygenic risk score to ensure equity in the clinic.

Authors:

E. Rosenthal[1], L. Hsu[2], M. Thomas[3], U. Peters[4], C. Kachulis[5], K. Patterson[6], G. P. Jarvik[7]; [1]Univ. of Washington Sch. of Med., Seattle, WA, [2]Fred Hutchinson Cancer Res. Ctr., Seattle, WA, [3]Fred Hutch, Redmond, WA, [4]Fred Hutchinson CA Res Ctr, Seattle, WA, [5]Broad Inst., Cambridge, MA, [6]Univ. of Washington, Seattle, WA, [7]Univ Washington Med Ctr., Seattle, WA

Colorectal cancer (CRC) is the 3rd most common cancer in the U.S., and is associated with the 2nd most deaths from cancer. CRC is a complex disease with monogenic, polygenic and environmental/lifestyle risk factors. Polygenic risk scores (PRS), which aggregate low risk variants across the genome, are currently being developed to identify individuals at high risk of developing CRC due to a polygenic component, and to inform the starting age of preventive screening. As causal loci, linkage disequilibrium and minor allele frequencies vary across ancestral groups, PRS must be developed using diverse ancestral data to be equitably applied in the clinic. Similarly, the distribution of raw PRS varies by ancestry, necessitating calibration so that risk stratification is valid regardless of genetic ancestry. We compared 4 calibration methods using the All of Us Research Program Whole Genome Sequence data (N=98,256) for a CRC PRS previously developed in participants of European and East Asian ancestry. Case (N=348) and control (N=12,378) status was determined using a validated algorithm (PheKB #514). Methods 1 and 2 use linear models and the entire dataset to center the raw PRS. Methods 3 and 4 train a linear model to estimate the expected mean and variance on an ancestrally diverse subset of the unassigned participants, which is then applied to case/control data. Methods 1 and 3 use the first 5 principal components of ancestry; methods 2 and 4 use admixture estimates for global ancestries: African (AFR), Admixed American (AMR), European (EUR), East Asian (EAS), Middle Eastern (MID), and South Asian (SAS). Participants were grouped into ancestral clusters based on admixture estimates: AFR (23%), AMR (9.5%), EUR (45%), MID (4.1%) and other (OTH) (16%). Participants clustering with EAS and SAS made up < 3% of the sample and did not have power for statistical testing. The AUC (95% bootstrap C.I.) of the PRS were > 0.5 in AFR 0.59 (0.51,0.67), AMR 0.71 (0.60,0.83), EUR 0.63 (0.59,0.67) and OTH 0.63 (0.56,0.70) but not for MID 0.53 (0.50,0.67). Calibration method 4 had the best performance overall, resulting in distributions that were closest to standard Normal with accurate upper tail frequencies for each ancestry. Sample size was sufficient only in the EUR cluster to show that participants with PRS in the top 20% had higher risk of CRC than participants in the middle (40-60%) with OR=2.3, C.I.=(1.5,3.4). Larger, more diverse datasets are required to further develop and validate the PRS. Ultimately, a more comprehensive ancestrally calibrated risk that includes environment, genetics, and social determinants of health is necessary for equitable application in the clinic.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4097 Compound heterozygous effects across common phenotypes in the UK Biobank.

Authors:

F. Lassen[1], S. Venkatesh[1], N. Baya[1], A. Bloemendal[2], B. Neale[3], B. Kessler[1], W. Zhou[3], C. Lindgren[4], D. Palmer[5]; [1]Univ. of Oxford, Oxford, United Kingdom, [2]Broad Inst., Cambridge, MA, [3]Massachusetts Gen. Hosp., Boston, MA, [4]Big Data Inst., Oxford Univ., Oxford, United Kingdom, [5]Univ. of Oxford, Big Data Inst., Oxford, United Kingdom

Exome-sequencing association studies have successfully linked rare protein-coding variation to risk of thousands of diseases. However, whilst heterozygous and homozygous variants have been well-studied, the phenotypic impact of rare deleterious compound heterozygous variation has not been investigated at scale.

Here, we leverage advances in statistical phasing to accurately phase rare variants (minor allele frequency (MAF)~0.001%) in exome sequencing from 175,587 UK Biobank participants. We then systematically annotate putatively deleterious compound heterozygous coding variation. We show that 6.5% of individuals carry such damaging variants in the compound heterozygous state, with 90% of variants occurring at MAF<0.34%.

Using a logistic mixed model framework, systematically accounting for relatedness, polygenic risk, nearby common variants, and rare variant burden, we investigate recessive effects in common complex diseases. We discover 8 gene-trait associations at 'exome-wide significance' ($P<1.68x10^{-7}$). These include well-established recessive gene-trait pairs with recessive mode of inheritance, including *MUTYH* and colorectal cancer ($\beta=10.8$, $P=2.5x10^{-12}$). We further incorporate age-at-diagnosis information from primary care electronic health records, to show that genetic phase influences lifetime risk of disease across 20 gene-trait combinations (FDR < 5%). Our findings show that for certain gene-trait pairs, individuals with a single disrupted gene copy have a risk of developing disease that is virtually indistinguishable from that of wildtypes, suggesting clear non-additive gene dosage effects. Through statistically shuffling haplotypes across cases and controls, we provide the first large scale cohort evidence that genetic phase is indeed the primary determinant of disease susceptibility for a range of gene-trait pairs, including *FLG*-asthma ($P=0.0027$).

Taken together, we demonstrate the utility of phasing large-scale genetic sequencing cohorts for robust identification of the phenome-wide consequences of compound heterozygosity.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4098 Comprehensive evaluation of CNV detection leveraging SV support on Illumina DRAGEN™ Bio-IT platform v4.2 using high quality human genomes and clinical samples

Authors:

Y. Gao, M. Fleharty, E. Malolepsza, D. Toledo, N. Lennon; The Broad Inst. of MIT and Harvard, Cambridge, MA

Copy Number Variations (CNVs) significantly contribute to human genetic diversity and disease predisposition, necessitating the development of precise and efficient detection methodologies in genomics research and clinical diagnostics. In this context, we present a study that evaluates the CNV detection performance of the CNV caller that leverages both depth and structural variant signals in the latest Illumina DRAGEN™ Bio-IT Platform v4.2. For the evaluation, we utilized high-quality data from the Telomere-to-Telomere (T2T) sequenced human genome HG002 and its associated dipcall variant calling results as a benchmark. Comparing CNV calls from DRAGEN™ for HG002 against this ground truth, we assessed the platform's ability to detect deletion (copy number losses) and duplication (copy number gains) events across an extensive range of CNV sizes. Further extending the robustness of our study, we incorporated a set of samples featuring known clinically significant chromosomal aberrations. This added dimension of real-world clinical data allows for a deeper examination of the platform's potential performance in a diagnostic context. To evaluate the platform's precision and potential for erroneous calls, we sequenced CHM13 using the NovaSeqX™ platform and aligned the data to the CHM13-T2T reference. As part of our evaluation process, we considered the CHM13-T2T reference to encompass all true CNV events. Consequently, any calls made in relation to this reference were classified as false positives. Our evaluation, stratified by CNV type and length, provides a nuanced understanding of the sensitivity and specificity of CNV detection from Illumina DRAGEN™ Bio-IT Platform v4.2. The insights drawn from this study offer valuable guidance for researchers and clinicians dealing with high-throughput genomic data and contribute to the continued enhancement of CNV detection methodologies. Please note that the results of this study are intended for research purposes only and are not for use in diagnostic procedures.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4099 † Comprehensive polygenic prediction of respiratory diseases: a cross biobank multi-trait and multi-ancestry approach

Authors:

Y. He, Y. Wang, K. Tsuo, A. Martin; Massachusetts Gen. Hosp., Boston, MA

Polygenic risk scores (PRS) are poised to become valuable risk prediction tools for improving precision medicine and health outcomes. However, they demonstrate poor trans-ancestry portability, thus hindering their clinical utility and potentially exacerbating disparities. To date, most PRSs have been developed for single traits and single ancestry groups without considering genetic correlation between traits or ancestry specific linkage disequilibrium (LD) and allele frequency patterns. Here, we develop PRS-xtra (cross TRait and Ancestry), a stepwise approach to model multi-trait and multi-ancestry features. We demonstrate its greater predictive ability compared to existing approaches for respiratory illnesses.

Respiratory diseases such as chronic obstructive pulmonary disease (COPD) and lung cancer are leading causes of morbidity and mortality globally and share many comorbidities and risk factors. We thus jointly modeled genome-wide association studies (GWAS) for 7 strongly correlated traits-COPD, asthma, lung cancer, FEV1, FVC, smoking status, and cigarettes/day-in African (AFR), Admixed American (AMR), East Asian (EAS), and European (EUR) ancestry groups. We leveraged the largest and most diverse GWAS from the Global Biobank Meta-analysis Initiative, Pan UK Biobank, and GWAS & Sequencing Consortium of Alcohol and Nicotine use. We employed the multi-trait analysis of GWAS (MTAG) method to model genetic correlation between traits, then PRS-CSx to model LD within and between populations. Altogether, we derived 56 PRSs in 120614 individuals from the All of Us research program (AoU). We synthesized PRS-xtra using elastic net regularization to penalize number of PRSs and their effects. We evaluated PRS-xtra in held out test sets from AoU and Mass General Brigham Biobank.

All traits were significantly correlated with each other, with the strongest correlations being between asthma and COPD ($rg=0.705$, $p<0.0001$) and lung cancer and cigarettes/day ($rg=0.614$, $p<0.0001$). PRS-xtra demonstrated significant improvements in predictive performance, especially in non-EUR populations. For example, PRS-xtra significantly improved predictive accuracy of COPD in AMR and AFR populations, with AUC increasing from 0.522 to 0.654 and from 0.504 to 0.535, respectively.

We conducted the most powerful multi-trait and multi-ancestry genetic analysis of respiratory diseases and auxiliary traits to date. We propose PRS-xtra as a method to model genetic correlation across traits and LD differences between ancestry groups that demonstrate significantly better disease prediction, thereby advancing more equitable and generalizable prediction models.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4100 Computational algorithms for phenotypic patient matching to achieve diagnoses in rare disease cohorts.

Authors:

I. Cooperstein[1], A. Ward[1,2], G. Marth[1]; [1]Univ. of Utah, Salt Lake City, UT, [2]Frameshift Genomics Inc, Boston, MA

Introduction: Patients with rare monogenic diseases often lack a clear disease-causing variant or mutation following sequencing. Computational approaches that efficiently identify functionally import variants for analysis are essential. However, incomplete knowledge of gene associations and diverse clinical phenotypes, which rarely perfectly align with disease descriptions in databases, limit the power of many existing variant prioritization tools. We present a computational approach that utilizes real patient data from deeply genotyped and phenotyped rare disease databases. Our algorithmic and visualization tools automate the identification of patients with similar genetic mutations and disease phenotypes, which can be crucial to inform diagnoses and significantly advance knowledge of a rare disease. **Methods:** The hierarchically organized Human Phenotype Ontology (HPO) terms describing patients' disease presentations enable calculations of phenotypic similarity scores between patient pairs. We apply a Bayesian statistical framework to determine the significance of high-scoring patient pairs. Our method prioritizes candidate variant lists of new or previously unsolved cases based on the diagnostic genes of high-scoring diagnosed patient matches. Furthermore, our algorithms identify clusters of phenotypically similar undiagnosed patient matches, then search for high-impact variants within shared genes across all cluster members to discover novel disease genes. **Results:** Applying our methods to the difficult-to-diagnose patient cohort within the NIH's Undiagnosed Diseases Network (UDN) demonstrate that patient pairs with matching clinical diagnoses exhibit significantly higher phenotypic similarity scores than when scoring the patients against the Orphanet disease description of their diagnosis. Further, patients with diagnoses in the same gene exhibit significantly higher phenotypic similarity scores than those with diagnoses in different genes. We have identified compelling candidate variants for undiagnosed UDN cases based on their phenotypic and genotypic similarity to other UDN patients. **Conclusion:** Our approach leverages the power of phenotypic and genomic data in existing datasets to automate patient matching, offering a promising pathway to improve diagnostic rates and generate new insights into gene function in rare monogenic diseases. This development is part of the larger Calypso project, which aims to provide longitudinal support to genomic medicine. Applying our algorithms to reanalyze data over time will allow us to continually revisit patient cohorts to discover new impacting and relevant genes.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4101 Contribution of rare variants to heritability of a disease is much greater than conventionally estimated: modification of allele distribution model

Authors:

Y. Nagao; Tokai Univ. Hosp., Isehara, Japan

Missing heritability is a current problem in human genetics. I previously reported a method to estimate heritability of a polymorphism (h_p^2) for a common disease without calculating the genetic variance under dominant and the recessive models. Here, I extended the method to the co-dominant model and carry out trial calculations of h_p^2. I also calculated h_p^2 applying the allele distribution model originally reported by Pawitan et al. for a comparison. Unexpectedly, h_p^2 calculated for rare variants with high odds ratios was much higher. I noticed that conventional methods use the allele frequency (AF) of a variant in the general population. However, this implicitly assumes that the unaffected are included among the phenotypes: an assumption that is inconsistent with case-control studies in which unaffected individuals belong to the control group. Therefore, I modified the allele distribution model by using the AF in the patient population. Consequently, the h_p^2 of rare variants was quite high. Recalculating h_p^2 of several rare variants reported in the literature with the modified allele distribution model, yielded results were 3.2 - 53.7 times higher than the original model. These results suggest that the contribution of rare variants to heritability of a disease has been considerably underestimated.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4102 Coordinated Epistasis reveals Symptom-Specific Polygenic Pathway Interactions in Major Depressive Disorder

Authors:

J. Rietkerk[1], L. Huang[1], V. Appadurai[2], V. Trindade-Pons[3], H. van Loo[3], L. Garvert[4], S. van der Auwera[4], M. Adams[5], H. J. Grabe[6], A. Schork[7], B. Müller-Myhsok[8], A. Dahl[9], N. Cai[1]; [1]Helmholtz Munich, Neuherberg, Germany, [2]Inst. of Biological Psychiatry, Roskilde, Denmark, [3]Univ. Med. Ctr. Groningen, Groningen, Netherlands, [4]Univ. of Greifswald, Greifswald, Germany, [5]Univ. of Edinburgh, Edunburgh, United Kingdom, [6]Univ. of Greifswald, Greifwald, Germany, [7]Res. Inst. for Biological Psychiatry, Roskilde, Denmark, [8]Max-Planck Inst. of Psychiatry, Munich, Germany, [9]UChicago, Chicago, IL

Major Depressive Disorder (MDD) is a highly heterogeneous disorder; patients often exhibit very diverse symptom profiles, some of which have no symptoms in common. Mounting literature suggests that MDD is likely the common outcome of many different etiological pathways. In this project, we aim to identify polygenic pathways that contribute to symptom heterogeneity in MDD. We do this through identification of interactions between their genetic effects on the disorder using Coordinated Epistasis (CE). CE tests for interactions between polygenic effects of etiological pathways by aggregating the effects into Polygenic Risk Scores (PRS). For a polygenic disorder like MDD, CE assumes that those genetic effects are distributed across the genome. Therefore, we approximate these pathways with random genome partitions. We validate this assumption through simulation: we randomly partition the genome into halves 100 times. The resulting 100 interaction effect estimates (γ) constitute a distribution over the different partitions and its standard deviation varies with polygenicity. Next, we apply this 100-partition test to MDD and its 14 worst-episode symptoms in the UK Biobank. We find that their γ distributions capture both the pathways' polygenicity and their interaction effects. To further characterize these, we test for CE between PRS of all chromosome pairs. We find 2 significantly interacting chromosome pairs in MDD and 64 significantly interacting chromosome pairs in 13 out of 14 MDD symptoms (at 10% FDR). We then follow-up on significantly interacting pairs through SNP-PRS interaction tests and identified 102 chromosome-wide significant loci. Strikingly, we find that the significant SNP-PRS interactions for specific worst-episode symptoms of MDD align with previously found GWAS hits for MDD and other MDD-related phenotypes. For example, chromosomes 1 and 11 show significant PRS interactions for the worst-episode symptom hypersomnia (P = 3e-04); the locus involved (rs4245154, SNP-PRS interaction test P = 1.2e-06) is in an intron of the gene DRD2, which has previously been associated with MDD and neuroticism. We identify, for the first time, polygenic pathway interactions in MDD and its worst-episode symptoms. Our simulations show that γ forms a distribution over partitions of the genome, and simultaneously indicates interaction effects and pathway polygenicity. We further find specific loci that contribute to pathway interactions for symptoms of MDD. These loci are previously found associated with MDD and a variety of MDD-related phenotypes. This demonstrates the relevance of these symptom interactions as part of MDD genetic architecture.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4103 Copy number variants associated with neuropsychiatric traits differ in prevalence across ancestry groups.

Authors:

L. M. Schultz[1], G. Huguet[2], M. Jean-Louis[2], Z. Saci[2], J. Mollon[3], E. Knowles[3], D. C. Glahn[3], S. Jacquemont[2], L. Almasy[4,1]; [1]Children's Hosp. of Philadelphia, Philadelphia, PA, [2]CHU Sainte-Justine, Montréal, QC, Canada, [3]Boston Children's Hosp., Boston, MA, [4]Univ. of Pennsylvania, Philadelphia, PA

Rare copy number variants (CNVs) have been associated with a variety of health outcomes and neurodevelopmental traits. We previously found that CNV prevalence varies across the European, African, and South Asian ancestry groups in the UK Biobank (UKBB). We now replicate these findings in SPARK, a United States cohort of individuals diagnosed with autism spectrum disorder (ASD) and their families, and we extend our analyses to examine the prevalence of recurrent CNVs in individuals of admixed American ancestry. We used PennCNV and QuantiSNP to call CNVs (≥ 50 kB pairs) for European (EUR; *n* =46,971), African (AFR; *n* = 3,697), and admixed American (AMR; *n* = 7,882) SPARK participants whose genetic ancestry we inferred using KING. Then, we subsampled two cohorts of EUR individuals who were propensity-score matched on age, sex, and ASD diagnosis to either the AFR cohort (EUR-AFR; *n* = 3,697) or the AMR cohort (EUR-AMR; *n* = 7,882). We calculated odds as the number of recurrent CNV carriers divided by the number of non-carriers for each of 11 recurrent CNVs that we previously selected for analyses in the UKBB. Then, we calculated AFR odds ratios (ORs) as the AFR odds / EUR-AFR odds and AMR ORs as AMR odds / EUR-AMR odds. Despite smaller sample sizes and ascertainment differences, we replicated our previous findings of reduced prevalence of *CRYL1* DEL (AFR OR = 0; 95% confidence interval:[0, 0.693]), 15q13.1 BP4.5-BP5 DUP (AFR OR = 0.248; [0.0988, 0.551]), and 16p13.11 DUP (AFR OR = 0.315; [0.103, 0.82]) as well as increased prevalence of 2q13 *NPHP1* DUP (AFR OR = 17.1; [2.67, 711]) in AFR- as compared to EUR-ancestry individuals. Additionally, the *NRXN1* DEL was less prevalent in the SPARK AFR cohort than in the matched EUR cohort (AFR OR = 0.055; [0.001, 0.351]). We also found that 4 of the 11 recurrent CNVs were less prevalent in the SPARK AMR cohort as compared to the matched EUR cohort: 15q11.2 DEL (AMR OR = 0.546; [0.334, 0.875]), *NRXN1* DEL (AMR OR = 0.436; [0.215, 0.842]), 16p13.11 DUP (AMR OR = 0.512; [0.278, 0.915]), and 1q21.1 TAR DUP (AMR OR = 0.071; [0.002, 0.469]). By establishing that ancestry-related differences in CNV prevalence are present in both unselected community populations (UKBB) and cohorts enriched with ASD-diagnosed individuals (SPARK), our results suggest that genetic ancestry should be a key consideration when probing associations between CNVs and neurodevelopmental outcomes.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4104 Cox proportional hazards mixed model enables accurate estimation of heritability for time to event outcomes

Authors:

K. Taraszka[1], S. Sankararaman[2], A. Gusev[1]; [1]Dana-Farber Cancer Inst., Boston, MA, [2]UCLA, Los Angeles, CA

Heritability is defined as the variance in disease susceptibility attributed to genetics and is a fundamental parameter for interpreting disease architecture and polygenic prediction. Multiple methods have been developed for estimating the heritability of continuous or case-control phenotypes (under the liability threshold model, LTM), however, the heritability of time-to-event (TTE) outcomes has been largely unstudied. This has limited the characterization of age-of-onset and treatment response phenotypes, the latter particularly relevant for clinical studies. Here, we propose and evaluate a Cox proportional hazard mixed model (COXMM) to enable accurate estimation of heritability for TTE phenotypes with censoring. We apply these methods to real data from the UK Biobank.

We implemented an efficient COXMM with a random effect modeled by the genetic relatedness across individuals, and benchmarked its performance in extensive simulations. We simulated phenotypes according to generative models reflecting either LTM or TTE models. For TTE phenotypes, data followed various Weibull distributions with independent censoring. Across TTE simulations, COXMM produced unbiased estimates of heritability while classic case-control heritability estimators (Haseman-Elston regression, HE-reg under the LTM) showed significant downward bias. Likewise, restricting only to cases and estimating heritability of age-of-onset as a normalized continuous phenotype with HE-reg produced biased estimates, due to the artificial exclusion of controls. For example, for an exponential distributed age-of-onset with heritability of 0.8 and 40% of individuals in the study being cases, COXMM inferred heritability was 0.80 (0.007), compared to 0.32 (0.004) for case-control HE-reg, and 0.09 (0.02) for case-only age-of-onset HE-reg.

In the UK Biobank, we analyzed ten cardiovascular related age-of-onset traits. Interestingly, we found that the TTE estimate was higher for eight traits while the other two traits better reflect the LTM. For example, myocardial infarction (N=275,841 with 6% as cases), the COXMM heritability was 0.38 compared to 0.05 for case-control HE-reg, and 0.26 for case-only age-of-onset HE-reg. Ultimately, disease susceptibility is a well studied phenomenon while the heritability of TTE phenotypes such as age-of-onset or progression remain understudied due to poor performance of existing methods.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4105 Craniofacial enhancers are enriched for *de novo* mutations in trios with orofacial clefts

Authors:

S. Curtis[1], J. Murray[2], E. Feingold[3], T. Beaty[4], S. Weinberg[3], M. Marazita[5], M. Epstein[1], D. Cutler[1], J. Cotney[6], A. Visel[7], E. Leslie[1]; [1]Emory Univ., Atlanta, GA, [2]U of Iowa, Iowa City, IA, [3]Univ of Pittsburgh, Pittsburgh, PA, [4]Johns Hopkins Univ, Sch PubHlth, Baltimore, MD, [5]Univ Pittsburgh, Ctr. for Craniofacial and Dental Genetics, Pittsburgh, PA, [6]UConn Hlth., Farmington, CT, [7]Lawrence Berkeley Natl Lab., Berkeley, CA

Identifying the contribution of non-coding variants to human disease is challenging due to the increased number of genetic variants relative to coding variants and the cell-type and timepoint specific function of non-coding elements, which are incompletely catalogued. Although it is well established that common genetic variants associated with human diseases are enriched in non-coding regions of the genome, less is knowns about the role of rare non-coding variants, including de novo mutations (DNMs). We investigated the role of non-coding DNMs in a cohort of 1,114 trios with orofacial clefts (OFCs), a common birth defect, sequenced by the Gabriella Miller Kids First Research Program. We developed a novel algorithm to estimate the expected mutation rate in any defined region of the genome based on mutation rates from gnomAD, and validated its performance for coding regions using existing software tools ($r = 0.88$, $p<2.2x10-16$). We then called 73,284 high confidence DNMs and curated a set of 316,823 enhancers active in embryonic human craniofacial tissue. A total of 10,346 unique DNMs occurred in 8,421 of these enhancer regions. After correcting for multiple tests, 8 enhancers showed a significant burden of DNMs ($p < 5.9$ x 10-6). The enhancer with the greatest enrichment of DNMs was a 250kb super enhancer region with 20 DNMs ($p=1.35$ x 10-7), located 200kb upstream of PTK2, a tyrosine kinase (FAK) involved in focal adhesions between growing cells. While PTK2 is ubiquitously expressed, the conditional knockout in neural crest cells in mice was previously found to cause cleft palate and other structural anomalies. We next tested whether the DNMs in the 8 significant enhancers disrupted any transcription factor (TF) binding sites. 62 of the DNMs were predicted to alter the binding of TFs, with 144 TF binding sites predicted to be destroyed in the presence of the mutation, including binding sites for TFAP2A, TWIST1, SOX9, and ZBTB7A, all TFs known to be involved in craniofacial development and/or OFCs. To demonstrate tissue relevance, we repeated the analysis using DNMs in 301 healthy trios from the 1000 Genomes Project and >21 million non-craniofacial enhancers from the Roadmap Epigenomics Consortium. None of the OFC-enriched enhancers were enriched for DNMs in the control trios. Six of the Roadmap enhancers were enriched for DNMs in OFC trios ($p < 5$ x 10-8), and further analysis showed they overlapped enhancers active in fetal and craniofacial development. Our overall results demonstrate that craniofacial enhancers are enriched for DNMs found in OFC trios and suggests a mechanism by which these variants could alter gene regulation and affect developmental pathways leading to OFCs.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4106 Creating a Simulation Framework and Testing Methods to Adjust for Fine-Scale Population Structure in Rare Variant Analyses

Authors:

K. Marker[1], R. Shemirani[2], M. Lin[1], E. Kenny[2], G. Belbin[3], C. Gignoux[1]; [1]Univ. of Colorado, Anschutz Med. Campus, Aurora, CO, [2]Icahn Sch. of Med. at Mount Sinai, New York, NY, [3]Gencove Inc., New York, NY

When there is a difference in average trait value or disease prevalence between subpopulations, fine-scale genetic differences in rare variants can induce population structure that cannot always be guaranteed to be corrected with traditional approaches. This is increasingly a problem as very large sample sizes are needed to detect rare variant associations. Fine-scale ancestry differences between cohorts will induce population structure bias that can confound rare variant association studies if the fine-scale population structure is not properly adjusted for. We aim to compare four methods for capturing and adjusting for fine-scale population structure under varying phenotypic and environmental structure to identify the best method/s for various scenarios. These methods include identity-by-descent (IBD) clusters, kinship clusters, rare variant PCA, and Uniform Manifold Approximation and Projection (UMAP) projections. We developed a genetic simulation framework to generate 9 populations from a continental European demographic model with stepping-stone migration using msprime and tskit. We first simulate four outer populations from an ancestral population, and the inner populations are admixed from those original four populations and add continuous migration between adjacent populations. For our phenotype simulations we extend prior work on simulating ancestry-by-environment interactions (APRICOT). We extend the method to allow for complex multi-way admixture, in addition to adding isolation by distance models. This allows us to model and control the fine-scale population structure to test various confounding scenarios. Using this simulation framework, we run rare variant association tests to compare adjustment by IBD clusters, kinship clusters, rare variant PCA, and UMAP projections under varying phenotypic structures.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4107 Credible set is sensitive to imputation quality and missing variants

Authors:

Y. Liang, 23andMe Research Team, A. Auton, X. Wang; 23andMe, Sunnyvale, CA

Performing Bayesian fine-mapping to obtain credible sets has been widely applied in GWAS to pinpoint causal variants. With most GWASs relying on genotype imputation, the imputed genotypes are often represented by genotype dosages that may contain errors resulting from imperfect imputation. Besides, the imputation panel may not include the causal variants. However, these uncertainties resulting from imputation are usually not taken into account in downstream fine-mapping analyses.

In this work, we consider a common use case in which a GWAS study is performed using a minority of genotype calls that originate from a genotyping chip and a majority of variants that are imputed as genotype dosages using a reference whole genome sequencing panel. We designed a simulation study to look into the effect of genotype imputation on the quality of fine-mapping results. In particular, we focused on the observed coverage of the credible set and how well the credible set could tag a missing causal variant. We found that the coverage is expected or better when the association p-value is around 1e-8 to 1e-10 or the causal variant is rare. In contrast, when the association p-value < 1e-10 and the causal variant is common, the observed coverage is significantly smaller than the expected coverage as a result of attenuation due to imputation error and linkage disequilibrium. We further show that this miscalibration is due to the fact that the posterior inclusion probabilities overly concentrate on nearby variants that are in high LD with the causal variant while the posterior probability of the causal variants is low. This miscalibration is exacerbated by low imputation quality. To resolve this calibration problem, we explored potential post-GWAS solutions that explicitly take imputation quality into account by weighting the summary statistics on the basis of imputation quality and the true LD of the locus.

In summary, this work highlights the usefulness of including imputed variants in fine-mapping analyses and points out the potential miscalibration of fine-mapping results due to the uncertainty introduced by genotype imputation. To maximize the utility of the fine-mapping analysis in the context of genotyping in combination with genotype imputation workflow, we suggest methodological improvements for considering imputation error in fine-mapping analyses in the future.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4108 † Cross-ancestry genetic architecture influences estimation of heritability and genetic correlation in diverse populations

Authors:

S-K. Jang[1], N. Zaitlen[2], TOPMed consortium, S. Vrieze[3], S. Basu[3]; [1]Univ. of California, Los Angeles, Los Angeles, CA, [2]UCLA, Los Angeles, CA, [3]Univ. of Minnesota, Minneapolis, MN

With global initiatives for diverse genomic research, diverse genetic data will be increasingly available. Understanding the extent to which genetic architecture overlaps across populations and its impact on SNP-based heritability (h^2_{SNP}) and cross-ancestry genetic correlation are important to leverage these datasets to advance the genetic etiology of complex traits. Here we examined behaviors of different analytic approaches in estimating h^2_{SNP} and cross-ancestry genetic correlation estimates in diverse populations, considering various cross-ancestry genetic architecture scenarios with varying allele frequency (AF) and allelic effects of causal variants across ancestries.Using up to 55,724 diverse whole-genome sequence data from the TOPMed consortium, we compared four approaches and different ways to standardize genotype in simulation and real anthropometric phenotype analysis: 1) single ancestry approach: Genomic Restricted Maximum Likelihood (GREML) /Haseman-Elston (HE) regression applied to individual ancestry, 2) combined ancestry approach: GREML/HE regression applied to ancestry- combined sample, 3) bivariate approach: bivariate GREML to estimate h^2_{SNP} and cross-ancestry genetic correlation, and 4) GxE approach: GxEMM to estimate h^2_{SNP} for shared and ancestry-specific genetic effects. In the simulation, we found that heterogeneity in AF and allelic effects significantly affect heritability and genetic correlation estimates. Enriching causal variants in variants with lower cross-ancestry AF difference than that of randomly chosen SNPs led to a downward bias in h^2_{SNP} estimates due to the overall low LD level of these variants. Additionally, enriching causal variants in variants with larger cross-ancestry AF differences led to an underestimation of cross-ancestry genetic correlation across all ancestry pairs. Overall, bivariate and GxE approaches using within-ancestry standardized GRM yielded robust estimates in a relatively wide range of cross-ancestry architectures than other methods. In real phenotype analysis, we observed that h^2_{SNP} of height in the AFR population is enriched in variants with larger AF differences with European ancestry even after accounting for overall allele frequency and LD level in individual ancestry (h^2_{SNP} estimates=.10 (.04) v.s., .19 (.04) for smaller v.s. larger AF difference). This study provides guidance for estimating h^2_{SNP} and cross-ancestry genetic correlation using diverse ancestry data and highlights the importance of considering cross-ancestry genetic architectures in interpreting the results of existing methods.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4109 Data-driven trait heritability-based extraction of human facial phenotypes in genome-wide association studies

Authors:

M. Yuan[1], S. Goovaerts[1], H. Hoskens[2], O. D. Klein[3], R. Spritz[4], B. Hallgrimsson[2], S. Richmond[5], S. Walsh[6], M. D. Shriver[7], J. R. Shaffer[8], S. M. Weinberg[8], H. Peeters[1], P. Claes[1]; [1]KU Leuven, Leuven, Belgium, [2]Univ. of Calgary, Calgary, AB, Canada, [3]Univ. of California, San Francisco, San Francisco, CA, [4]Univ Colorado Hlth Sci Ctr, Aurora, CO, [5]Cardiff Univ., Cardiff, United Kingdom, [6]Indiana Univ. Purdue Univ. Indianapolis, Indianapolis, IN, [7]Penn State Univ, University Park, PA, [8]Univ. of Pittsburgh, Pittsburgh, PA

Genome-wide association studies (GWAS) of complex, multi-dimensional traits, such as the human face, typically rely on predefined and simplified phenotypic measurements like inter-landmark distances and angles. These measures are predominantly designed by human experts based on biological or clinical knowledge. To circumvent manual crafting, alternative phenotypic descriptors, such as features derived from dimension reduction techniques (e.g., principal component analysis), are employed. While the features generated by computational algorithms capture the geometric variations of the biological shape, they are not necessarily genetically relevant. Therefore, genetically informed data-driven phenotyping is desirable. Here, we propose an approach where phenotyping is done by a data-driven optimization of trait heritability or estimates of the degree of variation in a phenotypic trait in a population that is due to genetic variation. The resulting phenotyping process consists of two steps: the first step involves constructing a feature-space that models shape variations using dimension reduction techniques, while the second step entails searching for directions in the feature-space where trait-heritability are high using a genetic search algorithm. This heritability-dependent optimization enables the extraction of phenotypes assumed to have a higher genetic value. We trained and validated our proposed phenotyping process using 3D facial surface scans of father-offspring pairs (n=770) and an independent cohort of unrelated European individuals (n=8,426) with genome-wide SNP data, respectively. The phenotypes resulting from the heritability-optimized training demonstrated higher mean trait heritability (0.399+/-0.088) compared to that of principal components (0.265+/-0.126), as expected. More notably, trait-heritability optimized phenotypes exhibited higher SNP-based heritability (0.36+/-0.062) compared to principal components (0.214+/-0.087). Furthermore, a smaller number (n=30) of effective heritability-optimized traits was sufficient to identify the same number of independent genetic loci (n=11) compared to using all principal components (n=70), that together explain 98% of the facial shape variation. Our results demonstrate that data-driven trait heritability-based optimization enables the automatic extraction of genetically relevant phenotypes, shown by their increased power in GWAS.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4110 *De novo* germline variants associated with congenital heart disease and neuroblastoma: Evidence of pleiotropic genetic effects from 1,320 parent-offspring trios.

Authors:

A. Renwick[1], Y. Yu[2], R. Harris[1], P. J. Lupo[1], C. Huff[2]; [1]Baylor Coll. of Med., Houston, TX, [2]Univ. of Texas M.D. Anderson Cancer Ctr., Houston, TX

There is emerging evidence that children with congenital heart disease (CHD) are more likely to develop neuroblastoma when compared to children without these conditions. While there are some lines of evidence to support these associations (e.g., common cells of origin), there have been few efforts to evaluate the genetic overlap between these phenotypes. Therefore, we evaluated pleiotropic effects of *de novo* variants in CHD and neuroblastoma parent-offspring trios.

Whole genome sequencing (WGS) data on CHD trios (n=711) and neuroblastoma trios (n=609) was obtained from the Gabriella Miller Kids First Data Resource Center as trio VCFs. Our filtering criteria to identify *de novo* variants included the number of high-quality reads, the fraction of alternative alleles, and likelihood ratio in both parents and probands. We excluded common variants and variants from genomic repeat regions. Following filtering, we tested for an enrichment of *de novo* single nucleotide coding variants in each cohort using the VARPRISM software. We then computed an omnibus p-value across the two cohorts using Fisher's method, and then manually inspected the BAM alignments of variants contributing to significant enrichments. We also incorporated findings from a previously reported exome study of the CHD cohort.

Probands in the neuroblastoma cohort were 54% male and 81% Non-Hispanic White, with a median age at diagnosis of 16 months. Probands in the CHD cohort were 59% male and 60% Non-Hispanic White. The most frequent CHD phenotypes were ventricular septal defects (46%), tetralogy of Fallot (32%), hypoplastic left heart syndrome (28%), atrial septal defects (25%), and right aortic arch (25%). In our trio-based analysis, we tested 18,662 genes and found six genes were nominally (but not genome-wide) significant for an enrichment of *de novo* variants in both cohorts: *CIC* ($p=2.2x10^{-3}$), *FAM76A* ($p=2.3x10^{-4}$), *IREB2* ($p=6.4x10^{-4}$), *PCDHGB5* ($p=3.8x10^{-4}$), *PELO* ($p=1.5x10^{-3}$), and *POGZ* ($p=8.2x10^{-6}$).

While our results do not suggest that de novo variants play a strong role in the overlap between CHD and neuroblastoma, we did identify genes that are involved in development and carcinogenesis. For example, *FAM76A*, *IREB2*, *PCDHGB5*, and *PELO* are all involved in gene regulation, while *CIC* is involved in between-cell communication. Additionally, *CIC*, *IREB2*, *PCDHGB5*, and *POGZ* are known to be involved in neurodevelopmental disorders, phenotypes which have been reported among children with CHD and neuroblastoma. Our assessment demonstrates the utility of evaluating the genetic underpinnings of these associated conditions.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4111 Decoding the genetic architecture of hearing acuity in a family-based design

Authors:

J. Duran[1], S. Mansour[1], D. Neklason[2], L. Jorde[1]; [1]Univ. of Utah Dept. of Human Genetics, Salt Lake City, UT, [2]Univ of Utah Dept. of Internal Med., Salt Lake City, UT

Genetic studies of hearing impairment and deafness have revealed over 223 genes that negatively impact hearing. The genetics of natural variation in hearing acuity is less studied and poorly understood. Identifying genetic variation responsible for an individual's hearing ability would provide future avenues to explore novel targets for treating hearing loss. In addition, it may identify genetic variation that contributes to exceptionally acute hearing and offer insight into protective variants. We report the first pedigree-based assessment of heritability and genetic correlations of hearing ability across a range of frequencies. Whole-genome sequencing and audiometric data from 39 three- and four-generation Utah CEPH families were used to examine the genetic architecture of hearing acuity. The narrow-sense heritability of hearing acuity within the CEPH cohort decreased from 56% at 250Hz to 35% at 8000Hz. The single nucleotide polymorphism (SNP) heritability decreased from 35% to 15% across the same frequency range. These patterns are consistent with a larger non-genetic component for high-frequency hearing acuity. To explain the effects of familial transmission on SNP heritability we estimated the SNP heritability of unrelated individuals in our cohort. As expected, the SNP heritability of unrelated individuals was lower than that of related individuals, with an average decrease of 8%, $p = 0.01$. A bivariate analysis was performed to understand the genetic correlation between any two frequencies. We found that the closer frequencies are in pitch, the more significant the genetic correlation between those frequencies, suggesting distinct genetic effects contribute differentially to low- and high-frequency acuity. Previous studies of overall hearing acuity, based on twin data, have provided heritability estimates that may be upwardly biased. Thus far they have been limited to values averaged across frequencies. A Swedish twin study estimated the narrow-sense heritability averaged across four high frequencies at 75%. Similarly, a twin study of U.S. veterans reported hearing acuity heritability upper limits of 73% for mid-range frequencies and 72% for high frequencies. By comparison, our family-based heritability estimates are 51% for mid-level frequencies and 43% for high frequencies. This difference suggests that previous studies may have overestimated the genetic contribution to hearing acuity, and highlights the importance of considering frequency-specific differences. Continued work on the genetic architecture of hearing acuity may open avenues for investigating protective variation and novel targets to treat hearing impairment.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4112 Deep multilayer proteomics and multi-omics integration reveal molecular networks related to Alzheimer's disease in diabetic brains

Authors:

V. Sarsani[1], A. Capuano[2], S. Tasaki[2], Z. Arvanitakis[2], L. Liang[1]; [1]Harvard T.H. Chan Sch. of Publ. Hlth., Boston, MA, [2]Rush Alzheimer's Disease Ctr., Chicago, IL

Type 2 diabetes mellitus (T2D) and Alzheimer's disease (AD) are two complex diseases that are prevalent in the aging population. Recent research suggests that insulin resistance, a key feature of T2D, is associated with AD pathology, including amyloid-β deposition and cognitive impairment. To better understand the molecular mechanisms of insulin resistance in the brain and periphery and its relationship to AD neuropathology and cognitive function, we performed a phosphoproteome and proteome profiling study using a multi-omics integration approach. We examined 192 post-mortem brain and muscle samples from individuals with and without T2D and with varying levels of AD pathology. We quantified 11,726 phosphopeptides and 29,955 non-phosphopeptides from 3,537 protein groups in frozen brain tissue (dorsolateral prefrontal cortex). In muscle, we quantified 7,801 phosphopeptides and 14,940 non-phosphopeptides from 1,383 protein groups. Our analysis revealed that subjects with higher levels of AD pathology exhibited hyperphosphorylation in proteins like MAPT, regardless of T2D status. Diabetic proteome exhibited hyperphosphorylation in c*ytochrome* oxidases subunit proteins like NDUFA4 and COX7B. We also identified 105 phosphorylation changes in proteins, including MAP2, SLC43A2, and GIT1, that were common to both AD and T2D. We then integrated GWAS, methylation, metabolomics, and transcriptomic profiles of matched samples to build predictive omics signatures using a novel semi-supervised machine learning (ML) framework. To examine the generalizability and utility of our results, we tested our framework-generated signatures on primarily peripheral blood omics data from AD and T2D subjects in UK Biobank. In conclusion, with our multilayer proteomics data and novel ML omics integration framework, we built predictive signatures and molecular network maps of AD and T2D phosphorylation signaling events, thus filling knowledge gaps in metabolism, insulin signaling, and AD research.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4113 DeepRVAT: Joint modelling of rare variant genetic effects using deep learning and data-driven burden scores.

Authors:

E. Holtkamp[1], B. Clarke[2], H. Öztürk[2], F. Brechtmann[1], F. Hölzlwimmer[1], J. Gagneur[1,3,4], O. Stegle[2,5,6]; [1]Technical Univ. of Munich, Garching bei München, Germany, [2]German Cancer Res. Ctr., Heidelberg, Germany, [3]Helmholtz Ctr. Munich, Neuherberg, Germany, [4]Inst. of Human Genetics, Munich, Germany, [5]European Molecular Biology Lab., Heidelberg, Germany, [6]Wellcome Sanger Inst., Hinxton, United Kingdom

Population-scale genome sequencing data provides unprecedented opportunities to survey the effect of rare variants on human traits, with relevance for effector gene detection and extreme phenotype predictions. Existing methods for rare variant association genetics require *a priori* assumptions about the variant categories that exhibit phenotypic effects, thus limiting their efficacy. Recent variance-based methods incorporating multiple annotations lack the ability to jointly consider all variant annotations, instead performing post-hoc combination of individual tests into meta-models.

Here, we propose DeepRVAT (Deep Rare Variant Association Testing), an end-to-end deep-learning model that jointly estimates gene impairment from rare variants and annotations to explain trait variation. DeepRVAT employs a set neural network architecture to flexibly handle varying numbers of rare variants per gene and individual and learns an unbiased, gene- and phenotype-agnostic rare gene impairment scoring function in a data-driven manner. It leverages rich variant annotations, can score variants unseen at training time, accounts for nonlinear and epistatic effects, and can be combined with classical statistical tests to maintain rigorous control over type I error rates.

We first applied DeepRVAT to exome sequencing data for 21 quantitative UK Biobank (UKBB) traits (N=167k). Using the generic DeepRVAT gene impairment score yields substantial power gains for gene discovery, from 9-27% compared to meta-modelling methods that use an entire panel of mixed effect tests, up to 124% compared to the methods most commonly used in practice. Strikingly, these gains come together with improved replication rates in hold-out data.

Moreover, unlike meta-models, which have limited applicability beyond association testing, DeepRVAT gene impairment scores empower the creation of polygenic rare variant risk scores. By complementing common variant polygenic risk scores (PRS) with DeepRVAT scores to stratify high-risk individuals (extreme one-percentiles) in the UKBB cohort, we achieve a relative improvement of up to 52% in AUPRC compared to using the common PRS alone, thereby significantly outperforming recently proposed rare variant PRS strategies based on single annotations. Altogether, using deep learning and rich variant annotations, DeepRVAT summarises rare variant effects into a generic gene impairment score, enabling substantial improvements to multiple downstream applications such as gene-based association testing and high-risk individual stratification.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4114 Deriving computationally-defined latent phenotypes from medical images and genetic data.

Authors:

J. Wang[1], R. Border[2], O. Avram[1], B. Durmus[1], J. Mefford[1], E. Halperin[2], S. Sankararaman[1], N. Zaitlen[1]; [1]UCLA, Los Angeles, CA, [2]Univ. of California Los Angeles, Los Angeles, CA

Medical images are a fundamental component of clinical care and contain a large number of basic and informative biomarkers. Many of these image-derived biomarkers are defined a priori by domain experts and measure clear physical quantities such as surface areas or volumes of various brain regions. However, these quantities are not necessarily the most predictive of desired outcomes, nor are they the most genetic. Indeed, SNP-based heritability estimates have been computed for many of these traits, but even with large cohorts, most studies have produced estimates less than half of those from twin studies. Among the potential drivers of this missing heritability and the ensuing difficulty of identifying causal variants is one of particular interest: extensive genetic overlaps between related image phenotypes. Multivariate modeling approaches have emerged as an effective way of mitigating this problem, but this strategy is not entirely satisfactory, as multivariate methods sacrifice computational efficiency and interpretability.

Here, we propose a framework to define and compute univariate latent phenotypes which are more heritable and have a more discoverable architecture. This is accomplished by using recent extensions of convolutional neural networks in two ways. In the first, we predict externally defined phenotypes such as BMI or psychiatric disorders from organ scans, thus obtaining the portions of these broad phenotypes related to particular organs; here, our objective function is squared error loss against the true phenotype. In the second, we output, in an unsupervised fashion, a phenotype which is maximally heritable; here, our objective function is the Haseman-Elston regression estimate of heritability. To support the latter approach, we verify that heritability is a tractable objective function by recovering geometric features from simulated images and genetic data, without prior knowledge of said features. To support the former approach, we obtain good estimates of BMI (R2=0.82) from UK Biobank liver MRIs (n=7230) and perform comparative GWAS analyses against the true BMI, identifying rs1642763 and rs74543191 as SNPs which may be related to BMI through liver-specific mechanisms (waist circumference and diabetes-associated liver disease, respectively). This evidence suggests that our method may be a useful alternative for genetic analysis of medical images.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4115 Detecting D4Z4 repeat contraction with 4qA haplotypes from short-read whole genome sequences

Authors:

J. Choi[1,2,3], O. D. King[4], L. M. Kunkel[1,2], E. A. Lee[1,2,3]; [1]Boston Children's Hosp., Boston, MA, [2]Harvard Med. Sch., Boston, MA, [3]The Broad Inst. of Harvard and MIT, Cambridge, MA, [4]Univ. of Massachusetts Chan Med. Sch., Worcester, MA

Facioscapulohumeral Muscular Dystrophy 1 (FSHD1) is a common form of muscular dystrophy caused by the reduced copy (1-10 copies) of D4Z4 tandem repeats at 4q35, along with a 4qA polymorphism distal to the repeat. D4Z4 repeats have a long repeat unit (RU) of 3.3 Kbp that conventional short sequencing reads (100-150 bp) cannot cover. Additionally, there are two alleles distal to D4Z4 (4qA and 4qB), only one of which (4qA) is pathogenic leading to FSHD1 with D4Z4 contraction. Moreover, there are nearly identical D4Z4 repeats on chromosome 10, which do not contribute to FSHD1 pathogenicity. Thus, it is extremely challenging to detect pathogenic contraction of D4Z4 repeats using short-read whole-genome sequencing (WGS) data.

To address these challenges, we propose a new method that utilizes two different types of loci: chromosome separating loci (CSL) and allele separating loci (ASL). These loci have different genotypes according to their chromosomes and distal alleles (4qA vs 4qB), allowing us to estimate the RUs of D4Z4 repeats on the 4qA haplotype. To evaluate our method, we applied it to WGS data from 34 FSHD1 patients and compared the estimated RUs with those measured through classical western blots. Additionally, we applied our method to nine samples with both short-read WGS and Bionano data where single-molecule optical mapping enables precise measurement of D4Z4 RUs. We observed a high correlation between our estimated RUs and those independent measurements confirming the accuracy of our method. Then, we applied our method to ~30X WGS data of 2,504 individuals across 26 populations from the 1000 Genome Project (1KGP). Our analysis revealed population-specific characteristics of D4Z4 repeats. Notably, African populations exhibited a lower frequency of the 4qB haplotype that has recently appeared in the evolution, consistent with the literature. Furthermore, we identified 126 asymptomatic individuals from the 1KGP cohort with predicted pathogenic D4Z4 contraction by our method. Through our preliminary Genome-wide Association Study (GWAS) analysis, we identified two clusters of variants enriched in those asymptomatic carriers. Although more investigation is warranted, this approach might be useful to identify potential disease-modifying or protective variants against FSHD1.

Overall, our study presents a novel methodology to estimate D4Z4 repeat units on a pathogenic haplotype from short-read WGS data and provides a comprehensive characterization of D4Z4 repeats across different populations. These findings enhance our understanding of genetic architecture of FSHD1 and may offer insights into potential protective variants against the disease.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4116 Detecting Perturbation Effects of Gene Expression in Single-cell CRISPR Screening

Authors:

X. Wang, M. Lalli, J. Buxbaum; Icahn Sch. of Med. at Mount Sinai, New York, NY

We present a novel method for detecting perturbation effects of gene expression in single-cell CRISPR screening using low MOI CRISPR data. Our method utilizes subspace perturbation modeling to capture the effects of CRISPR perturbations and identifies consistent cell type shifts for a subset of Autism risk genes. In our study, we performed single-cell RNA-sequencing (scRNA-seq) on human-induced neural progenitor cells (iNPCs) that were subjected to CRISPR perturbations at a low multiplicity of infection. By analyzing the scRNA-seq data, we observed distinct subgroups of cells, including glial progenitor cells, neuronal progenitor cells, and maturing neurons. However, these changes in cell compositions can obscure the effects of CRISPR perturbations on gene expression. To overcome this challenge, we developed a method that separates the effects on cell composition from the effects on gene expression. By modeling the subspace perturbations, we were able to capture the specific changes in gene expression caused by CRISPR perturbations. Our method detected consistent shifts in cell types for a subset of Autism risk genes, providing valuable insights into their functional roles. Overall, our method enables the detection of perturbation effects of gene expression in single-cell CRISPR screening, even with low MOI CRISPR data. By utilizing subspace perturbation modeling, we can uncover the underlying effects of CRISPR perturbations and identify consistent cell type shifts for Autism risk genes. This approach contributes to our understanding of gene function in the context of neuro-psychiatric disorders.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4117 Developing prediction model for the age at natural menopause with the population-based Health Examinee cohort study

Authors:

M. Yuk, S. Yang, K. Oh, Y. Kwak, J. Lee, J. Youn, N. Song; Coll. of Pharmacy, Chungbuk Natl. Univ., Cheongju, Korea, Republic of

BACKGROUND: Most women experience menopausal transition inducing various physical and emotional changes that could even cause diseases. Hence, predicting the timing of menopause helps women prepare and cope menopausal transition-related diseases. In this study, we developed a prediction model for age at natural menopause and assessed its predictive performance.
METHODS: The study subject was drawn from the Korean Genome and Epidemiology Study_Health Examinee (KoGES_HEXA) study. Among women with genome-wide single-nucleotide polymorphism (SNP) data, after excluding women who experienced natural or artificial menopause at baseline of the study, total 10,729 (4,009 postmenopausal, 6,720 premenopausal) women who also had the follow-up data were included in the analysis. Cox proportional hazards regression, accelerated failure time (AFT), artificial neural network (ANN), XGBoost, and support vector machine (SVM) models were applied to determine the optimal model to predict age at natural menopause. The 52 variables of health screening-related data including anthropometrics, reproductive health, biochemistry, and diseases history, and 41 SNPs (31 known and 10 identified by genome-wide association studies) were used as predictors for age at natural menopause. The performance of five models was evaluated using mean absolute error (MAE) and R-squared (R^2) for all models and additional concordance index (C-index) for time-dependent models.
RESULTS: Cox model was presented as the most appropriate model for predicting age at natural menopause with the lowest MAE and the highest R^2 (MAE=0.71, R^2=0.86, AFT model; 2.07, 0.30, ANN; 1.16, 0.78, XGBoost; 1.12, 0.79, SVM; 1.24, 0.71) among the five models. In this time-dependent model, mean corpuscular hemoglobin concentration (MCHC) had the most significant effect on the age at natural menopause (HR=4.22, CI=3.67-4.86). When the Cox model was also evaluated with the genetic biomarkers, the predictive performance was better compared to the baseline model without genetic biomarkers (Harrell's C-index difference *P*=0.001).
CONCLUSION: The time-dependent Cox model was developed to predict age at natural menopause based on clinical and/or genetic data from the population-based Health Examinee cohort study. This time-dependent prediction model for age at natural menopause would be utilized to support making clinical and personal decisions related to menopausal transitions.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4118 † Differential cell-type-specific gene expression by type 2 diabetes status in human skeletal muscle

Authors:

D. Ciotlos[1], S. Hanks[1], A. Varshney[1], M. Erdos[2], N. Manickam[1], A. Jackson[1], H. Stringham[1], N. Narisu[2], L. Bonnycastle[2], P. Orchard[1], M. Laakso[3,4], J. Tuomilehto[5], T. Lakka[3], K. Mohlke[6], M. Boehnke[1], H. Koistinen[5], F. Collins[2], S. Parker[1], L. Scott[1]; [1]Univ. of Michigan, Ann Arbor, MI, [2]NHGRI/NIH, Bethesda, MD, [3]Univ. of Eastern Finland, Kuopio, Finland, [4]Kuopio Univ. Hosp., Kuopio, Finland, [5]Finnish Inst. for Hlth.& Welfare, Helsinki, Finland, [6]Univ North Carolina, Chapel Hill, NC

Skeletal muscle tissue plays a role in regulation of body glucose, and skeletal muscle insulin resistance is a hallmark of type 2 diabetes (T2D). Previous skeletal muscle bulk RNA-seq studies have identified association of gene expression with T2D status; however, these studies could not distinguish gene expression changes that differed between fast twitch (types 2A and 2X) and slow twitch (type 1) muscle fiber types.
To understand the relationship between T2D status and gene expression in different cell types, we performed single nucleus RNA-sequencing of skeletal muscle biopsied from 282 Finnish adults (41.5% female, 25.9% T2D) from the FUSION Tissue Biopsy Study. We observed 12 main cell- and subcell- type clusters including muscle fiber types. We tested for associations between cell-type composition and T2D-related traits using a negative binomial model. Individuals with T2D, higher fasting glucose, or higher fasting insulin had greater proportions of neuromuscular junction nuclei than individuals with normal glucose tolerance (min p=0.0020).
For each cell type, we tested for associations between gene expression and T2D using a negative binomial model (n genes tested=23,766). In each muscle fiber type, we found 4 to 9 genes differentially expressed by T2D status. However, using gene set enrichment analysis, we observed 6.4-10.1% of 5641 gene ontology gene sets were enriched for different levels of gene expression by T2D status across muscle fiber types. For most but not all GO terms, muscle fiber cell types and bulk RNA-seq data show concordant enrichment of expression differences between individuals with T2D or normal glucose tolerance (NGT). In each muscle fiber type and bulk skeletal muscle, genes associated with cellular respiration were enriched for higher expression in individuals with NGT than those with T2D (OR=0.51-0.65, min p=6.7e-6). In contrast, in type 2X muscle fibers, genes associated with positive regulation of vasoconstriction were enriched for higher expression in individuals with T2D than those with NGT (OR=2.84, p=3.84e-4) but this difference was not observed in bulk tissue (OR=1.01, p=0.93).
Overall, T2D-associated differences in gene expression in bulk RNA-seq are detected in all muscle fiber types, indicating similar biological processes across muscle fiber types. However, we also identified gene sets for which enrichment of expression differences are only detected in one muscle fiber type and not in bulk skeletal muscle, suggesting potential fiber-type-specific biology.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4119 Discovery of core genes for systemic lupus erythematosus via genome-wide aggregation of*trans*-effects.

Authors:

A. Iakoviev[1], O. Castellini-Pérez[2], M. Alarcón-Riquelme[2], P. McKeigue[1], E. Carnero-Montoro[2], A. Spiliopoulou[1]; [1]Univ. of Edinburgh, Edinburgh, United Kingdom, [2]GENYO. Ctr. for Genomics and Oncological Res. Pfizer/Univ. of Granada/Andalusian Regional Government, Granada, Spain

Aim: Systemic lupus erythematosus (SLE) is an autoimmune disorder in which defective degradation and clearance of DNA lead to upregulation of Type 1 interferon signaling and a B-cell mediated autoimmune response. The objective of this study was to test whether the genetic architecture of SLE fits an "omnigenic" model in which the polygenic *trans*-effects of common SNPs coalesce on a relatively sparse set of "core" effector genes. **Methods**: Using expression and protein level quantitative trait loci (eQTL and pQTL) summary statistics from 31,684 samples in eQTLGen and 35,559 samples in deCODE respectively, we computed genome-wide aggregated *trans*-scores to predict gene expression and levels of circulating proteins in a case-control GWAS of SLE cohort consisting of 5267 SLE cases and 4909 controls. We identified putative core genes based on strength of association of the genome-wide *trans*-score with SLE, the effective number of *trans*-eQTLs contributing to the genome-wide score, and specificity of the association (lack of correlation with scores for other genes). **Results**: SLE was strongly associated with *trans* scores for at least 12 co-expressed Type 1 interferon-stimulated genes (ISGs); these genes share a common set of five *trans*-eQTLs, so that the aggregated *trans*-scores are correlated ($r > 0.8$). Of these ISGs, *MX1* has the strongest effect (Log OR=0.13, $P = 4 \times 10^{-10}$), *ISG15* (Log OR=0.1, $P = 1 \times 10^{-6}$) and *RABGAP1L* (Log OR=-0.1, $P = 6 \times 10^{-7}$) are known monogenic causes of SLE. The score for *STAT1* was also associated with SLE (Log OR=0.1, $P = 2 \times 10^{-6}$) but uncorrelated with detected ISGs. From pQTL analysis, *LGALS2* ($P = 2 \times 10^{-14}$), *TNFRSF17* ($P = 2 \times 10^{-17}$), *ALKBH2* ($P = 1 \times 10^{-12}$), *CRTAM* ($P = 2 \times 10^{-11}$) and *CD5L* ($P = 6 \times 10^{-15}$) were identified as putative core genes. **Conclusions**: Some genetic effects on SLE risk are mediated via *trans*-effects on basal expression of a cluster of ISGs. These do not appear to be mediated by effects on the JAK/STAT signaling pathway. Of the other putative core genes, *TNFRSF17* encodes a receptor for a therapeutic target for SLE. *LGALS2*, *CD5L* and *CRTAM* are in immune-related pathways, and *ALCBH2* is a DNA repair enzyme. These results support the conclusion of a recent study of type 1 diabetes that genome-wide aggregated *trans*-QTL analysis is able to identify core genes for disease that are not detected in a conventional SNP-by-SNP GWAS analysis.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4120 Discovery of kidney function loci and MR-based associated causal effects of lipid traits in 80k African ancestry individuals

Authors:

C. Kintu[1], S. Opeyemi[2], M. Richard[1], J. Daudi[3], C. Tinashe[4], F. Segun[5]; [1]MAKERERE Univ., Kampala, Uganda, [2]The African Computational Genomics (TACG) Res. Group, Entebbe, Uganda, [3]Makerere Univ. Coll. of Hlth.Sci., Kampala, Uganda, [4]The African Computational Genomics (TACG) Res. Group, Kampala, Uganda, [5]Dept. of Non-Communicable Disease Epidemiology, London Sch. of Hygiene and Tropical Med. (LSHTM), United Kingdom, London, United Kingdom

Background: Estimated glomerular filtration rate (eGFR) is a proxy for kidney function and lower eGFR (<60ml/min/1.73m^2) is indicative of chronic kidney disease (CKD). Although people of African ancestry (AA) have a greater burden of CKD, most genome-wide association studies (GWASs) have been conducted in Europeans. These findings might not be transferable to Africans due to genetic differences. We therefore set out to perform a meta-analysis of African-ancestry datasets to identify novel loci associated with CKD in Africans, and to determine the causal effects of lipid traits on kidney function in this population. **Methods:** We conducted a meta-analysis of creatinine-based eGFR in 80,027 AA participants from three global population-based cohorts: The Chronic Kidney Disease Genetics Consortium (CKDGen, n = 16474), UK Biobank (UKBB, n = 6217), and the Million Veteran Program (MVP, n = 57336). A fixed-effect inverse-variance weighted meta-analysis was performed using METAL software. We used the Bayesian fine-mapping approach to localize putatively causal variants. We further performed a Two-sample MR; instrument variables (IVs) for the predictor (lipid traits) were derived from a meta-analyzed African lipids GWAS (MALG, n = 24,215). The outcome IVs were computed from the MVP dataset. A random-effects inverse variance method was used as our primary analysis while adjusting for pleiotropy.
Results: The meta-analysis identified 403 significant Single Nucleotide Polymorphisms (SNPs) (at $p<5x10^{-8}$), 8 SNPs were lead variants and 3 of these were novel. The novel variants; rs200950799, rs10084572 and rs77408001 mapped to genes: Solute Carrier Family 15 member 5 (*SLC15A5*), 1-acylglycerol-3-phosphate O-acyltransferase 3 (AGPAT3), and Frizzled Class receptor 9 (*FZD9*) respectively. SLC15A5 and AGPAT3 are both diferentially expressed in the kidney-medulla and kidney-cortex. The rs200950799 variant was the putatively causal variant with a 71% posterior probability.
Our univariable MR analysis indicated a significant association between low-density lipoprotein (LDL) cholesterol and eGFR in African ancestry individuals (β= 1.1, 95% CI = [0.411-1.788]; p=0.002). The multivariable (MVIVW) method indicated similar evidence for LDL and TC with eGFR (β = 1.228 [0.477-1.979]; p=0.001) and (β = 1.357 [0.444-2.27]; p=0.004) respectively. A protective effect of Triglycerides on eGFR was observed in the MVIVW analysis (β = -1.412 [-2.714— -0.109]; p=0.034). **Conclusion:** We report ancestry-specific genetic variants associated with kidney function and show evidence of a possible causal effect of lipid traits on eGFRcrea in Africans.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4121 Disease-Aware Machine Learning Pipeline for Drug Target Prediction

Authors:

A. Stiemke[1], A. Musolf[1], B. Guo[1], T. Rajasekar[1], K. Shkura[2], J. Sul[1], S. Longerich[1]; [1]Merck & Co., Inc, Rahway, NJ, [2]MSD, London, United Kingdom

Identifying effective and safe drug targets in drug development is a long, arduous, and expensive process. Recently, machine learning (ML) techniques have been deployed for drug development leading to increased throughput and decreased failure rates of drug discovery. ML methods have been applied to identify and prioritize high quality, druggable gene targets for multiple diseases. However, these methods and their predictive features are usually built specifically for an indication; utility beyond the indication is limited. The goal of this study is the construction of a ML-based pipeline that can be utilized across diverse disease indications.

To create adaptability across disease, our pipeline takes advantage of disease-agnostic features like cellular location/function and druggability/tractability of targets. To improve predictability, we incorporate several disease-specific features which are automatically generated when applied to different diseases. One major component of our approach is incorporation of genetic evidence for diseases, which is largely absent from other ML methods. Genetic evidence is important because the probability of successfully developing a drug is much higher when the target is supported by human genetic evidence. This new ML pipeline containing disease-specific and disease-agnostic features is termed “disease-aware”.

We analyze our array of features through a new ML ensemble method. This ensemble approach uses 11 different machines (random forest, neural networks, etc.) to select the best predictive machine and achieves optimal prediction across many diseases. This overall model adapts and uses the appropriate features and best-fit machine based on the disease of interest. Simulations show that our method is robust and maintains high power while varying the number of causal features and noise factors. We demonstrate this ensemble method matches and surpasses the individual models and can be used effectively across disease types to identify new, high-confident druggable gene targets in multiple disease areas, such as glaucoma and autism.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4122 Disentangling the shared genetic aetiology of type 2 diabetes and schizophrenia.

Authors:

A. Arruda[1], L. M. Huckins[2], E. Zeggini[3]; [1]Helmholtz Zentrum München, Neuherberg, Germany, [2]Yale Univ. Sch. of Med., New Haven, CT, [3]Helmholtz Zentrum Munich, Neuherberg / Munich, Germany

Multimorbidity represents an increasing public health challenge with far-reaching implications for health management and policy. Among adults, the coexistence of metabolic and psychiatric diseases is a commonly observed comorbidity. In this study, we leverage summary statistics from large-scale trans-ancestry genome-wide association studies to investigate the genetic intersection between type 2 diabetes and schizophrenia, two positively epidemiologically associated conditions. We perform colocalization analysis and identify robust evidence of shared association signals at 11 genomic loci, six of which have opposing directions of effect for type 2 diabetes and schizophrenia. To elucidate these colocalizing signals, we integrate multi-omics data from bulk and single-cell data from disease-relevant tissues (pancreatic islets, adipose tissue, liver and brain), along with functional information. We identify high-confidence effector genes, each of which is supported by at least four orthogonal lines of evidence for their involvement in the diseases. The top-ranking effector gene is *NUS1*, which plays a role in lipid trafficking regulation. Mendelian randomization analysis suggests that increased expression of *NUS1* in the brain is causal for schizophrenia liability and protective against type 2 diabetes liability. This causal evidence is supported by data from differential expression analysis in the brain, indicating overexpression of *NUS1* in schizophrenia patients. The result also aligns with observations from *NUS1* knockout mice, displaying elevated levels of circulating triglycerides, which are associated with an increased risk of type 2 diabetes. Our findings offer insights into the underlying biological mechanisms contributing to the co-occurrence of type 2 diabetes and schizophrenia, shedding light on the complex nature of this comorbidity.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4123 Dissecting the shared aetiology of immunologically-driven disease through genetic correlation analysis

Authors:

M. Boyton[1], T. Gaunt[1], T. Richardson[2], V. Walker[1]; [1]Univ. of Bristol, Bristol, United Kingdom, [2]MRC Integrative Epidemiology Unit (IEU), Bristol, United Kingdom

Numerous comorbidities and altered risk profiles exist between prevalent immunologically-driven disorders in the observational literature. However, little is known of the underlying genetic basis for these diverging relationships between often complex phenotypes. We investigated the shared global and local genetic architecture of a spectrum of immune-associated diseases via Linkage Disequilibrium Score Regression (LDSC) and Local Analysis of (co)Variant Association (LAVA). In LDSC, the heritability and genome-wide genetic correlation between pairs of phenotypes is computed through cross-trait LD-score regression, whereas in LAVA, the bivariate analysis of region-specific genetic correlations is calculated. Our results demonstrate the existence of shared global and local genetic correlations between immunologically-relevant disorders, with diverging relationships spanning autoimmune, allergic and cancer phenotypes. Early results indicate a shared genetic basis for diseases such as Lupus and Rheumatoid Arthritis ($rG = 0.523$, $p < 0.001$) and Kidney Cancer and Sepsis ($rG = -0.892$, $p < 0.001$), as well as numerous loci of bivariate genetic importance spanning known inflammatory genes such as BACH2 and IL6R. Further investigation into the local genetic contributions to these relationships is ongoing, with the aim of identifying gene and protein instruments to guide drug target prioritisation and repurposing opportunities.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4124 Dissection of a known gene-physical activity interaction using metabolomics

Authors:

K. Westerman[1], M. Sevilla[1], C. Gu[2], T. O. Kilpelaïnen[3], J. I. Rotter[4], D. I. Chasman[5], A. K. Manning[1], CHARGE Gene-Lifestyle Interactions Working Group; [1]Massachusetts Gen. Hosp., Boston, MA, [2]Washington Univ., St. Louis, MO, [3]Univ. of Copenhagen, Copenhagen, Denmark, [4]Lundquist Inst., Harbor-UCLA Med Ctr, Torrance, CA, [5]Brigham & Women s Hosp., Boston, MA

Improved molecular understanding of gene-environment interactions (GEIs) from large-scale discovery efforts is necessary to facilitate often-elusive replication and create the confidence necessary for clinical implementation. Here, we aimed to use metabolomics to refine the relevant exposures and outcomes for a previously discovered GEI modifying the relationship between physical activity (PA) and HDL-cholesterol (HDL-C). We explored this GEI, originally identified by Kilpeläinen et al. (PMID: 30670697), in the Women's Genome Health Study (WGHS; the strongest cohort-specific GEI signal from the original meta-analysis), the UK Biobank (UKB), and the Multi-Ethnic Study of Atherosclerosis (MESA). In each cohort, we retrieved self-reported PA (MET-hrs/wk) and genotypes at rs295849 (nearest gene: *LHX1*). Plasma metabolomics data were available from NMR (WGHS, N=23,000 women; UKB, N=67,738) and LC/MS (MESA, N=877) platforms. As initially reported, minor allele carriers of rs295849 in WGHS had a stronger positive association between PA and HDL-C (p_{int}=0.002). To refine the HDL-C outcome, we tested the effect of this interaction on 45 NMR metabolites (primarily lipoprotein and lipid subfractions), finding a stronger effect on medium-sized HDL particle concentrations (M-HDL-P; $p_{int}=1.0\text{x}10^{-4}$) than HDL-C. In the UKB, the basic GEI was not significant for HDL-C or M-HDL-P. However, GEI effects were stronger both in females (consistent with the female-only WGHS cohort) and using vigorous PA as the specific exposure. Taken together, the rs295849-vigorous PA interaction in females reached $p_{int}=3.7\text{x}10^{-4}$ for M-HDL-P. Having narrowed the relevant PA subtype, lipoprotein quantity, and cohort subgroup, we examined the interaction in MESA. While lipoprotein particle concentrations were not available, the rs295849-vigorous PA interaction for HDL-C replicated in females (p_{int}=0.008). This replication was weaker in the full (male and female) population and not significant using a combined PA measure. Further exploring potential molecular pathways mediating the GEI, we found 8 LC/MS metabolites associated with vigorous PA. Of these potential molecular proxies for PA, 2 (both unknowns) showed significant interactions with rs295849 for HDL-C, with mass-to-charge ratios pointing to a bacterial lipopolysaccharide component and a very long-chain fatty acid. Our work demonstrates an improvement in replication of a GEI by refining the exposure, outcome, and relevant population subgroup. Such omics-based follow-up is critical to bridge the gap between large-scale GEI discovery and downstream experimental studies and clinical translation.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4125 † DNA language models: A new paradigm for predicting the impact of human genome variation

Authors:

G. Benegas, C. Albors, A. Aw, C. Ye, Y. Song; Univ. of California at Berkeley, Berkeley, CA

Introduction: Remarkable progress (e.g., ChatGPT) has been made in natural language processing by training large neural network models on massive text corpora. This notion of "self-supervised" learning has emerged as a powerful paradigm in biology as well; e.g., protein language models have proven to achieve outstanding performance in missense variant effect prediction. Regulatory non-coding variants also underlie numerous complex traits and diseases, but the existing DNA language models for the human genome have not yet demonstrated a competitive edge in predicting the effects of genetic variants at a genome-wide scale. These models often grapple with a number of challenges such as the diverse nature of genomic regions and the high prevalence of repetitive sequences, while requiring extremely large GPU resources. To address these challenges, we have developed a novel DNA language model that incorporates several key innovations to reduce the compute resource requirement by several orders of magnitude while achieving state-of-the-art results on genome-wide variant effect prediction.
Results: We have evaluated our model on three distinct metrics relevant to human genetics:
1. Classification of ClinVar variants.
In the case of missense variants (3 stars and above), our unsupervised model achieves an AUROC of 92.4%, comparable to CADD (AUROC=92.0%), which is a supervised model trained on proxy-deleterious and proxy-neutral variants. In contrast, the Nucleotide Transformer, a 30 times larger DNA language model trained using 4800 times more GPU hours than our language model, achieves an AUROC of only 53.2%.
2. Assessment of deleterious score enrichment in rare versus common gnomAD variants.
The top 1% of most likely functional variants predicted by our model are substantially more enriched in rare variants compared to CADD, suggesting that our model has an improved ability to detect variants under purifying selection.
3. Enrichment of UK BioBank GWAS hits in regions of extreme scores.
The top 0.1% of most likely functional variants predicted by our model are 15-fold enriched in GWAS hits, improving over CADD's 9-fold enrichment.
Conclusion: Our research underscores the promise of DNA language models for genome-wide variant effect prediction, bridging advancements in natural language processing with important insights from evolutionary biology. Our unsupervised model substantially improves genome-wide variant effect prediction over state-of-the-art methods, and it has the potential to benefit a number of statistical tasks in human genetics, including genome-wide rare variant burden testing, fine-mapping, and polygenic risk score prediction.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4126 DNA methylation GrimAge version 2.

Authors:

A. Lu[1], Q. Yan[1], S. Horvath[1,2,3]; [1]Altos Labs, SDI, San Diego, CA, [2]Dept. of Human Genetics, David Geffen Sch. of Med., Univ. of California Los Angeles, Los angeles, CA, [3]Dept. of Biostatistics, Fielding Sch. of Publ. Hlth., Univ. of California Los Angeles, Los angeles, CA

We previously described *DNAm GrimAge*, a DNA methylation-based biomarker of human mortality risk. In this study, we present version 2 of GrimAge (trained on individuals aged between 40 and 92),which leverages two new DNAm-based estimators of (log transformed) plasma proteins: high sensitivity C-reactive protein (logCRP) and hemoglobin A1C (logA1C). We evaluate GrimAge2 in 13,399 blood samples across nine study cohorts. After adjustment for age and sex, GrimAge2 outperforms GrimAge in predicting mortality across multiple racial/ethnic groups (meta $P=3.6x10^{-167}$ versus $P=2.6x10^{-144}$) and in terms of associations with age related conditions such as coronary heart disease, lung function measurement FEV1 (correlation= -0.31, $P=1.1x10^{-136}$), computed tomography based measurements of fatty liver disease. We present evidence that GrimAge version 2 also applies to younger individuals and to saliva samples where it tracks markers of metabolic syndrome.DNAm logCRP shows a positive correlation with morbidity count ($P=1.3x10^{-54}$). DNAm logA1C is highly associated with type 2 diabetes ($P=5.8x10^{-155}$). DNAm PAI-1 outperforms the other age-adjusted DNAm biomarkers including GrimAge2 in correlating with triglyceride (cor=0.34, $P=9.6x10^{-267}$) and visceral fat (cor=0.41, $P=4.7x10^{-41}$). Pedigree based polygenic models of GrimAge2 adjusted for age and gender, yielded significant narrow sense heritability estimates (=0.30, *P*=0.03), comparable with the previous version GrimAge (=0.29, *P*=0.02). Finally, we applied the blood-based mortality clock, GrimAge2, to a previously published reprogramming dataset in Human Dermal Fibroblasts (HDFs). Notably, we observed a decrease in GrimAge2 in partially reprogrammed cells after day 7, mirroring observations with Horvath's epigenetic pan tissue clock. Similarly, several estimates from GrimAge2 components, including DNAm logA1C, changed with different phases of reprogramming to reflect aging reversal post-OSKM transduction. Overall, we demonstrate that GrimAge version 2 is an appealing epigenetic biomarker for assessing human mortality and morbidity risk.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4127 DNA structural information underpins mutation rate variations and regulatory site polymorphisms in the human genome

Authors:

Z. Liu, M. Samee; Baylor Coll. of Med., Houston, TX

Single nucleotide mutation rates have critical implications for human evolution and genetic diseases. Importantly, the rates vary substantially across the genome and the principles underlying such variations remain poorly understood. A recent model explained much of this variation by considering higher-order nucleotide interactions within sequence contexts around mutated nucleotides. This model's success implicates a connection between DNA shape and mutation rates. DNA shape, i.e., structural properties like helical twist and tilt, is known to capture interactions between nucleotides within a local context. Thus, we hypothesized that changes in DNA shape features at and around mutated positions can explain mutation rate variations in the human genome. Indeed, DNA shape-based models of mutation rates showed similar or improved performance over current nucleotide sequence-based models. These models accurately characterized mutation hotspots in the human genome and revealed the shape features whose interactions underlie mutation rate variations. DNA shape also impacts mutation rates within putative functional regions like transcription factor binding sites where we find a strong association between DNA shape and position-specific mutation rates. This work demonstrates the structural underpinnings of nucleotide mutations in the human genome and lays the groundwork for future models of genetic variations to incorporate DNA shape.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4128 Efficient Bayesian inference of genome-wide genealogies for large samples

Authors:

Y. Deng[1], R. Nielsen[2], Y. Song[1]; [1]UC Berkeley, Berkeley, CA, [2]Univ California Berkeley, Berkeley, CA

The Ancestral Recombination Graph (ARG) is a structure that describes the joint genealogies along DNA sequences. Accurate inference and sampling of the ARG, especially for large datasets, has the potential to substantially improve analyses in population genetics. Several previous methods have been developed for this purpose (e.g. ARGweaver, Relate, and tsinfer), but they are either not computationally scalable or do not allow accurate sampling of both topology and time. Here we provide a new method, SINGER (Sampling and INference of GEnealogies with Recombinations), for ARG inference and sampling based on Sequentially Markovian Coalescent (SMC) model. Our method provides a balance between inference accuracy and computational speed, and can also allow full sampling of ARGs in terms of both coalescence time and topology. The new method has better inference accuracy than Relate and tsinfer, while achieving a significant speedup over ARGweaver, and with more efficient MCMC convergence and mixing.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4129 Elucidating genetic susceptibility to orofacial clefting within a multi-ancestry cohort

Authors:

Z. Koenig[1,2], I. Wong[3,4], Y. Wang[1,2,4,5], A. A. Alade[6,7], L. J. Gowans[8], M. A. Eshete[9], W. L. Adeyemo[10], C. J. Buxo[11], C. D. Padilla[12,13], E. Feingold[14,15], E. J. Leslie[16], F. A. Poletta[17,18,19], G. L. Wehby[20], J. T. Hecht[21], J. C. Murray[22], J. R. Shaffer[14,15], L. M. Moreno[23], N. Mukhopadhyay[15], S. M. Weinberg[14,15], A. Butali[6], M. L. Marazita[14,15,24], H. Brand[4,3,5], A. R. Martin[2,1,4,5]; [1]Stanley Ctr. for Psychiatric Res., The Broad Inst. of MIT and Harvard, Cambridge, MA, [2]Analytic and Translational Genetics Unit, Massachusetts Gen. Hosp., Boston, MA, [3]Dept. of Neurology, Massachusetts Gen. Hosp. and Harvard Med. Sch.,, Boston, MA, [4]Ctr. for Genomic Med., Massachusetts Gen. Hosp., Boston, MA, [5]Program in Med. and Population Genetics, The Broad Inst. of MIT and Harvard, Cambridge, MA, [6]Dept. of Oral Pathology, Radiology and Med., Coll. of Dentistry, Univ. of Iowa, Iowa City, IA, [7]Iowa Inst. for Oral Hlth.Res., Univ. of Iowa, Iowa City, IA, [8]Komfo Anokye Teaching Hosp. and Kwame Nkrumah Univ. of Sci. and Technology, Kumasi, Ghana, [9]Surgical Dept., Sch. of Med., Addis Ababa Univ., Addis Ababa, Ethiopia, [10]Dept. of Oral and Maxillofacial Surgery, Coll. of Med., Univ. of Lagos, Lagos, Nigeria, [11]Dental and Craniofacial Genomics Core, Sch. of Dental Med.,, San Juan, PR, [12]Dept. of Pediatrics, Coll. of Med., Univ. of the Philippines, Manila, Philippines, [13]Philippine Genome Ctr., Univ. of the Philippines System, Manilla, Philippines, [14]Dept. of Human Genetics, Sch. of Publ. Hlth., Univ. of Pittsburgh, Pittsburgh, PA, [15]Ctr. for Craniofacial and Dental Genetics, Dept. of Oral and Craniofacial Sci., Sch. of Dental Med., Univ. of Pittsburgh, Pittsburgh, PA, [16]Dept. of Human Genetics, Emory Univ. Sch. of Med., Emory Univ., Atlanta, GA, [17]CEMIC: Ctr. for Med. Ed. and Clinical Res., Buenos Aires, Capital Federal, Argentina, [18]Lab. of Congenital Malformation Epidemiology, Oswaldo Cruz Inst., FIOCRUZ, Rio de Janerio, Brazil, [19]ECLAMC (Latin American Collaborative Study of Congenital Malformations) at INAGEMP (Natl. Inst. of Population Med. Genetics), Buenos Aires, Argentina, [20]Dept. of Hlth.Management and Policy, Coll. of Publ. Hlth., Univ. of Iowa, Iowa City, IA, [21]Dept. of Pediatrics, McGovern Med. Sch. and Sch. of Dentistry UT Hlth.at Houston, Houston, TX, [22]Dept. of Pediatrics, Carver Coll. of Med., Univ. of Iowa, Iowa City, IA, [23]Dept. of Orthodontics, Coll. of Dentistry, Univ. of Iowa, Iowa City, IA, [24]Clinical and Translational Sci., Sch. of Med., Univ. of Pittsburgh, Pittsburgh, PA

Orofacial clefting (OFC) is one of the most prevalent birth defects, affecting 1 in 700 births globally. It is characterized by the abnormal formation of the lip and/or palate during prenatal development. Though previous OFC studies have identified several associated loci, they have not fully characterized the genetic architecture across heterogeneous phenotypes.

We have harmonized genetic and phenotypic data from a large multi-ancestry OFC cohort collected across 30 global sites. The cohort comprises 7,896 cases and 16,299 unaffected family members and controls from 5 genetic ancestry groups, including 3,653 East Asians, 4,408 South Asians, 5,183 Admixed Americans, 6,050 Europeans, and 4,901 Africans. We focused on nonsyndromic OFC, which accounts for the majority of cases but has a less clear delineation of their etiology than syndromic cases. We performed pre-imputation quality control and genotype imputation using GWASpy, a scalable and open source pipeline that we recently developed using the SHAPEIT/EAGLE and IMPUTE5 algorithms. Next, we are using GWASpy to infer genetic relatedness among individuals, which uses the pc_relate method that adjusts for population structure using principal components analysis (PCA) while inferring relatedness. Then, we are assessing population structure, identifying potential outliers within the dataset, and characterizing ancestry patterns by performing PCA. Furthermore, to more comprehensively understand the genetic architecture, we are estimating single nucleotide polymorphism (SNP)-based heritability using the genome-based restricted maximum likelihood method, allowing us to quantify the contribution of genetic variation to the observed phenotypic variance based on individual-level genotype data. Then, we will employ linear mixed models implemented in the SAIGE software to identify genetic loci associated with isolated OFC through genome-wide association study (GWAS), which offers robust and accurate association testing while accounting for relatedness of the dataset.

By conducting comprehensive analyses integrating genetic and phenotypic data from a diverse multi-ancestry OFC cohort, we aim to enhance our understanding of the genetic underpinnings of isolated OFC and contribute valuable insights into the etiology of this complex craniofacial anomaly. We anticipate that our study will have the potential to improve risk stratification via polygenic scores and enhance understanding of risk across diverse populations.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4130 Enhancing Genome-Wide Association Meta-Analysis in Multi-Ancestry and Admixed Population for Tobacco and Alcohol Use Phenotypes using 3.2 Million Individuals

Authors:

X. Wang, GWAS and Sequencing Consortium of Alcohol and Nicotine Use; Penn State Coll. of Med., Hershey, PA

Genome-wide association meta-analysis (GWAMA) has been instrumental in identifying genetic variants associated with various complex traits and diseases. As we push the boundaries of knowledge, it becomes crucial to explore beyond European populations and include individuals with diverse ancestries, including those of admixed origin. However, existing methods fail to adequately account for ancestry heterogeneity and the unique mosaic local genetic structure found in admixed populations. We propose improved GWAMA methods that account for global and local ancestry information. By harnessing principal components (PCs) of allele frequencies from each study as a proxy for global ancestry, we capture the nuanced variations in genetic effects. As for local ancestry, we model allele frequencies in admixed ancestry as a weighted combination of allele frequencies from continental ancestry backgrounds, reflecting the admixed population proportions. A grid-search window-based method is proposed to estimate these proportions by comparing the admixed population with reference panels from the 1000 Genome Project and Human Genome Diversity Project. Our approach provides a comparable estimation of ancestry proportion compared to individual-level local ancestry estimation method RFMix. We then model the genetic effects from each study as a mixture of models, incorporating PCs and ancestry proportion information, thus accounting for varying degrees of heterogeneity and local genetic structure. Additionally, considering ancestry proportion effects allows for a more accurate estimation compared to results obtained through ancestry stratification results. Our methods exhibit controlled type I error and enhanced power across diverse simulation scenarios. Applying them to the GSCAN study, encompassing 24 cohorts (78% European, 9% East Asian, 9% Admixed American, 4% African American), we unveil 838 and 830 significant loci using global and local ancestry information, respectively. Notably, 162 of these loci are novel compared to other commonly used methods, including FE, RE, RE2, BE, MR-MEGA, and TransMeta. Furthermore, we identify genetic signals exhibiting differences in effect sizes by comparing the ancestry proportion effects with ancestry stratified results. Collectively, our findings represent a significant advancement in comprehending the genetic architecture underlying tobacco and alcohol use in multi-ancestry samples and admixed populations.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4131 Ensemble Polygenic Risk Score Development for Coronary Heart Disease in Middle Eastern Populations

Authors:

M. Saad[1], K. Nahin[1], K. Kunji[1], A. El-Menyar[2], E. Ullah[1], I. Kullo[3], J. Al Suwaidi[2]; [1]Qatar Computing Res. Inst., Hamad Bin Khalifa Univ., Doha, Qatar, [2]Hamad Med. Corp., Doha, Qatar, [3]Mayo Clinic, Rochester, MN

Background: We previously validated four existing polygenic risk scores (PRSs) for coronary heart disease (CHD) in Middle Eastern (ME) populations using whole genome sequencing (WGS). Here, we aim at developing ME PRSs, and integrating them with existing PRSs to obtain an ensemble PRSs that outperforms individual PRSs.
Methods: Our cohort comprised of 1014 CHD patients and 6009 controls with 30x WGS in a Middle Eastern cohort. We split the data into training and testing datasets (70% and 30%). We developed PRSs using pruning and thresholding (P+T), LDpred2, and machine learning models (e.g., XGBoost). We downloaded 35 PRSs from the PGS catalog and evaluated their performance in the testing dataset. We combined developed and existing PRSs into an ensemble PRS using summing and machine learning techniques.
Results: P+T model performed better than LDpred2 (OR=1.8, AUC=0.664 vs OR=1.7, AUC=0.656). The 3 existing PRSs that performed the best in our data were PGS000337 (OR=1.8, AUC=0.657), PGS003356 (OR=1.6, AUC=0.64), and PGS003355 (OR=1.6, AUC=0.637). Summing these 3 PRSs improved the performance (OR=1.9, AUC=0.667). Summing these 3 PRSs, and our P+T and LDpred2 improved the performance further (OR=2.2, AUC=0.698). The best XGBoost model outperformed P+T and LDpred2 (OR=1.8, AUC=0.67).
Conclusions: Our ME PRSs performed better than existing PRSs to predict CHD. Machine learning models showed good performance but not strikingly better than P+T and LDpred2. Combining PRSs developed with different datasets and methods improved prediction performance, which suggests a greater transferability across ancestries.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4132 Estimating association effects of Copy Number Variants using penalized regression with Lasso and weighted fusion penalties.

Authors:

Y. Si[1], W. Lu[2], A. A. Tucci[1], H. Wang[3], Y. Cheng[1], L-S. Wang[3], G. Schellenberg[3], W-P. Lee[3], J-Y. Tzeng[1,2]; [1]Bioinformatics Res. Ctr., North Carolina State Univ., Raleigh, NC, [2]Dept. of Statistics, North Carolina State Univ., Raleigh, NC, [3]Penn Neurodegeneration Genomics Ctr., Dept. of Pathology and Lab. Med., Perelman Sch. of Med., Univ. of Pennsylvania, Philadelphia, PA

Background: Copy number variants (CNVs) are DNA gains or losses involving at least 50 base pairs. By analyzing one CNV locus at a time, genome-wide scanning of CNVs has provided valuable insights into their association with complex traits and diseases. However, accurately estimating the effect size of CNVs on disease risks requires considering several factors. First, the variations in CNV dosage and CNV length need to be addressed. CNVs can involve duplications or deletions with lengths ranging from a few hundred base pairs to several kilobases. The dosage and length of a CNV can impact gene expression, influencing its functional impact on the disease. Second, it is crucial to incorporate all CNV events in a genomic region and simultaneously assess the joint effect of multiple CNVs to gain a more comprehensive understanding of their combined effects. **Methods**: We propose a statistical framework that treats CNVs as consecutive events over the genome. Specifically, we model an individual's CNVs as a piecewise constant curve, which allows us to naturally account for CNV length and dosage and capture the CNV effects as a continuous function over the genome. To perform joint modeling of all CNVs in a genomic region and select CNVs associated with the outcome of interest, we adopt penalized regression with a Lasso penalty. Additionally, we integrate a weighted fusion penalty to encourage genomic regions impacted by the same CNV events to have similar effects when supported by the data. This framework avoids the arbitrary definition of a CNV locus, allows for adjusting potential confounding factors, and accommodates both continuous and binary outcomes. We conduct numerical studies using simulations and real data analysis to evaluate the performance of the proposed framework. **Results and Conclusion**: The simulation studies show that the proposed framework can more effectively identify outcome-associated CNVs (i.e., higher true positive rates) without introducing additional noise signals (i.e., comparable false positive rates) compared to the baseline methods (standard Lasso regression). Moreover, the proposed methods yield more precise effect size estimation (i.e., smaller mean squared errors in the true signal regions and equivalent mean squared errors in the non-signal regions). Real data analyses further indicate that the proposed framework provides a useful tool to detect and estimate CNV effects. In conclusion, the proposed statistical framework, modeling an individual's CNVs as a piecewise constant curve and adopting Lasso and weighted fusion penalties, demonstrates better variable selection and effect size estimation performance for CNV association studies.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4133 Estimating family-to-offspring effects from genetic data of first-degree relatives

Authors:

L. Darrous[1,2,3], Z. Kutalik[1,2,3]; [1]Univ. of Lausanne, Lausanne, Switzerland, [2]Univ. Ctr. for Primary Care and Publ. Hlth., Lausanne, Switzerland, [3]Swiss Inst. of Bioinformatics,, Lausanne, Switzerland

Motivation: While past studies investigated direct genetic effects of complex trait, we aimed at estimating the parental/family effects that arise when the genetics of an individual affect the trait of a family member. These parental/family effects also allow the estimation of parent-to-offspring causal effects via Mendelian Randomization (MR).
Methods: Using data from the UK Biobank, we estimated the direct and family genetic effects by regressing an individual's phenotype jointly onto its own genotype and that of its first-degree relatives. As a proof of concept, we examined educational attainment (EDU) and body mass index (BMI) and distinguished their direct and parental genetic basis via contrasting their genetic correlation to ~200 representative traits. To understand the parental determinants of EDU and BMI, we estimated the multivariable causal effects of >100 parental traits (proxying rearing environment) on these two focal traits by employing MR.
Results: We meta-analyzed the family effect estimates from parent-offspring and sibling pairs (N = 39'000) across ~8.5 million SNPs. As expected, the heritability of the family effects for EDU (0.06, se = 0.02) and BMI (0.01, se = 0.02) were modest since they represent environmental rearing.
Interestingly, we see no genetic correlation between offspring and parental obesity. On the other hand, high offspring BMI is genetically anticorrelated with healthy parental dietary habits (increased dried fruit/cereal, reduced beef/pork/poultry intake). In contrast to BMI, parental and offspring education extensively share genetic basis. While parental intelligence is not, parental longevity is strongly genetically correlated to offspring EDU (rG=1.33, P=2.2E-4).
MR revealed no parent-to-offspring transmission of BMI, but rather parental dietary habits and socio-economic status being the causal drivers of offspring obesity. Intriguingly, MR revealed that both parental BMI ($\boldsymbol{\alpha}$;=-0.04, P=2.5E-6) and leisure time habits (Time spent watching TV: $\boldsymbol{\alpha}$;=-0.08, P=3.7E-4) have a significantly negative causal effect on offspring EDU.
Conclusion: Using samples from the UK Biobank, we were first able to dissect direct and family genetic effects acting on BMI and EDU. These estimates fed into genetic correlation and causal inference analyses shed light onto the aspects of the rearing environment that play a key role in shaping an offspring's BMI and education level. These results highlight that a high socioeconomic environment and healthy parental diet have a favorable effect on offspring BMI, and other family habits, such as excessive TV watching, tend to decrease the offspring's educational attainment.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4134 Estimating genetic prevalence of autosomal recessive disorders in the Korean population using 5,000 genome sequencing data.

Authors:

W. Jeong, S. Ryu, G. Seo, C. Keum, H. Lee; 3billion, Seoul, Korea, Republic of

Patients with rare genetic conditions (RGC) could go through a long diagnostic odyssey when the symptoms overlap with common disorders. Oftentimes, this is due to healthcare providers not considering the patients may have an RGC either because of unawareness or because they think the disease prevalence (DP) is too low. However, DP may not be as low as there are many RGCs with underestimated DP. To break this negative cycle, it is important to accurately estimate DP. DP is measured by sampling individuals with a certain condition but estimating DP for RGCs is challenging due to the limited number of patients. However, with large genomic population data becoming available, it is now possible to calculate genetic prevalence (GP) instead, which may be more accurate than the estimated DP. GP is based on carrier frequency (CF) of the pathogenic variants of a condition. CF varies by populations so it is important to use the appropriate sub-population data to calculate.
Here, we calculated CF/GP for rare OMIM AR diseases in 5,000 'healthy' South Koreans using 102 million variants from genome data obtained through the Korean Genome and Epidemiology Study (KoGES). Using an internally developed ACMG guideline-based variant classification system, EVIDENCE, 6,099 putative pathogenic variants were identified in 1,899 genes. CF was calculated from allele frequency data for each variant. GP was calculated by aggregating the product of all possible homozygous and compound heterozygous variant combinations.
We compared our results to published data of 5 RGCs prevalent (CF>=1/200) in South Korea: Wilson Disease, Glycogen Storage Disease Ia, Phenylketonuria, Congenital hypothyroidism and Congenital lipoid adrenal hyperplasia. CF concordance was high at r=0.94 but the concordance between our GP and reported DP was r=0.25 with GP being higher than DP for 3 RGCs and vice versa for 2. Another example is cystic fibrosis, a tier 1 disorder in the ACMG carrier screening guideline. The published data in 2011 showed that there are 10 patients in South Korea diagnosed with cystic fibrosis. However, our calculated CF was 1/250, consistent with the published CF 1/261 for Asians. From this, we could estimate GP to be 1/250,000. Since the South Korean population in 2011 was ~50 million, there should have been ~200 patients with cystic fibrosis, which is 20 times more than reported.
These findings highlight the utility of population-matched large-scale genomic data in calculating GP to estimate DP and how with more accurate DP, each sub-population would be able to develop more customized healthcare plans to increase RGC awareness and diagnostic rate and develop appropriate carrier screening programs.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4135 † Estimating SNP-pair effect correlations across functional annotations.

Authors:

M. Zhang[1], A. Durvasula[2], C. Chiang[3], P-R. Loh[4], S. Gazal[5], S. Sunyaev[6], A. Price[7]; [1]Harvard Univ., Roxbury Crossing, MA, [2]Harvard Med. Sch., Cambridge, MA, [3]Boston Children's Hosp., Boston, MA, [4]Brigham and Women s Hosp. / Harvard Med. Sch, Boston, MA, [5]USC, Los Angeles, CA, [6]Harvard Med. Sch., Boston, MA, [7]Harvard Sch Pub Hlth, Boston, MA

Inferring genetic architectures of disease and partitioning heritability across functional annotations are fundamental goals with far-reaching implications. However, virtually all published studies assume that nearby SNPs have independent causal effects on disease—an assumption that warrants careful scrutiny. Here, we propose a method, LD SNP-pair effect correlation regression (LD-SPEC), to estimate correlations of causal effect sizes (of derived alleles) for pairs of nearby SNPs, depending on their functional annotations. Roughly, LD-SPEC determines that a SNP-pair annotation has positive (resp. negative) correlation of causal effect sizes if SNPs with concordant signed LD to pairs of SNPs in the SNP-pair annotation have higher (resp. lower) chi-square statistics than SNPs with discordant signed LD. LD-SPEC produces robust estimates across a wide range of genetic architectures in simulations with real genotypes.
We applied LD-SPEC to 29 UK Biobank diseases/traits (avg. *N*=298K), estimating effect correlations for common (MAF>5%) and low-frequency (0.5%<MAF<5%) SNP pairs depending on their functional annotations. We reached 3 conclusions. First, low-frequency SNP pairs had significantly negative effect correlations that decayed with distance (e.g., −0.39±0.09, −0.14±0.04, −0.06±0.01 for 0-100bp, 0-1kb, and 0-10kb low-frequency SNP pairs; non-significant for common SNP pairs). The negative SNP-pair effect correlations can potentially be explained by linkage masking, whereby linked SNPs with opposite effects on disease have reduced effects on fitness and escape negative selection. Second, positive-LD SNP pairs had significantly more negative effect correlations than negative-LD SNP pairs (e.g., −0.18±0.05 vs. +0.19±0.06 for common 0-100bp SNP pairs and -0.38±0.10 vs. +0.21±0.18 for low-frequency 0-100bp SNP pairs), consistent with linkage masking; forward simulations under an evolutionary model involving stabilizing selection recapitulated these findings. Third, SNP pairs with shared functional annotations had significantly more negative effect correlations (e.g., −0.22±0.04 for low-frequency same-exon SNP pairs vs. -0.02 for random distance-matched low-frequency SNP pairs; avg. distance of 3.7kb), consistent with a model in which SNP pairs with shared function and opposite effects on disease are likely to have opposite effects on pleiotropic traits driving fitness. Overall, our results advocate against models in which nearby SNPs have independent causal effects on disease, and motivate new directions in modeling the impact of selection on genetic architectures and improving fine-mapping.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4136 † Evaluating genetically-predicted causal effects of lipoprotein(a) in human diseases: a phenome-wide Mendelian randomization study

Authors:

J. Shi[1], S. Shringarpure[1], 23andMe Research Team, A. Auton[2], M. V. Holmes[1]; [1]23andMe, Inc., Sunnyvale, CA, [2]23andMe, Inc., South San Francisco, CA

Lipoprotein(a) (Lp(a)) is a circulating plasma lipoprotein that is emerging as an important risk factor for vascular disease. Lp(a) levels are 75-95% heritable and predominantly determined by SNPs at the LPA gene. As such, genetic variants at the LPA locus can serve as instrumental variables for investigating the potential causal effects of circulating Lp(a) levels.
In 6.5M individuals from the 23andMe database, we constructed a weighted Lp(a) genetic risk score (GRS) using instrumental variables identified to be GWAS significant for Lp(a) from the UK Biobank data. Using the Lp(a) GRS, we conducted two-stage least square phenome-wide MR analyses across a spectrum of 448 phenotypes from the 23andMe database. In addition, we compared effects of genetically-predicted higher Lp(a) to the equivalent standardized difference in LDL-C using LDL-C instruments derived from the UK Biobank data.
Our phenome-wide MR analysis identified strong effects of genetically-predicted higher Lp(a) on multiple cardiovascular endpoints. Higher genetically-predicted Lp(a) increased the risk of coronary artery disease, with the magnitude of effect being very similar to that for a comparable standardized increase in LDL. Compared to LDL-C, we found genetically-predicted Lp(a) to have a directionally consistent and yet larger magnitude of effect on risk of heart failure, aortic stenosis and peripheral vascular disease. Both Lp(a) and LDL-C increased the risk of aortic aneurysm to a similar extent. Interestingly, genetically-predicted Lp(a) was found to lead to a higher risk of arrhythmia, specifically atrial fibrillation whereas such a relationship was not seen for LDL-C. Consistent effects of Lp(a) with LDL-C were identified for risk of stroke and transient ischemic attack.
Our findings demonstrate genetically-predicted deleterious effects of Lp(a) on a broad range of cardiovascular endpoints, including some traits (e.g atrial fibrillation) that are not impacted by LDL-C. Future work will explore the independent effects of Lp(a) and LDL-C using multivariable approaches, and estimate the non-linear effects of these potential causal relationships.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4137 Evaluating the contributions of rare and common variants in disease risk

Authors:

Z. FULLER, W. Wang, K-H. Lin, J. O'connell, A. Auton, X. Wang; 23andMe, South San Francisco, CA

Over the last decade, the dramatic increase of sample sizes in genetic association studies have demonstrated the clear polygenic role of common variants in complex disease risk. While individually any associated common variant tends to have a very small effect size, the contribution of many such variants across the genome in aggregate explains a much larger fraction of the observed heritability for many complex traits. At the same time, these large sample sizes have also enabled the discovery of rare variants (e.g. < 1%) with individually large effect sizes on complex traits, yet it remains unknown how such variants contribute to population trait variance. How the combination of common and rare variants influences human disease remains an open question. Here, taking advantage of over 6 million samples from 23andMe research participants, we quantify the contributions of common and rare variants to phenotypic variance for several quantitative and case-control human traits as a proof-of-principle, including height, body mass index, and obesity. We first construct polygenic risk scores based on common variants and then generate per-gene risk scores for rare loss of function variants based on estimates from gene-based burden tests. While overall rare variants do not explain a large amount of phenotypic variance and tend to yield minimal increases in predictive accuracy (incremental AUC < 10^{-3}), we find that they tend to be enriched in individuals whose true phenotype deviates the most from their predicted phenotype based on common genetic variants. Moreover, we find that although incremental AUC tends to be small, there is a significant proportion of individuals who are successfully reclassified with the addition of a per-gene rare variant polygenic score. Finally, we demonstrate how detecting genes enriched for large effect rare variants in individuals who deviate from their common variant derived predicted phenotype may be a useful approach for identifying therapeutic targets. Together, these results shed light on the joint influences of rare and common genetic variants on human traits and provide a framework for interpreting their contributions under a liability-threshold model.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4138 Evaluating therapeutic targets for Nonalcoholic Fatty Liver Disease using human genetic and phenotypic data from the UK Biobank

Authors:

M. Luth, G. McInnes, Y. Han, D. Jakubosky, B. Cajes, S. Bruse, O. Gottesman; Empirico Inc., San Diego, CA

Nonalcoholic Fatty Liver Disease (NAFLD) is a chronic and potentially life-threatening liver condition that affects more than 25% of the global adult population. Though many investigational therapies for NAFLD are the subject of active clinical development or have been previously evaluated in clinical trials, no drugs have yet been approved specifically to treat NAFLD. Given the attrition rate in clinical development across all therapeutic indications is approximately 90% and that previous studies have demonstrated a two-fold increase in drug approval rates if the drug target for an indication is supported by human genetics evidence, we sought to evaluate whether there was supportive human genetic evidence for multiple NAFLD targets that have entered clinical trials using the genetic and phenotypic data from 452,401 individuals in the UK Biobank. We evaluated associations between nonsynonymous variants in each target gene, either individually or in aggregate, and phenotypes associated with liver disease. Specifically, we assessed the associations between individual genetic variants and gene burdens with quantitative measures of liver function or composition and diagnosis codes pertaining to NAFLD or its sequelae. While we found supportive human genetics evidence for many of the targets of investigational NAFLD drugs, particularly in the form of associations with liver enzymes, we were unable to find supportive human genetics evidence for several targets that have failed in clinical development. For example, we observed no significant associations with NAFLD-related traits for variants or burdens in MAP3K5 (ASK1), the target of Selonsertib, which failed to meet its primary efficacy endpoints during Phase 3 clinical trials in patients presenting advanced liver fibrosis due to NAFLD. We also conducted phenome-wide association studies to predict potential adverse effects related to pharmacological modulation of each target. This study provides valuable insights into possible efficacy and safety concerns for potential NAFLD targets and highlights the utility of incorporating human genetics data early in the drug discovery and development process. This research has been conducted using the UK Biobank Resource under Application Number 34229.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4139 Evaluation of epistasis detection methods in simulated quantitative phenotype data.

Authors:

S. Listopad, Q. Peng; Scripps Res., San Diego, CA

Background: Epistasis, that is gene-gene interaction, plays a role in susceptibility to common human diseases such as Alzheimer's. There are two types of epistasis: biological epistasis and statistical epistasis. Biological epistasis involves physical interaction between two or more biological components. Statistical epistasis refers to the departure from additive effects of genetic variants at different loci with regard to their global contribution to the phenotype. There are different types of statistical epistasis: dominant, recessive, multiplicative, and exclusive or (XOR). The former two are considered biologically plausible, while the latter two are not. Epistasis detection tools detect statistical epistasis. There are many epistasis detection tools currently available. Several publications have compared the performance of epistasis detection methods suitable for case-control data. However, little is known about the performance of epistasis detections tools suitable for quantitative phenotype data. **Methods:** We have identified seven implemented epistasis detection tools suitable for quantitative phenotype data: Matrix Epistasis, MIDESP, Plink Epistasis, WISH-R, SNPassoc, FRGEpistasis, and EpiSNP. Three of these methods were well documented, easy to use, and highly optimized: Matrix Epistasis, MIDESP, and Plink Epistasis. In order to evaluate these tools, we have generated 46 simulated datasets using EpiGEN: 8 dominant, 8 recessive, 20 multiplicative, 4 XOR, and 6 control datasets. All interactions were second order, that is between pairs of disease SNPs. Each dataset contained 1000 SNPs, 1000 individuals and 1-8 pairs of interacting disease SNPs. Half of the datasets were pure, that is they had no marginal effects, and half were impure, that is they had marginal effects. The strength of interaction effect was also varied across datasets. **Results:** Plink and Matrix Epistasis tools excelled at detecting dominant epistasis with 100% detection rate. They also detected 50% of recessive epistasis pairs, 16%-18% of multiplicative epistasis pairs, and 0% of XOR epistasis pairs. MIDESP detected 50% of XOR, 41% of multiplicative, 33% of dominant, and 0% of recessive epistasis pairs. All three tools were able to process a single 1000 SNP dataset in less than a minute. All three tools exhibited better detection rates for datasets with stronger interaction effects and pure interaction. **Conclusion:** Plink and Matrix Epistasis were most effective at detecting dominant and recessive epistasis, while MIDESP was most effective at detecting multiplicative and XOR epistasis. All three tools were well documented, easy to use, and highly optimized.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4140 Evaluation of input data modality choices on functional gene embeddings

Authors:

F. Brechtmann[1], T. Bechtler[1], S. Londhe[1], C. Mertes[2], J. Gagneur[1]; [1]Technical Univ. of Munich, Garching, Germany, [2]Technical Univ. of Munich, Garching bei München, Germany

Functional gene embeddings, numerical vectors capturing gene function, provide a promising way to integrate functional gene information into machine learning models. These embeddings are learnt by applying self-supervised machine-learning algorithms on various data types including quantitative omics measurements, protein-protein interaction networks, and literature. However, downstream evaluations comparing alternative data modalities used to construct functional gene embeddings have been lacking. Here we benchmarked functional gene embeddings obtained from various data modalities for predicting disease-gene lists, cancer drivers, phenotype-gene associations, and scores from genome-wide association studies. Off-the-shelf predictors trained on precomputed embeddings matched or outperformed dedicated state-of-the-art predictors, demonstrating their high utility. When predicting curated gene lists, embeddings derived from genome-wide experimental data (transcriptomics, deletion screens, and protein sequence) were outperformed by embeddings based on literature and protein-protein interactions inferred from low-throughput experiments and manual curation. However, they performed similarly for predicting genome-wide association signals and were not biased towards highly-studied genes. These results indicate that embeddings derived from literature and low-throughput experiments may appear favourable in many existing benchmarks because they are biased towards well-studied genes and should therefore be considered with caution. Altogether, our study and precomputed embeddings will facilitate the development of machine-learning models in genetics and related fields.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4141 † Exome sequencing reveals genes shaping the plasma and urine metabolome and connections to human diseases.

Authors:

N. Scherer[1,2], D. Fässler[3], O. Borisov[1], K-U. Eckardt[4,5], M. Wuttke[1], J. Hertel[3,6], A. Köttgen[1,7]; [1]Inst. of Genetic Epidemiology, Faculty of Med. and Med. Ctr. - Univ. of Freiburg, Freiburg, Germany, [2]Spemann Graduate Sch. of Biology and Med. (SGBM), Univ. of Freiburg, Freiburg, Germany, [3]Dept. of Psychiatry and Psychotherapy, Univ. Med. Greifswald, Greifswald, Germany, [4]Dept. of Nephrology and Hypertension, Univ. Hosp. Erlangen, Friedrich-Alexander-Univ. Erlangen-Nürnberg, Erlangen, Germany, [5]Dept. of Nephrology and Med. Intensive Care, Charité - Univ.smedizin Berlin, Berlin, Germany, [6]Sch. of Med., Univ. of Galway, Galway, Ireland, [7]Dept. of Epidemiology, Johns Hopkins Bloomberg Sch. of Publ. Hlth., Baltimore, MD

Background
Rare inborn errors of metabolism have revealed many genes underlying metabolic homeostasis. We hypothesized that systematic, exome-wide genetic studies of paired plasma and urine metabolomes can reveal known and novel metabolite-associated genes and variants, differentiate between systemic and kidney-specific effects, and establish connections to related diseases.
Methods
We performed gene-based burden tests using whole-exome sequencing data, focusing on the effects of rare (MAF <1%), putatively damaging variants on the levels of 1,294 plasma and 1,396 urine metabolites (781 overlapping) from 4,737 participants of the GCKD study. Two masks reflecting different genetic architectures were defined to select qualifying variants (QVs). Identified gene-metabolite associations were integrated with common variant signals from the respective GWAS to investigate allelic series, and related to health outcomes using database annotations and PheWAS. The contribution of individual QVs was prioritized via forward selection.
Results
We identified 192 significant gene-metabolite associations across plasma (P<5.04e-9) and urine (P<4.46e-9), with 43 overlaps. Overall, 73 unique genes were involved, where 22 and 17 were detected exclusively in plasma and urine. Whereas plasma-specific signals predominantly arose from lipids, urine-specific signals mainly arose from amino acids. For instance, carriers of putative loss-of-function variants in *SLC7A9* showed significantly altered levels of urine amino acids that are transported by the encoded protein in the kidney. GWAS detected common-variant associations for 68% of the respective gene-metabolite pairs, enabling the generation of allelic series. QVs revealed different genetic architectures across genes and their associated metabolites, as well as for heterozygous and hemizygous carriers. For example, male compared to female carriers of QVs in X-chromosomal *TMLHE* showed markedly higher plasma and urine levels of the encoded enzyme's substrate N6,N6,N6-trimethyllysine. The identified genes and variants were associated with many traits and diseases in the UK Biobank, including kidney-enriched transporters with kidney function measures, implicating conditions for which their cognate metabolite(s) can serve as intermediate readouts of effect sizes across the allele frequency spectrum.
Conclusion
This genetic study identified known and novel gene-metabolite relationships, differentiated between systemic and kidney-specific processes, prioritized rare damaging variants involved in disease-causing processes, and established molecular links between genes and diseases.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4142 Exome-wide assessment of pediatric rhabdomyosarcoma suggests a novel susceptibility gene related to telomere function

Authors:

Y. Yu[1], S. Pagna[2], A. Sabo[2], J. Kim[3], M. Scheurer[4], S. Dugan-Perez[2], D. Marquez-Do[2], R. Gibbs[4], J. Khan[5], J. F. Shern[5], D. Malkin[6], C. Linardic[7], B. D. Crompton[8], R. Venkatramani[2], D. Hawkins[9], S. Plon[4], D. Stewart[3], C. Huff[10], C. Huff[11], P. Lupo[2]; [1]UT MD Anderson Cancer Ctr., Houston, TX, [2]Baylor Coll. of Med., Houston, TX, [3]Natl. Cancer Inst., Rockville, MD, [4]Baylor Coll. Med., Houston, TX, [5]Natl. Cancer Inst., Bethesda, MD, [6]Hosp for Sick Children, Toronto, ON, Canada, [7]Duke Univ. Sch. of Med., Durham, NC, [8]Harvard Med. Sch., Boston, MA, [9]Univ. of Washington, Seattle, WA, [10]U.T. MD Anderson Cancer Ctr, Pearland, TX, [11]Univ. of Texas MD Anderson Cancer Ctr., Houston, TX

While previous studies have characterized the frequency of pathogenic variants in established cancer predisposition genes (CPGs) among children with rhabdomyosarcoma (RMS), there have been few attempts to evaluate gene-level estimates of relative risk or to discover novel RMS susceptibility genes. Exome sequencing from two previously published RMS cohorts (n = 689) was combined and compared to population-based controls (n = 2,218). Following an extensive data harmonization and QC procedure we utilized VAAST 2.1 to integrate predicted variant function and frequency differences between cases and controls in gene-level association tests across rare variants (MAF < 0.5%). Analyses were conducted by ancestry group (non-Hispanic European [NHE], African, and Hispanic) and then meta-analyzed. Gene-level ORs and 95% CIs were calculated among those of NHE ancestry based on sample size. Results were also stratified by RMS histology (embryonal, alveolar). We observed significant associations for established CPGs including *TP53* (meta-p = 5.9 x 10^{-5}), *NF1* (meta-p = 0.001), *BRCA2* (meta-p = 0.002), and *HRAS* (meta-p = 0.02). When considering gene-level ORs for rare coding variants irrespective of known pathogenicity, we observed associations for *NF1* (OR = 2.5, 95% CI: 1.3-4.8) and *HRAS* (OR = 4.9, 95% CI: 1.1-21.6), which were driven by variants in those with embryonal RMS. We also observed a strong association for *TP53* (OR = 5.7, 95% CI: 2.0-16.1), which was consistent for those with embryonal RMS (OR = 5.5, 95% CI: 1.7-17.9) and alveolar RMS (OR = 7.1, 95% 1.7-29.6). We also identified a potentially novel RMS susceptibility gene, *ACD* (meta-p = 3.2 x 10^{-5}), which was significant across ancestry groups and encodes the telomere protein TPP1. It was strongly associated with RMS risk (OR = 5.5, 95% CI: 2.3-13.5) and was consistent for embryonal RMS (OR = 6.4, 95% CI: 2.4-17.6) and alveolar RMS (OR = 5.7, 95% CI: 1.5-22.1). In a sensitivity analysis of *ACD* ORs by ancestry, we observed similar trends for those of Native American ancestry (OR = 1.8, 95% CI: 0.5-6.1) and African ancestry (OR = 2.4, 95% CI: 0.5-10.3), but the 95% CI included the null. Overall, our findings further characterize the role of rare germline variants on RMS susceptibility. Specifically, our exome-wide assessment 1) confirmed the contribution of CPGs on RMS risk, including *TP53*; 2) demonstrated subtype specific effects, including *NF1* and embryonal RMS; and 3) implicated a potentially novel RMS susceptibility gene - *ACD*. Notably, *ACD* is involved in telomere function, which is emerging as an important predisposing factor for other sarcomas. Additional analyses are underway to further explore and validate our findings.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4143 Exploiting latent gene-environment Interaction in the analysis of binary traits

Authors:

Z. Zhang[1], J. Lawless[2], A. Paterson[3], L. Sun[1]; [1]Univ. of Toronto, Toronto, ON, Canada, [2]Univ. of Waterloo, Waterloo, ON, Canada, [3]Hosp for Sick Children, Toronto, ON, Canada

In genome-wide association studies (GWAS), it is desirable to include and test the interaction effect (GxE) between single-nucleotide polymorphism (SNP,G) and environmental variable (E). However, directly accounting for this interaction effect is often infeasible, because the environmental variable E is latent. The latent GxE, however, can be identified through *indirect* interaction analysis. For quantitative traits (Y) that are approximately normally distributed, it has been shown that indirect testing on GxE interaction can be done by testing the heteroskedasticity of Y between genotypes. However, when traits are binary, the existing methodology based on testing for the heteroskedasticity between genotypes cannot be generalized. In this work, we propose a solution for indirect testing hence leveraging latent (GxE) interaction effects for binary traits. Through numerical experiments, we show that the proposed approach can better detect associated SNPs compared to the traditional additive testing method. We illustrate the use of the proposed method by applying it to the UK Biobank dataset for a complete GWAS study. Our method reveals potential latent interaction effects in various SNPs and genes, which would not be detected using the traditional GWAS method.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4144 † Exploring plateau and saturation of PRS accuracy in the multi-populations and multi-trait model

Authors:

Y. Ruan[1], H. Zhang[2], A. Rodriguez[3], R. Madduri[4], W. Hornsby[1], P. Natarajan[5]; [1]Broad Inst., Cambridge, MA, [2]Natl. Cancer Inst., Cambridge, MA, [3]Argonne Natl. Lab., Lincolnwood, IL, [4]Argonne Natl. Lab., Naperville, IL, [5]Massachusetts Gen. Hosp., Boston, MA

Background: Polygenic risk score (PRS) is a valuable tool for basic research and its accuracy increases with increasing sample size of discovery Genome-wide association studies (GWAS). PRS is typically based on common Single Nucleotide Polymorphisms (SNPs), and its accuracy of PRS will theoretically be saturated when it reaches the level of common SNP heritability. Previous studies have projected the risk-prediction ability of PRS across sample sizes for naive clumping and P-value thresholding methods and single discovery GWAS scenario. However, contemporary PRS utilize GWAS from multiple populations and more complex Bayesian methods, potentially leading to different saturation estimates. **Methods:** To evaluate efficient schemes for future GWAS data collection and PRS methodology development, we simulated genotype data for up to 1 million individuals across five populations (African, American, East Asian, European, and South Asian) using 1000 Genome linkage disequilibrium references. The performance of prevalent Bayesian PRS methods (e.g. PRS-CS and LDpred2) in single-GWAS and multi-trait/population scenarios was benchmarked usingthe simulated data and large-scale real-world cohorts. **Results:** Our results showed that for within-trait prediction, combining GWAS from multiple populations requires less overall sample size to reach saturation than single-population GWAS. Even when the target data and the single-population GWAS are of the same population, the combination of different population discovery GWAS holds a marginal advantage. Particularly, this is notable when the trait is under strong selection. For example, when predicting adjusted LDL-C in a subset of UKBB EUR samples, partial R^2 of PRS based on the combination of GWAS of 4 populations is close to the estimated common SNP heritability when overall N= 672k while the PRS based on only EUR data is below this level when N> 1M. Besides, our results show the usefulness of the cross-trait PRS model, particularly for binary-traits with low prevalence, where discovery GWAS might be sparse or unavailable. The PRS model leveraging multiple trait GWAS and including additional information such as annotation and local genetic correlation can further increase PRS accuracy when the PRS sample size is limited. This study offers valuable insights for the future development of PRS methodologies and efficient schemes for GWAS data collection.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4145 Exploring prevalence and cancer risk in adult individuals with pathogenic and likely pathogenic variants in RASopathy genes.

Authors:

G. Ney[1], J. Kim[1], M. Frone[1], E. Astiazaran-Symonds[1], J. Haley[2], U. Mirshahi[2], G. Urban[2], H. Rao[2], M. Shadrina[3], A. Pemov[1], A. Gross[4], M. Yohe[5], D. Carey[2], B. Gelb[3], D. Stewart[1]; [1]Natl. Cancer Inst., Rockville, MD, [2]Geisinger Clinic, Danville, PA, [3]Icahn Sch. of Med. at Mount Sinai, New York, NY, [4]Natl. Cancer Inst., Bethesda, MD, [5]Natl. Cancer Inst., Frederick, MD

RASopathies are clinical syndromes due to germline pathogenic variants in the RAS/mitogen-activated protein kinase (MAPK) pathway, leading to clinical sequelae including cardiac anomalies, neurologic conditions and failure to thrive. Notably, some of the RASopathies, such as Costello syndrome, have increased cancer incidence in childhood but very little is known about cancer risk later in life. To investigate cancer incidence in adults with germline pathogenic or likely pathogenic (P/LP) variants in canonical RASopathy genes, variants were extracted from exome data using the UKB, DiscovEHR and BioMe cohorts and clinical phenotypes were compiled. Variants were filtered and annotated with snpEFF, ClinVar (retrieved 12/08/2022) and InterVar (v.2.1.2). Variants with a read depth ≤ 10, ABHet ≤ 0.25, alternate allele read ≤ 3, or predicted loss-of-function (except for *SPRED1*) were removed. Variants were then classified by ClinVar and/or InterVar as P/LP and reviewed by a member of ClinGen's RASopathy Expert Panel. If there was consensus for P/LP, the variant was included. For analyses, non-carriers were selected as individuals with any variant status in any gene but did not harbor any P/LP variants in RASopathy genes. The most common variants identified were in genes associated with Noonan syndrome (1:2669, 1:1321, and 1:1406 in UKB, DiscovEHR, and BioMe, respectively), with *PTPN11* accounting for ~ 45% of the variants. Cancer incidence in *CBL*-associated Noonan syndrome, Noonan syndrome (NS) and cardiofaciocutaneous syndrome was similar in carriers compared to non-carriers across the UKB and DiscovEHR biobanks. However, a roughly 4-fold increased cancer incidence was observed in *SPRED1* carriers in UKB (OR 4.54; p = 4.61E-4) with earlier cumulative cancer risk than non-carriers, while no difference was observed in NS. Notably, individuals with P/LP germline variants in Noonan-associated genes had inferior survival compared to non-carriers, with a decrease in mean age of about three years (Cox p value = 8.84E-5). Surprisingly, while there was no difference in cancer incidence or cumulative incidence of cancer in NS, cancer-cause mortality was significant (Cox p-value = 0.0013), and remained after removing individuals with P/LP variants in *PTPN11* (cox p-value = 1.02E-5). Individuals who harbor germline P/LP variants in Noonan-associated genes do not have increased cancer incidence in adulthood, while data in *SPRED1* carriers in UKB suggests increased cancer incidence and earlier onset. However, carriers of germline P/LP variants in Noonan syndrome-associated genes demonstrate earlier all-cause mortality as well as cancer-cause mortality.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4146 Exploring the genetic background of facial variation in genetic disease to investigate the genotype to phenotype relationship, using achondroplasia as a model.

Authors:

K. Tran[1], H. Hoskens[1], D. J. Aponte[2], C. Unger[1], P. Claes[3], C. Rolian[4,1], B. Hallgrimsson[1]; [1]Univ. of Calgary, Calgary, AB, Canada, [2]Deep Surface AI, Calgary, AB, Canada, [3]KU Leuven, Leuven, Belgium, [4]McGill Univ., Montreal, QC, Canada

The complexity of the genotype-phenotype (GP) relationship poses a major issue in developmental biology and quantitative genetics; that is, changes in gene expression do not necessarily translate into changes in the phenotype. One trait that demonstrates this complexity is craniofacial shape, which is extraordinarily variable and is determined by interactions between thousands of genes. Craniofacial dysmorphism is characteristic among 30-40% of Mendelian diseases. Due to expressivity and penetrance, genetic diseases present with a wide range of craniofacial variability. This suggests that non-pathogenic genetic background variants interact with the disease-causing mutation, producing a continuum of phenotypic severity within the same disease. Craniofacial diseases thus offer a powerful tool to disentangle the role of background variants in the complexity of the GP relationship. One such disease is achondroplasia, the most common form of skeletal dysplasia. Patients with achondroplasia and mouse models present with shortened stature, macrocephaly, and an underdeveloped midface. Interestingly, the Longshanks mouse model, which has enhanced bone growth, has been shown to have opposite craniofacial features to achondroplasia mice. It is thus possible that the same gene networks might produce this long axis of phenotypic variability, from the Longshanks to the achondroplasia phenotype. Similarly, these genes likely contribute to the range of phenotypic severity in achondroplasia. To investigate this, I will map variants associated with the achondroplasia and the Longshanks phenotype in mice. Using geometric morphometrics, I will identify the facial shape axis associated with each phenotype and map a population of Diversity Outbred (DO) mice onto the shape axes. A genome-wide association study will reveal non-pathogenic background variants that may contribute to the achondroplasia and/or Longshanks phenotype. To confirm these effects, I will model the joint contribution of these genes to craniofacial shape using the process-MGP, a tool developed by our lab. This work will identify background variants that likely contribute to the severity of genetic disease and ultimately, improve our current understanding of expressivity, and penetrance, and the GP relationship.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4147 Expression QTL informed genetic risk score prediction in different genetic ancestries.

Authors:

S. Chhetri[1], D. Dutta[2], A. Omdahl[1], G. Qi[1], M. Arvanitis[1], A. Battle[1]; [1]Johns Hopkins Univ., Baltimore, MD, [2]Natl. Cancer Inst., Rockville, MD

Polygenic risk scores (PRS) estimate genetic predisposition for complex traits using weighted sums of trait-associated alleles and, with improved estimates based on recent large genome-wide association studies (GWAS), offer the potential to be incorporated into clinical applications in the near future. However, their limited predictive performance for populations that differ in genetic ancestry remains a substantial obstacle to accuracy and equitable usage, as most current well-powered GWAS are biased toward individuals of European ancestry.
To mitigate the poor generalizability of PRS performance, we asked whether including multi-ancestry expression quantitative trait loci (eQTL) as a functional annotation in PRS would improve portability. Using eQTLs from European and African genetic ancestry, we determine, for each regulatory locus, whether the underlying genetic architecture is shared between ancestries and prioritize shared causal signals in the genetic risk score calculation. Using lymphoblastoid cell lines eQTLs from the GENOA consortia, we applied Bayesian colocalization to identify the shared causal variants responsible for the association in both cohorts. We identified 2256 regions with colocalizing cis-genetic effects across both ancestries and 298 cis-loci with distinct causal eQTL variants with posterior probability > 0.7.
To determine the trait heritability captured by GWAS loci with different patterns of eQTL causal variant sharing, we applied stratified-linkage disequilibrium score regression on 102 immunological, hematological, and anthropological traits. We hypothesize that GWAS effect sizes, and thus PRS weights, should be shared between ancestries more strongly at loci where eQTLs also share causal variants. We observe eQTLs shared by both ancestries were strongly enriched for heritability in blood-related traits and outperformed existing functional annotations in the LDscore baseline annotation model. We calculated the mean posterior effect size for cis-variants in colocalized regions with LD-pred-funct using coloc-weighted functional priors. We trained and evaluated a Systolic Blood Pressure (SBP) PRS in the WHIMS Study with 5550 European and 6503 African cohorts. We evaluate the relative utility of ancestry-shared and ancestry-specific eQTLs as features in the PRS for each cohort. We also compare eQTL features to baseline features. Overall, we have explored the ancestry-specificity of eQTLs and the utility of prioritizing shared causal genetic signals from eQTLs to compute weighted PRS in multiple ancestries.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4148 Extending Factor Models of Pleiotropy to Incorporate Epistasis Improves Prediction Accuracy in Complex Traits

Authors:

H. Kim, A. Dahl; Univ. of Chicago, Chicago, IL

Pleiotropy is pervasive in complex traits, and biobanks provide an opportunity to jointly model pleiotropic effects across thousands of related traits. Characterizing pleiotropy is necessary for understanding shared vs specific genetic effects. Specific effects are important to understand the core biology of a trait, especially in heterogeneous traits such as major depressive disorder (MDD). Exploiting shared effects, on the other hand, improves power to detect genetic effects and polygenic prediction accuracy. However, the standard approach to understand pleiotropy-genetic correlation-is overly simplistic as it only captures genome-wide average and only applies to two traits. More recent approaches have extended genetic correlation to locus-level measures or factor models spanning many traits. Still, it remains challenging to integrate the locus-level signatures or derive biologically meaningful factors. Moreover, such models assume that pleiotropy is additive, missing potentially complex relationships between loci and factors. Here, we develop an approach to address these limitations by partitioning high-dimensional pleiotropic relationships into shared and specific interacting factors. We applied our approach to a yeast dataset with 46 traits measured on 1008 samples. First, we computed latent factors using softImpute to accommodate missing data. Then we build polygenic scores for each trait (ordinary PS) and factor (factor-PS). We found that adding the factor-PS improved prediction accuracy on top of the ordinary PS for all 46 traits (using a calibrated train-test-validate scheme). Next, we found that the epistatic interaction between the ordinary PS and factor-PS was significant for 21 traits (FDR < 10%), and that adding these terms substantially improved prediction for a subset of traits (maximum of average R2 increase across factor-PS = 17%). Notably, we found statistically significant epistatic interactions between SNPs and ordinary PS for one trait. Finally, we will present results from an application to 216 MDD-relevant traits in the UK Biobank. We previously showed how multi-trait factor analyses of this data improve prediction and specificity in MDD GWAS; here, we show how our novel approach further improves prediction and specificity using interpretable priors, advancing our understanding of MDD genetic architecture. Additionally, our factor-level PS enables powerful detection of epistatic interactions in complex human traits. We will present results evaluating the utility of factor-PS in diverse ancestries. Overall, our approach is a step toward more realistic and holistic models of phenome-wide genetic effects.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4149 † Factor analysis adjusted for cohort overlap reveals genetic components shared across GWAS traits.

Authors:

A. Omdahl[1], S. Chhetri[2], R. Keener[1], G. Qi[3], A. Battle[3]; [1]Johns Hopkins Sch. of Med., Baltimore, MD, [23]Johns Hopkins Univ., Baltimore, MD

Joint analysis of genome-wide association studies (GWAS) elucidates the contribution of pleiotropic genetic variants to correlated phenotypes. Studies performing GWAS summary statistics factorization have identified distinct genetic disease etiologies and prioritized genes relevant to correlated phenotypes. However, existing factorization methods do not account for levels of sample sharing between studies, which may inflate estimates of genetic relatedness. Though well understood in bivariate genetic correlation analyses, the confounding impact of shared participants has received limited consideration in the context of GWAS matrix factorization. To evaluate the effects and prevalence of cohort overlap in this setting, we simulated GWAS summary statistics from overlapping cohorts. Increasing sample sharing between studies inflated the correlation of summary statistics (even in the absence of genetic signal) and reduced SVD's ability to reconstruct the genetic latent space. We examined the effect of cohort overlap on estimates of trait relatedness among 4,111,707 pairs of GWAS studies from the UKBB and found non-zero cohort overlap effects in 3.8% of study pairs (Bonf. p &lt 0.05). Many such phenotype pairs (including traits like BMI and limb impedance) with confounding shared-cohort effects contribute to leading latent factors in published GWAS factorization work. To address these concerns, we present GLEANER, a method that produces **G**WAS **L**atent **E**mbeddings while **A**ccounting for **N**ois**e** and **R**egularization. This factorization method uses a regularized generalized-least-squares approach to identify genetic components from summary statistics while simultaneously accounting for estimated cohort overlap and effect size uncertainty. GLEANER is initialized to detect pleiotropic effects while enforcing component sparsity. In simulated GWAS with correlated phenotypes and cohort overlap, GLEANER outperforms SVD and FLASH in latent space reconstruction (R2). Applying GLEANER to 55 diverse UKBB GWAS (including anthropometric, ocular, behavioral, respiratory, cardiovascular, and blood-related phenotypes), we identified distinct latent components enriched for central nervous system, cardiovascular, and musculoskeletal-specific tissue markers. These components differ from those estimated without cohort-overlap correction and exclude potentially inflated relationships. We also demonstrate how our approach can evaluate GWAS from various cohorts and meta-analyses in the context of women's reproductive health phenotypes. Our work offers an interpretable workflow for jointly analyzing many GWAS, even with overlapping samples.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4150 Family history of developmental dysplasia of the hip is a risk factor for the progression of hip osteoarthritis.

Authors:

S. Yoshino[1,2], R. Yamaguchi[2], H. Tanaka[2], S. Ikegawa[1], Y. Nakashima[2], C. Terao[1]; [1]RIKEN IMS, Yokohama, Japan, [2]Dept. of Orthopaedic Surgery, Graduate Sch. of Med. Sci., Kyushu Univ., Fukuoka, Japan

Background: Developmental dysplasia of the hip (DDH) is considered to have genetic predisposition and presents many intrafamilial occurrences. However, there is no report that evaluates the effect of DDH family history on the progression after the onset of hip osteoarthritis (OA). Methods: Medical interviews about detailed clinical information including family history were conducted on 298 consecutive patients who had undergone surgery for OA due to DDH. Clinical or radiographic items that are associated with the severity of DDH (total hip arthroplasty (THA), involvement of bilateral DDH, onset age of hip pain, and three radiological indices of DDH; center-edge angle, sharp angle, and acetabular roof obliquity) were collected and evaluated in the multivariate analyses for their associations with DDH family history in a qualitative or quantitative manner. Survival time analysis for THA as the endpoint was also performed to evaluate the effects of DDH family history on the progression of OA. Results: DDH family history showed significant associations with bilateral involvement of DDH (OR=2.09 [95% CI 1.05 to 4.16]; p=0.037), early onset of hip pain (p=0.0065) and radiological severity of DDH (p=0.016). It also showed a significant association with undergoing THA (OR=2.25 [95% CI 1.09 to 4.66]; p=0.029), further supported by the Cox regression analysis (HR=1.56 [95% CI 1.15 to 2.11]; p=0.0044). Conclusions: DDH family history is a risk factor for the progression of hip OA. Stronger genetic predisposition to DDH leads to earlier onset and faster progression of hip OA.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4151 Fast and accurate heritability partitioning using association summary statistics and sparse representation of LD

Authors:

H. Li[1], R. Mazumder[2], L. O'Connor[3], X. Lin[4]; [1]Harvard Univ., Cambridge, MA, [2]Massachusetts Inst. of Technology, Cambridge, MA, [3]Broad Inst. of Harvard and MIT, Cambridge, MA, [4]Harvard T.H. Chan Sch Pub Hlth, Boston, MA

When analyzing a new genomic dataset, a common question is: are prioritized genes or regions of the genome enriched for disease heritability in GWAS? For common variant heritability partitioning, the most widely used method is stratified LD score regression (S-LDSC; Finucane et al. 2015 NG), which can jointly model a large number of overlapping functional annotations due to its computational efficiency. However, S-LDSC is inaccurate compared with slower and less versatile likelihood-based methods like GCTA (Yang et al. 2010 NG), and a particular limitation is that its results are unreliable for individual SNPs. We address these limitations with a new method, graphREML, which combines the accuracy of likelihood-based methods with the speed and versatility of S-LDSC.

graphREML utilizes linkage disequilibrium graphical models (LDGMs; Salehi, Wohns et al. bioRxiv) and GWAS summary statistics. Like S-LDSC, our method has the versatility to model numerous annotations jointly; like GCTA, it maximizes a likelihood function, leading to increased accuracy. Its use of LDGMs makes the estimation computationally efficient. It also features a flexible link function for the relationship between functional annotations and the per-SNP heritability, leading to well-calibrated heritability estimates for individual SNPs.

In simulations, we find that graphREML produces unbiased heritability enrichment estimates, with much smaller standard errors than S-LDSC (relative efficiency averages ~2.3 under realistic settings). Moreover, the estimates from graphREML remain unbiased and more precise than those from S-LDSC under model misspecification. We apply graphREML to analyze the heritability enrichments of a wide range of complex traits and diseases in the UK Biobank, using 96 functional and LD-related annotations. Our estimates are concordant with S-LDSC but have much smaller standard errors.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4152 Fast and powerful mixed-model association analysis for genome-wide association studies.

Authors:

J. Hof; Radboud Univ. Med. center, Nijmegen, Netherlands

In recent years, mixed model association analysis (MMAA) has emerged as the preferred method for performing a genome-wide association study (GWAS). MMAA can control type 1 error by accounting for population structure and familial relatedness, and can increase statistical power by conditioning on effects of loci distal to the SNP being tested. However, existing MMAA software often require long run times and substantial memory. We introduce LDAK-KVIK, a new tool for MMAA of quantitative and binary phenotypes. LDAK-KVIK first constructs a genetic prediction model, then includes this model as an offset when testing genetic variants for association with the phenotype. LDAK-KVIK has three novel features. Firstly, it includes a flexible elastic net algorithm that produces state-of-the-art genetic prediction models (e.g., more accurate than those from BayesR, glmnet and Bolt-LMM). Secondly, LDAK-KVIK never reads in more than 256 genetic variants at once, and therefore has very low computational requirements. Thirdly, we have implemented an empirical saddlepoint approximation method for robust association analysis of unbalanced binary traits, which is orders of magnitude faster than existing methods. As a result, LDAK-KVIK is both powerful and efficient. For example, LDAK-KVIK takes approximately 12 CPU hours to analyse GWAS data for 420k individuals and 10M SNPs, and requires less than 10Gb memory. By contrast, REGENIE and Bolt-LMM, two leading methods for MMAA, require 31 and 150 CPU hours, respectively, to analyse the same data. Applied to 20 traits from UK Biobank, LDAK-KVIK finds 15% and 7%, respectively, more genome-wide significant loci than Bolt-LMM and REGENIE.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4153 Fine mapping disease-causing variants through principled integration of expression quantitative trait locus results via SuSiE

Authors:

X. Zhang[1], H. Zhao[2]; [1]Yale Univ., New Haven, CT, [2]Yale Univ. Sch. of Publ. Hlth., New Haven, CT

Over the past decades, Genome-wide association studies (GWASs) have achieved remarkable success in associating thousands of genetic variants with complex traits. However, the presence of linkage disequilibrium (LD) makes it challenging to identify the true causal variants. To address this critical gap from association to causation, many fine mapping methods have been proposed to assign well-calibrated probabilities of causality to candidate variants, taking into account the underlying LD pattern. In this study, we introduce a new framework that incorporates expression quantitative trait locus (eQTL) information to fine mapping, utilizing the sum of single-effects (SuSiE) regression model. Our new method, SuSiE^2, connects two SuSiE models, one for eQTL analysis and one for genetic fine mapping. By leveraging eQTL information, SuSiE^2 enhances both the accuracy and efficiency of association studies by prioritizing functional variants within the candidate region. This is achieved by first computing the posterior inclusion probabilities (PIPs) from an eQTL-based SuSiE model with the expression level of the candidate gene as the phenotype. These calculated PIPs are then utilized as prior inclusion probabilities for risk variants in another SuSiE model for the trait of interest. Through simulations conducted on 10,000 UKBiobank samples genotyped at 20,000 SNPs, we observed that SuSiE^2 improved the power of detecting causal SNPs by 10% while reducing false positives by 50% when utilizing the in-sample LD matrix. Moreover, when the LD matrix was calculated based on an external reference panel, SuSiE^2 achieved even better performance with a 50% increase in power and a 30% reduction in the proportion of false positives. We applied SuSiE^2 to the functional SNPs associated with Alzheimer's disease (AD) predicted from single-cell epigenomic data. Among the 23 AD risk loci studied, SuSiE successfully identified the most likely mediator for five of them, while SuSiE^2 additionally captured the mediator for four risk genes: CTSH, TSPAN14, PICALM, and BIN1. While the original SuSiE detected 27 credible sets, with an average size of 8.1, SuSiE^2 identified 31 credible sets and reduced the average size to 6.6. Evaluations of AD risk genes like CTSH and C14orf93 indicated that SuSiE^2 improved the PIPs for causal mediators and achieved superior performance in distinguishing causal SNPs from non-causal variants. In conclusion, the integration of eQTL information with fine mapping through SuSiE^2 enhances the power of detecting causal SNPs while reducing false positives and the average size of credible sets.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4154 Fine-mapping causal variants using summary data with heritability-induced Dirichlet decomposition prior

Authors:

X. Li[1], P. Sham[2], Y. Zhang[3]; [1]The Univ. of Hong Kong, Hong Kong, China, [2]The Univ. of Hong Kong, Pokfulam, China, [3]The Univ. of Hong Kong, Hong Kong, Hong Kong

The goal of fine-mapping is to identify the causal genetic variants responsible for complex traits or diseases. Existing Bayesian fine-mapping methods based on discrete mixture priors rely on a prespecified maximum number of causal variants, and are more likely to be trapped into suboptimal solutions. In this work, we develop h2-D2, a novel fine-mapping method based on continuous global and local shrinkage prior. We also introduce a Bayesian inference approach that can define credible set of variants in the framework of continuous priors. Simulation studies demonstrate that h2-D2 can outperform state-of-art fine-mapping methods, SuSiE and FINEMAP, both in terms of variable selection and credible sets. We applied h2-D2 to fine-map prostate cancer using summary data from meta-analysis and identified novel causal variants. Functional enrichment analysis highlighted the importance of mediation effects of gene expression on prostate cancer risk.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4155 For the case-only design how rare must a disease be to control type I and II errors

Authors:

R. Dong[1], G. Wang[1], A. Dewan[2], S. Leal[3]; [1]Columbia Univ. Med. Ctr., New York, NY, [2]Yale Sch. of Publ. Hlth., New Haven, CT, [3]Columbia Univ., New York, NY

The case-only design can be a powerful approach to identify gene-gene and gene-environment interactions for rare diseases. It has not been investigated how rare a disease must be for the case-only design to be a valid approach. Through extensive simulations studies, we investigated the rare disease assumption and show that for diseases with prevalence <0.05 the case-only design has well controlled type I error and is substantially more powerful to detect interactions than the case-control design, but for higher disease prevalence, both type I and II errors are inflated. For a case-only study (N=10,000) where the disease prevalence is 0.20, and the genetic variant has a MAF=0.20 and a main effect odds ratio (OR)=1.2 and the environmental exposure has a prevalence of 0.1 and a main effect OR=2.0, under the null of no interaction for alpha=0.05 the type I error is 0.07. For the same scenario, but where there is an interaction (OR=1.2) the power to detect an interaction is 0.49, while for the case-control design (N=10,000 cases and 10,000 controls) the power is 0.85. For a disease prevalence of 0.04 instead of 0.20 the power for the case-only design is 0.88 and for the case-control design 0.55. We also evaluated the case-only design when there are no main effects or only one main effect, and type I and II errors are well controlled even for diseases with a high prevalence. However, the increase in power for the case-only design compared to the case-control design decreases with increasing prevalence. For example, under the same settings but where the main effect ORs are 1.0 and 2.0 for the genetic and environmental factors, respectively with an interaction OR=1.2 for a disease prevalence of 0.10 the power is 0.80 for the case-only design and 0.52 for the case-control design and for a disease prevalence of 0.20 the power is 0.59 for the case-only design and 0.49 for the case-control design. In the absents of two main effects, caution should be used when implementing the case-only design for common diseases, since main effects may be present but go undetected, which could lead to an increase in type I and II errors. In conclusion, although the case-only design is a powerful method to detect interactions, in practice it might not be beneficial for most complex traits given that usually their prevalence violates the rare disease assumption.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4156 Framework for detecting oligogenic patterns in undiagnosed clinical cohorts.

Authors:

S. Pajusalu[1,2], V. Pata[1,3], M. Marandi[4], K-M. Voovere[4], K. Ounap[1,2]; [1]Dept. of Clinical Genetics, Inst. of Clinical Med., Univ. of Tartu, Tartu, Estonia, [2]Genetics and Personalized Med. Clinic, Tartu Univ. Hosp., Tartu, Estonia, [3]Anesthesiology and Intensive Care Clinic, Tartu Univ. Hosp., Tartu, Estonia, [4]Inst. of Computer Sci., Univ. of Tartu, Tartu, Estonia

We propose a framework for studying oligogenic patterns in previously sequenced clinical cohorts to test the hypothesis that the actual underlying disease mechanism for some patients with the suspected monogenic disease is the summed effect of multiple variants across different biologically relevant genes.
The workflow is meant to reanalyze variant call files and phenotypic information from referral forms. It employs Hail software as the primary data manipulation layer and is compatible with Nextflow integration for processing large datasets in parallel. Variants are reannotated with gene labels, variant impact, and maximum population allele frequency (AF) using Ensembl VEP. Low-quality variants are filtered out to reduce noise.
Our workflow compares the variant burden between phenotype positive and negative sets (e.g., the presence of breast cancer) by collating variants with unique gene labels and gene sets. We hypothesize that phenotype-positive samples without a reported single high-impact diagnostic variant would show an overrepresentation of low to moderate-effect variants in biologically relevant genes. Conversely, as a positive control, a cohort with reported pathogenic single variants would exhibit an increased burden of rare high-impact variants in relevant genes. Variants are analyzed across four impact categories (modifier, low, moderate, high) and three frequency bins ($AF<0.01$, $0.01<AF<0.05$, and $AF>0.05$). Monte-Carlo simulation for selecting random gene sets is performed within each bin to extract variant count ratios between phenotype-positive and phenotype-negative samples. The distribution of ratios is then compared to the ratio calculated for predefined biologically relevant gene sets. Randomly selected gene sets are expected to show a distribution of ratios around 1. In contrast, predefined relevant gene sets should show an increased ratio (i.e., variant burden) if an oligogenic contribution to the disease is present.
Our current but growing database from Tartu University Hospital includes 4229 samples tested with a 113-gene inherited cancer risk panel and 4469 samples with a wide 6700-gene panel associated with various monogenic diseases, grouped and tagged with phenotypic information and reported pathogenic variants if available. Preliminary analysis focusing on breast cancer confirms an increased burden of high-impact variants in already diagnosed samples, supporting the utility of this framework. Further analysis on testing the oligogenic hypothesis will be presented for the complete cohort and different phenotypes.
Funding: Estonian Research Council grants PSG774, PRG471

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4157 From genomic blueprint to epigenetic portrait: Predicting CpG methylation from individual-level whole-genome sequencing data.

Authors:

Z. Xiao[1], Y. Gong[1], K. Su[1], A. Liu[1], C. Qiu[1], Z. Luo[1], Q. Tian[1], H. Shen[1], C. Wu[2], H-W. Deng[1]; [1]Tulane Univ., New Orleans, LA, [2]UT MD Anderson Cancer Ctr., Houston, TX

Whole-genome sequencing (WGS) offers a comprehensive view of an individual's genetic variations, while whole-genome bisulfite sequencing (WGBS) uncovers DNA methylation patterns. Genetic variations, such as single nucleotide polymorphisms (SNPs), are known to play an important role in the formation of DNA methylation marks. Generating multi-omics data from the same individual can be costly and limited. Therefore, trans-omics imputation/prediction is in critical need, as it enables us to gain deeper insights of molecular layers and uncover complex relationships and interactions within biological systems. In house, we have 310 Caucasians and 185 African Americans with paired WGS and WGBS (as training dataset), and ~7,000 individuals with only WGS (as testing dataset). As the genetic effects may be non-linear, we proposed using both linear and non-linear approach to predict individual-level WGBS data from WGS profiles at a single-nucleotide resolution. First, we developed a machine learning framework, based on elastic net regression, to integrate genetic features around individual CpG sites to build predictive models. Out of total 25,724,565 CpG sites, we obtained ~9% (2,211,322 CpGs) and ~11% (2,819,377 CpGs) well-predicted models (R^2 >= 0.01) in Caucasians and African Americans, respectively. Next, we will also build predictive models based on a non-linear approach. We will validate our approach using individuals for whom only WGS data is available. In the end, we expect our approach to have high accuracy and reliability in predicting individual-level DNA methylation patterns. Finally, we will explore the potential utility of our predictions in identifying disease-associated DNA methylation alterations and predicting disease risk based on epigenetic signatures.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4158 Genetic ancestry-based case-control matching to improve power for trait-specific association analysis.

Authors:

A. Moscati, S. Gaynor, L. Ganel, T. Joseph, GHS-RGC DiscovEHR Collaboration, Regeneron Genetics Center, Regeneron Pharmaceuticals Inc., A. Locke, E. Stahl, G. Abecasis, T. Thornton; Regeneron Pharmaceuticals, Tarrytown, NY

Studies of the genetic basis of rare diseases can often be challenging. Many such studies rely on case-only designs for efficiency - suitable for natural history studies but potentially challenging for genetic association. Large biobanks, which have revolutionized studies of common traits and diseases, can be unsuitable for analysis of rare diseases due to low prevalence. The combined analysis complementing studies that contribute predominantly cases (when enriched or ascertained for specific outcomes) with biobanks that contribute predominantly controls can achieve well-powered case-control analysis, in principle. However, these studies are challenging in practice because to avoid false positives and ensure validity of association results, the process for selecting appropriate controls for analysis with cases must account for such systematic differences between studies as heterogeneity in ancestry composition and differences in the genotyping and/or sequencing technologies used.

We present an approach for genetic association testing in case-control studies derived from the aggregation of multiple contributing studies. The pipeline optimizes the effective sample size while simultaneously controlling for bias between data sources, and benefits analyses where, in a dataset of interest, including additional controls would improve statistical power. We implement a comprehensive approach that jointly prepares array, exome, or both data types to unify a potential analytic sample and reduce noise in matching and association analyses. We then perform case-control matching based on ancestry-representative principal components using a reference panel, plus other demographic and user-specified features where relevant. The approach, implemented in a cloud-based pipeline, matches controls to cases based on a specified case-control ratio using a greedy algorithm for efficiency.

We apply the pipeline to generate a case-control cohort and evaluate genetic associations with severe obesity (BMI>40) using the Geisinger Health System DiscovEHR dataset (GHS) and UK Biobank (UKB). Using the 19K severe obesity cases and 34K controls from GHS alone results in 64% power to detect a 3% allele frequency variant with a strong genotypic relative risk of 1.20, at a P-value threshold less than 5e-8. We show that matching GHS severe obesity cases to additional UKB controls increases the effective sample size compared to using GHS alone. By matching 66K additional controls from UKB to the GHS analytic sample, power for the same variant increases to 86.3%, displaying the benefits of our approach for improved genetic association yield, while controlling type 1 error.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4159 Genetic and epigenetic insights into aging of the human retina

Authors:

T-Y. Lin[1,2], J. Advani[1], M. English[1], D. A. Ferrington[3], A. Segre[4,5], A. Swaroop[6]; [1]Natl. Eye Inst., NIH, Bethesda, MD, [2]Natl. Yang Ming Chiao Tung Univ. and Academia Sinica, Taipei, Taiwan, [3]Dept. of Ophthalmology and Visual NeuroSci.s, Univ. of Minnesota, Minneapolis, MN, [4]Ocular Genomics Inst., Dept. of Ophthalmology, Massachusetts Eye and Ear, Harvard Med. Sch., Boston, MA, [5]The Broad Inst. of MIT and Harvard, Cambridge, MA, [6]Natl. Eye Inst., NIH, Bethesda, MD

Aging is a major contributing risk factor for age-related diseases. The complex interplay of genetics, environment, and stochasticity alters local accessibility and regulation of genetic information, resulting in transcriptomic dysregulations. Aberrant gene expression patterns manifest by a gradual decline of normal physiological functions chronologically, leading to increased susceptibility to many diseases. Although genome-wide association studies (GWAS) of age-related diseases have suggested genes and pathways contributing to pathophysiology, the effects of aging in these study designs are challenging to account for and are often forgone when using age as a covariate. Quantitative trait loci (QTL) are genomic regions statistically correlated with traits (e.g., gene expression (GE)/mCpGs) and are largely non-coding regulatory sites that mediate changes in gene expression. We hypothesize that aging-associated gene expression changes are determined, at least in part, by genetic variations and epigenetic changes (especially DNA methylation) resulting from environmental factors and lifestyle. We present a novel framework to study the effect of aging in age-related diseases by longitudinally characterizing the physiologic aging of the retina by integrating diverse “omics” datasets across the lifespan. First, we modeled GE changes across age with linear and non-linear regression to identify age-related differential GE. Secondly, we will establish retinal QTLs (eQTL, mQTL) to study natural genetic variations’ impact on expression and methylation. With the QTLs, we (1) dissect genetic variation in GE/mCpGs with QTLs accounting for aging and natural variability of genetic polymorphism on GE/mCpGs; (2) explore GE/mCpGs effect on aging by analyzing the differential GE and methylated regions by comparing across the lifespan; and (3) distinguish the role of epigenetics in aging transcriptome by integrating the GE and mCpGs associated with aging with eQTM to capture the dynamics of epigenetic and indirectly environment in the aging transcriptome.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4160 Genetic architecture of acute COVID-19-related diarrhea in 23andMe research participants

Authors:

S. Aslibekyan[1], N. Chaudhary[2], C. H. Weldon[1], P. Nandakumar[2], J. Shelton[2], 23andMe Research Team, M. Holmes[2]; [1]23andMe, Mountain View, CA, [2]23andMe, Sunnyvale, CA

Diarrhea is a commonly reported extrapulmonary symptom during acute SARS-CoV2 infection. Understanding the genetic architecture of COVID-19-related diarrhea can provide insights into relevant biological mechanisms. Consented 23andMe, Inc adult participants were invited via email to complete an initial COVID-19 baseline and three follow-up surveys, distributed one month apart. Participants who self-reported testing positive for COVID-19 and experiencing diarrhea during their acute infection in any of these surveys were included as cases (N=59,010), and those who self-reported testing positive for COVID-19 and not experiencing diarrhea during their acute infection were included as controls (N=117,786). We conducted GWAS in the European population adjusting for age, sex, age*sex, and principal components. Genetic variants associated (P<=5*10-8) with diarrhea experienced during acute COVID-19 illness were tested for association with other COVID-related features and 1000+ phenotypes available within the 23andMe research database.

Mean age of the cohort was 45 years (SD=14). Female participants were more likely to report diarrhea during acute COVID-19 infection (72.4% vs 62.9%). Cases were more likely to be hypertensive (27.0% vs 23.3%), hypercholesterolemic (32.3% vs 31.5%), hospitalized for COVID-19 (6.3% vs 2.9%), and to subsequently experience long COVID (44% vs 23%). We identified four statistically significant index variants associated with COVID-related diarrhea. These index variants have been previously shown to be associated with severe COVID infection, electrocardiographic abnormalities and lipid metabolism. On phenome-wide analysis within the 23andMe research database, the index variant on chromosome 8 (negatively associated with COVID-related diarrhea) was linked to irritable bowel syndrome (IBS) with directionally opposite relationships for IBS subcategories: IBS-Constipation (positive association, p=1x10-13) and IBS-Diarrhea (negative association, p=1x10-133).

In summary, genetic associations of COVID-related diarrhea indicate potential relationships with IBS, cardiac electrophysiology and lipid metabolism. Future analyses will include other ancestry groups and functional characterization of top hits.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4161 Genetic control of gene expression in the brain links to individual variability in neuroimaging phenotypes and disease.

Authors:

N. Hoang[1], N. Sardaripour[1], G. Ramey[2], X. Bledsoe[1], E. Liao[1], Y. Chen[1], J. Park[1], B. A. Landman[1], E. R. Gamazon[3,4], K. G. Schilling[1], M. Benton[5], J. A. Capra[2], M. Rubinov[1]; [1]Vanderbilt Univ., Nashville, TN, [2]Univ. of California San Francisco, San Francisco, CA, [3]Vanderbilt Univ. Med. Sch., Nashville, TN, [4]Univ. of Cambridge, Cambridge, United Kingdom, [5]Baylor Univ., Waco, TX

Motivation: Neuroimaging phenotypes, such as regional volume and correlated functional activity, are heritable and linked to individual variability in human brain organization. Association studies have identified genetic variants associated with these phenotypes; however, the limited availability of neuroimaging and transcriptomic data from the same individuals has been a barrier to understanding the molecular mechanisms underlying human brain individuality.
Methods: We propose to bridge this gap by leveraging transcriptome-wide association studies (TWAS) to impute the genetically regulated component of gene expression (GREx) for hundreds of genes in eight brain regions for 39,565 individuals in the UK Biobank (UKB; 64.43 ± 7.69 mean age, 52% female) and 772 individuals in Human Connectome Project (HCP; 28.97 ± 3.58, 52% female). We applied Joint Tissue Imputation to impute GREx using models trained on *cis*-regulatory genetic variants and assayed expression from both the target brain region and other highly correlated regions. We then associated the individual-level, regional GREx to corresponding regional measures of volume in the UKB and connectivity in the HCP.
Results: Our TWAS on the UKB revealed 50 unique genes whose GREx associates with regional volume ($p \leq 0.005$, FDR). Several of these genes did not have genome-wide association study (GWAS) hits nearby, underscoring the potential for TWAS to discover new associations. Furthermore, comparing TWAS and GWAS results showed a strong correlation between the associations ($r_s = 0.67$, $p < 0.001$) but higher absolute effect sizes in the TWAS ($\bar{x}_{TWAS} = 52.63$, 95% CI [45.24, 60.03]; $\bar{x}_{GWAS} = 13.11$, 95% CI [12.91, 13.31]). We also discovered links between regional GREx and connectivity phenotypes in HCP. At nominal $p \leq 0.005$, we found 57 unique genes associated with regional homogeneity. Variation in the GREx of these genes also associated with multiple neurological and psychiatric phenotypes in BioVU, the Vanderbilt University biobank with linked electronic health records (EHRs).
Conclusions: Our analysis integrates TWAS with large-scale neuroimaging data and EHR-linked biobanks to complement traditional GWAS. Our analyses present a new avenue for studying brain-specific gene expression in neuroimaging genomics and, supporting their potential, uncover mechanisms of transcriptomic variation that contribute to individual variability in human brain organization and disease.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4162 † Genetic determinants of metabolic traits in middle-aged women predict cardiometabolic conditions across their life course.

Authors:

H. Pan[1,2], S. Ern[3,4], E. Lau[1], V. Gupta[1,2], A. Teh[1,2], P. Tan[1,2], P. Mishra[1], Y. Chen[5], R. Foo[3,4], K. Tan[1,6], J. Chan[7,8], D. Cameron-Smith[9], D. Wang[1,2,10]; [1]Singapore Inst. for Clinical Sci., Agency for Sci., Technology and Res., Singapore, Singapore, [2]Bioinformatics Inst., Agency for Sci., Technology and Res., Singapore, Singapore, [3]Yong Loo Lin Sch. of Med., Natl. Univ. of Singapore, Singapore, Singapore, [4]Cardiovascular Disease Translational Res. Program Inst., Natl. Univ. of Singapore, Singapore, Singapore, [5]Dept. of Mathematics and Computational Biology Programme, Natl. Univ. of Singapore, Singapore, Singapore, [6]Dept. of Lab. Med., Natl. Univ. Hosp., Singapore, Singapore, [7]Dept. of Reproductive Med., KK Women's and Children's Hosp., Singapore, Singapore, [8]Academic Clinical Program in Obstetrics and Gynaecology, Duke-NUS Med. Sch., Singapore, Singapore, [9]The Coll. of Hlth., Med. and Wellbeing, The Univ. of Newcastle, Newcastle, Australia, [10]Natl. Heart & Lung Inst., Imperial Coll. London, London, United Kingdom

Cardiometabolic diseases, encompassing a spectrum of disorders such as hypertension, coronary artery disease, stroke, and diabetes, are amongst the leading causes of morbidity and mortality worldwide. Mechanism studies of these cardiometabolic abnormalities reveal a complex interplay between genetic, environmental, and lifestyle factors throughout an individual's lifespan. Metabolic syndrome (MetS) is a complex cluster of traits associated with cardiometabolic diseases. The prevalence of severe MetS increases to 40% in peri-menopausal and post-menopausal women, resulting in a significant rise in heart failure and type 2 diabetes mellitus (T2DM). Despite MetS being a well-established risk factor, the impact of genetics on these traits throughout an individual's lifespan has been understudied.

To address this issue, we developed GenMetS, a polygenic model which was learned from MetS observed in middle-aged women. The model was composed of 7,850 SNPs selected from 17 GWAS studies related to the traits including waist circumference, glucose, triglycerides, HDL cholesterol and blood pressure. The relative contributions of the SNPs to MetS were learned from a multi-ethnic Asian cohort of women during pre- and post-pregnancy (N=1368, age=33.98 ± 5.72 years). Validation of GenMetS was performed on three independent cohorts: the UK Biobank (N=394,919), a Singaporean heart failure cohort (N=2310), and children from the discovery cohort at the age of 6 (N=1073).

GenMetS consistently explained 5-12% of the observed MetS in women of Asian ancestry across different age groups and nationalities. The GenMetS predicted by our genetic model was associated with the onset of six cardiometabolic diseases at different ages, including T2DM (odds ratio (OR) =1.83), hypertension (OR=1.49), and heart failure (OR=1.25). Moreover, it contributed to the prediction of multimorbidity (area under the curve = 69.3%) in conjunction with lifestyle and socioeconomic factors. The heritability of GenMetS was demonstrated by its associations with the growth trajectories of children (p-value = 1.3×10^{-7}). Our results highlight that there exists a genetic signature for metabolic traits in mid-life that persist for women, throughout their lifespan and in their offspring.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4163 Genetic determinants of the progression of Lewy body pathology

Authors:

L. Wu; Univ. Coll. London, London, United Kingdom

Parkinson's Disease (PD), PD dementia (PDD) and Dementia with Lewy bodies (DLB), which we describe here jointly as Lewy body diseases (LBD), are characterized pathologically by alpha-synuclein aggregates forming Lewy bodies and Lewy neurites. Cognitive and neuropsychiatric complications observed in LBDs correlate with the development of cortical Lewy body (LB) pathology. LB pathology can be divided into subcortical, limbic and neocortical stages. We have studied a large series of Brain Bank LBD cases to define the genetic drivers of regional LB pathology. We included 742 pathologically defined LB cases with summary pathological data together with genotyping using the Illumina Neurobooster array (NBA) and whole genome sequencing with 150 base-pair paired-end sequencing with 30x coverage generated in the Global Parkinson's Genetic Program (gp2.org). 454 (61%) cases had developed cortical LB pathology at post-mortem with a median disease duration of 12 years (interquartile range 8-20 years). We evaluated the genetic features of the most rapidly progressing cortical LB cases (n=132; disease duration <=7 years) as compared to those who did not develop cortical pathology, or developed it with a disease duration of 8 years or greater. We compared variant frequencies between cases and controls using Fisher's exact test. In common variants: 39 % of the rapid progression group carried the ApoEe4 allele compared to 32% in controls, 71 % carried the MAPT H1/H1 haplotype compared to 64% in controls (not significant) and 64 % carried the SNCA 5' PD risk allele compared to 78% in controls ($p < 0.01$). We evaluated the PD polygenic risk score and there was no difference between PD PRS between cases and controls. Cortical LB pathology in well characterized cases represents a sensitive method to define progression to advanced LB disease; and further analysis of common and rare risk variants is underway to provide new insights into the pathogenesis of these diseases.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4164 Genetic mapping power analyses for single-cell transcriptomics.

Authors:

Z. Akras[1,2], N. Mars[1,3], R. K. Walters[1,2], B. M. Neale[1,2]; [1]Broad Inst. of MIT and Harvard, Boston, MA, [2]Analytic and Translational Genetics Unit, Ctr. for Genomic Med., Massachusetts Gen. Hosp., Boston, MA, [3]Inst. for Molecular Med. Finland, FIMM, HiLIFE, Univ. of Helsinki, Helsinki, Finland

Single-cell transcriptomics paired with genetic information can elucidate mechanisms driving variability in cellular phenotypes. An important consideration in the design of these single-cell experiments is the statistical power to detect genotype-phenotype associations. Simulations of single-cell data can provide insight into the power properties of different experimental designs, but existing simulation methods focus on detecting changes in cell type composition or gene expression between sample types and do not address the genetic contribution to these differences. Furthermore, tools explicitly designed for mapping do not model the genetic effects on cell identity, nor do they consider power implications of the analysis method following data generation, such as whether to pseudobulk cells or perform a global comparison. We present a simulation framework for comprehensive power analysis of single-cell transcriptomic experiments in the context of genetic mapping. Our mixed model-based approach uses pilot data including genotype, cell state, and transcriptomic readout to estimate genetic effects on individual cell type composition and expression landscape as well as sources of variance. We then generate cells with simulated genotypes from a predetermined number of samples and assign cell states based on each cell's genotype according to the relationship learned from the pilot data. Finally, we simulate expression profiles as a function of each cell's state and genotype, with the strength of the genetic effects specified based on heritability. We examine the dependence of statistical power on parameters including number of samples, number of cells per sample, genetic and cellular heterogeneity, and analysis strategy. Our framework can inform the successful design of single-cell genetic mapping experiments to help decipher global and cell type-specific gene regulatory mechanisms.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4165 Genetic prediction of early adolescent chronotype: Effects of sex and pubertal status.

Authors:

L. Hernandez[1], M. Gandal[2], A. Fuligni[1]; [1]Univ. of California, Los Angeles, Los Angeles, CA, [2]Univ. of Pennsylvania, Philadelphia, PA

Adolescence is characterized by changes in sleep-wake cycles and a shift to later sleep onset and wake times, indicative of a more evening chronotype. While genome wide association studies (GWAS) in adult populations have shown that individual differences in chronotype are heritable and causally related to mental health outcomes, it remains unclear when during development genetic effects on sleep-wake cycles first emerge and whether these effects are stable during adolescence. To address these questions, here we examined the extent to which polygenic scores (PGS) derived from adult GWAS predict adolescent sleep-wake cycles and mental health. Data were obtained from 2,625 unrelated 10-13-year-old youth of European ancestry who participated in the Adolescent Brain and Cognitive Development study. Chronotype PGS were tested for associations with 1) chronotype as assessed with the Youth Munich Chronotype Questionnaire, and 2) 50 indices of youth mental health and cognitive functioning. We found that female youth were more pubertally advanced relative to male youth ($p<0.001$) and had later chronotype ($p<0.05$). As sex-differences were observed in both chronotype and pubertal status, we examined potential sex-differences in the predictive accuracy of chronotype PGS. The interaction between PGS and sex in predicting chronotype was significant ($p<0.05$), suggesting that sex moderates the relationship between PGS and sleep-wake cycles in adolescent youth. Follow-up analyses were performed in males and females separately; here, we found that PGS significantly predicted chronotype in female youth only ($p<0.001$). To determine potential biological moderators influencing the strength of the relationship between PGS and chronotype, we next performed regressions in each sex grouped by pubertal stage. Remarkably, we found that PGS was a significant predictor of chronotype only in females who were in mid- or late-pubertal stages ($p<0.05$), suggesting that puberty may be a critical transition point at which genetic variants discovered in adult populations become applicable to pediatric samples. At the behavioral level, in male youth, higher PGS was associated with lower levels of sleep-related disturbances and cognitive functioning (FDR $q<0.05$). In contrast, female youth showed a distinct association between higher PGS and lower levels of psychosis-like experiences (FDR $q<0.05$). Together, these data suggest that the predictive power of chronotype PGS emerges during puberty and emphasize the importance of sex and developmental stage as key moderators of genotype-phenotype associations.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4166 Genetic relationships between BMI, height velocity and pubertal timing in an Indigenous study population

Authors:

M. Ramirez Luzuriaga, S. Kobes, W-C. Hsueh, L. Baier, R. Hanson; Phoenix Epidemiology and Clinical Res. Branch, NIDDK, Phoenix, AZ

Early pubertal maturation is often observed in boys and girls with high BMIs, and both age of puberty and BMI are heritable. In this study, we aimed to determine whether these phenotypes are associated via genetic variation that contributes to both traits. We, thus, examined the associations of a polygenic score (PS) for BMI and a PS for age of menarche (AAM) with maximum BMI and four parameters of adolescent growth in a longitudinally studied Indigenous population from the Southwestern US with high prevalence of obesity.
Pre-existing genome-wide imputed genotypes and anthropometric data were available from a longitudinal study. Adolescent growth parameters were derived from the Preece-Baines growth model, a parametric growth curve for longitudinal height data. PSs for the BMI-PS (798 markers) and the AAM-PS (275 markers) were constructed by selecting common variants (allele frequency >0.01 in the Indigenous sample) with a high imputation confidence level (score >0.5) which achieved genome-wide significance in European meta-analyses (GIANT/UK Biobank Consortium and Reprogen Consortium, respectively). Associations between PS and maximum BMI (N=6789) were adjusted for birth year, sex, and the first 5 genetic principal components (PCs). Associations with parameters of adolescent growth (N=787) were adjusted for birth year, sex, maternal diabetes and the first 5 genetic PCs. Associations were calculated using linear mixed models, accounting for familial relationships.
The BMI-PS had a modest but significant association with maximum BMI ($r=0.14$, $p=5.73\times10^{-27}$) and velocity at take-off ($r=0.10$, $p=0.003$). It also had a weak but significant inverse association with age at peak velocity ($r=-0.05$, $p=0.03$). We found no significant association of BMI-PS with age at take-off ($r=-0.04$, $p=0.06$) and age at maturation ($r=-0.03$, $p=0.14$). Genetically predicted later puberty onset (AAM-PS) had a weak but significant inverse association with maximum BMI ($r=-0.03$, $p=0.005$) and velocity at take-off ($r=-0.10$, $p=0.002$). It also had a modest but significant positive association with age at peak velocity ($r=0.13$, $p=1.8\times10^{-04}$), age at take-off ($r=0.11$, $p=0.0013$) and age at maturation ($r=0.13$, $p=1.3\times10^{-04}$). Our results suggest that many variants that associate with AAM and BMI in large European cohorts also affect these traits in this Indigenous population. They also support the notion that obesity and early puberty onset share genetic determinants.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4167 Genetic Risk Factors for Late Bladder and Bowel Toxicity after Prostate Cancer Radiotherapy

Authors:

M. Farazi[1], G. Barnett[2], N. Burnet[3], D. Dearnaley[4], A. Dunning[5], D. Azria[6], A. Choudhury[7], T. Rancati[8], D. Ruysscher[9], P. Seibold[10], C. Talbot[11], L. Veldeman[12], A. Webb[11], M. Aguado-Barrera[13], A. Carballo[13], A. Gómez-Caamaño[13], P. Peleteiro[13], A. Vega[13], H. Ostrer[14], B. Rosenstein[15], T. Imai[16], S. Saito[17], M. Parliament[18], N. Usmani[18], K. DeRuyck[12], W. Hall[1], M. Janelsins[19], C. West[7], P. Auer[1], S. Kerns[1]; [1]Med. Coll. of Wisconsin, Milwaukee, WI, [2]Cambridge Univ. Hosp. NHS Trust, Cambridge, United Kingdom, [3]Christie Hosp., Manchester, United Kingdom, [4]Royal Marsden NHS Fndn. Trust, Surrey, United Kingdom, [5]Cambridge Univ., Cambridge, United Kingdom, [6]L'Inst. du Cancer de Montpellier, Montpellier, United Kingdom, [7]Univ. of Manchester, Manchester, United Kingdom, [8]Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy, [9]Maastricht Univ. Med. Ctr., Maastricht, Netherlands, [10]German Cancer Res. Ctr. (DKFZ), Heidelberg, Germany, [11]Univ. of Leicester, Leicester, United Kingdom, [12]Ghent Univ. Hosp., Ghent, Belgium, [13]Fundación Pública Galega de Med. Xenómica-Servizo Galego de Saude (SERGAS), Santiago de Compostela, Spain, [14]Albert Einstein Coll. of Med., Bronx, NY, [15]Icahn Sch. of Med. at Mount Sinai, New York, NY, [16]Natl Inst Radiological Sci, Chiba, Japan, [17]Natl. Tokyo Med. Ctr., Tokyo, Japan, [18]Cross Cancer Inst., Univ. of Alberta, Edmonton, AB, Canada, [19]Univ. of Rochester, Rochester, NY

Prostate cancer (PCa) is the most frequent cancer in men, with about 1 in 8 diagnosed during his lifetime. Non-metastatic PCa is highly curable with standard care that includes radiotherapy, but long-term side effects impact survivors. The purpose of this study was to investigate the genetic basis of late bladder and bowel toxicities following radiotherapy for PCa that limit curative doses and have an adverse effect on survivors. We performed a multi-ethnic genome wide association study (GWAS) of late bladder (hematuria) and bowel (rectal bleeding) toxicities including >7,000 participants in multiple cohorts to discover new risk loci. SNP-array data from each cohort were imputed into the TOPMed reference panel and used genome build GRCh38 with standard quality control procedures. We additionally developed a polygenic risk score (PRS) for each toxicity using GWAS results from a training set and tested each PRS in the remaining participants. All analyses used Cox proportional hazards regression to model genetic association with time to onset of grade 2 or worse toxicity and were adjusted for cohort and ancestry captured by principal components. The GWAS identified six new risk loci (rs571126149 on 1p36.11, HR=2.40, $p=2.67x10^{-8}$; rs56260582 on 4p15.1, HR=1.44, $p=3.11x10^{-9}$; rs4134357 on 10q11.22, HR 1.51, $p=2.02x10^{-8}$; rs72672122 on 4q28.2, HR=2.51, $p=2.71x10^{-8}$; rs12434768 on 14q24.2, HR 2.86, $p=1.51x10^{-8}$; and chr16:31181061:C:G on 16p11.2, HR=3.05, $p=7.35x10^{-9}$). The PRSs were comprised of 4,863 and 3,705 SNPs for hematuria and rectal bleeding, respectively. Multivariable models combining the PRS and clinical factors (age at radiotherapy, prior prostatectomy, receipt of hormone therapy) show the PRS contributes independently to risk of developing late toxicity in the test set (hematuria: per-allele adjusted HR=1.02, p=0.001; rectal bleeding: per-allele adjusted HR=1.02, p=0.018). Patients in the upper 20% of the hematuria PRS had an approximately 2-fold increased risk of developing that toxicity (adjusted HR 1.98, p=0.010) and patients in the upper 20% of the rectal bleeding PRS had an approximately 1.6-fold increased risk of developing that toxicity (adjusted HR 1.63, p=0.021) compared with patients in the lower 20% of the respective PRS. PRSs could be incorporated into dosimetric models used in treatment planning for PCa to personalize treatment by prioritizing avoidance of organs at high risk and guiding selection of patients who might benefit from MR-guided radiotherapy or proton therapy.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4168 Genetic risk score in predicting Age-related Macular Degeneration

Authors:

X. Zhang, N. Parikshak, G. Abecasis, Regeneron Genetics Center, DiscovEHR, Mayo Clinic Project Generation, Penn Medicine Biobank, G. Coppola, E. Jorgenson; Regeneron Genetics Ctr., Tarrytown, NY

Age-related Macular Degeneration (AMD) is the leading cause of irreversible blindness worldwide and can be separated clinically into geographic atrophy (GA, dry AMD) and choroidal neovascularization (CNV, wet AMD). Current treatments are effective in preventing vision loss from CNV; however, no currently approved treatment exists to prevent vision loss from GA. Understanding the genetic risk in AMD subtypes is one of the crucial considerations in designing clinical trials. We calculated a genetic risk score (GRS) using the summary statistics of 60 loci from the Million Veteran Program, which has the largest number of AMD cases in a population-based cohort. We characterized the GRS for 4 major age groups: Age 46-55; Age 56-65; Age 66-75 and Age 76 and above, in 8 biobank cohorts for AMD (UK Biobank, Geisinger Health System, Mayo Clinic Project Generation, Colorado, UCLA, Penn Medicine Biobank, MALMO and SINAI). In addition, we characterized AMD subtypes (CNV and Late AMD) in 4 cohorts with sufficient phenotypic data. We present the distribution of AMD prevalence across GRS decile within each age category. Comparing 1^{st} to the 10^{th} GRS decile, we observed the smallest increase of AMD risk for age below 65; and the strongest increase for age 76 and above for all scenarios (All AMD, CNV and Late AMD). Specifically for age 76 and above, there is a 3x risk increase for all AMD; in late AMD, the risk is 6x. The strongest increase is in wet AMD, for which risk increased 10x between the two extreme deciles. The observed GRS trend is consistent with disease pathology - meaning that genetic risk load contributes more to the most severe AMD form - wet AMD. Meanwhile, age greater than 76 has a striking increase in disease prevalence compared to other age groups. Our results present the landscape of GRS distribution in AMD, which provides insights into the understanding of genetic risk across different age groups within AMD subtypes. These findings could also aid in the selection of participants for clinical trials of AMD treatments.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4169 Genetic Susceptibility of Deletions or Excess Homozygosity to Head and Neck Cancer in Whole-Genome Studies.

Authors:

C-C. Wu[1], C-H. Chen[2], R. Yu[3], S. Shete[3]; [1]Natl. Cheng Kung Univ., Tainan, Taiwan, [2]Academia Sinica, Taipei City, Taiwan, [3]The Univ. of Texas MD Anderson Cancer Ctr., Houston, TX

Squamous cell cancer of the head and neck includes cancers of the oral cavity (including the gums and tongue), pharynx, and larynx and is the sixth most common malignancy worldwide. Genetic susceptibility to developing oropharyngeal and oral cancer incidence has been established as several loci, including alcohol-related genes (ADH1B and ADH7) and HLA region11, have been discovered, using whole-genome association analyses. However, the roles of deletion variants or excess homozygosity implicated in head and neck cancer susceptibility are poorly understood. Here, we performed two distinct whole-genome scans of associations between deletion variants or excess homozygosity and head and neck squamous cell carcinoma susceptibility and a meta-analysis synthesizing the association summary statistics from these two whole-genome scans. We used a logistic regression framework extension of the genome-wide statistical method we developed previously to assess the excessive homozygosity, permitting us to adjust for covariates. We used 733,202 SNPs in 2185 patients with head and neck squamous cell carcinomas from the MD Anderson Cancer Center in Houston, Texas, recruited from 1998 to 2012. The two whole-genome scans of disease-associated deletions include batch 1, with 1154 cases and 1542 controls, and batch 2, with 1031 cases and 2965 controls, respectively. We detected a 1.1-kb segment on chromosome 3p in batch 1 and 3 distinct 0.45-14kb segments on chromosome 1p, 5q, and 11q in batch 2. They were all significant at a 0.05 nominal significance level, adjusted for multiple comparison procedures. The corresponding meta-analysis identified five distinct segments on chromosome 5p, 6p, 9q, and 12q (2 segments).

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4170 Genetic variants associated with immune cell population abundances in single-cell data.

Authors:

L. Rumker[1,2,3], S. Sakaue[1,2,3], J. B. Kang[1,2,3], Y. Reshef[1,2,3], C. Valencia[1,2,3], S. Yazar[4], J. Alquicira-Hernández[4], P. Joseph[4], S. Raychaudhuri[1,2,3]; [1]Harvard Med. Sch., Boston, MA, [2]Brigham and Women's Hosp., Boston, MA, [3]Broad Inst. of MIT and Harvard, Cambridge, MA, [4]Garvan Inst. of Med. Res., Sydney, Australia

Genome-wide association studies have identified many variants that influence disease risk. To investigate the mechanisms mediating these associations, previous functional studies have primarily sought co-localizing associations of variants to univariate intermediate phenotypes, like the expression of individual genes (eQTLs) or proteins. To fully understand these disease risk mechanisms, we must also seek associations to more complex intermediate traits, like shifts in the abundance of functional cell states. However, there is limited precedent for GWAS of traits strategically defined within high-dimensional functional data, like single-cell transcriptional profiling.

In this study, we have developed a novel approach to identify fine-grained cell states in single-cell transcriptomic data that change in abundance with allelic dose of a genetic variant: cell state abundance QTLs (csa-QTLs). Our approach avoids pre-specifying candidate cell types to test so that we can more flexibly characterize genetically-associated cell states.

Applied in a genome-wide survey to a published dataset of ~1M peripheral blood cells from 969 Australians of European ancestry, our method reveals six independent loci (lead P-values less than 2e-8) associated with changes in immune cell population abundances. For example, one locus on chromosome 15 (lead SNP rs11632488, P-value 1.1e-8) is associated with increased ratio of classical to non-classical monocytes. Another locus on chromosome 12 (lead SNP rs3003, P-value 2.4e-11) is an eQTL for the NK cell lectin-like receptor gene *KLRC1*, key to innate immune destruction of malignant and virally-infected cells. We find that this SNP is associated with increased abundance of NK cells expressing TNF-alpha response genes. Interestingly, this signal co-localizes with a risk allele for psoriasis, a disease for which anti-TNF treatments have well-established efficacy. Therefore, csa-QTLs may provide disease-relevant mechanistic insight. Because we detect csa-QTLs in a population cohort without acute disease, characterizing these effects may illuminate how our genetic background can set the immunological stage for disease risk.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4171 Genetic variation for longitudinal change in anthropometric traits.

Authors:

K. Kemper[1], J. Sidorenko[1], H. Wang[1], B. Hayes[1], N. Wray[1], L. Yengo[1], M. Keller[2], M. Goddard[3,4], P. Visscher[1]; [1]The Univ. of Queensland, Brisbane, Australia, [2]Univ. of Colorado, Boulder, CO, [3]BioSci.s Res. Div., Bundoora, Australia, [4]The Univ. of Melbourne, Melbourne, Australia

Most genetic studies use cross-sectional data, that is a single observation on an individual at a given time. These studies cannot address questions relating to the degree of genetic control for within-person trait change without multiple measurements per person. In this study we used 50,117 European ancestry individuals from the UK Biobank with 2 repeated observations on 8 anthropometric traits (including height and weight). Individuals were 40-79 years at baseline with an average follow-up time of 7.4 years. We estimate an average loss from peak values of about 3.8 cm (females) and 3.4 cm (males) for height, about 3.7 kg (females) and 4.1 kg (males) for weight. The SNP-based heritability for rate of trait change was small but significantly greater than zero for height (0.015, s.e. 0.006), weight (0.031, s.e. 0.007) and body mass index (0.035, s.e. 0.007). We constructed polygenic scores (PGS) for the rate of trait change and found that a low PGS for height change (i.e. predicted height loss) was significantly associated with musculoskeletal disease diagnosis in an independent sample of the UK Biobank (P = 0.001). A genome-wide association study for rate of trait change found that the Alzheimer's APOE-E4 risk allele was significantly associated with weight loss (-0.047 kg per year, s.e. 0.007, $P = 2.2 \times 10^{-11}$). Further, we used an approximate random regression approach and independent cross-sectional data to show that the genetic correlation between weight measured at a young (40-50 years) or older (65-79 years) age is significantly less than unity (r_g = 0.92, s.e. 0.023, $P = 3 \times 10^{-4}$). In summary, we provide evidence at the genome and locus level for some degree of genetic control on longitudinal trait change for anthropometric traits, and show that genetic predictors of longitudinal trait change are associated with disease. Measuring large numbers of individuals over longer intervals will increase the power of future studies to investigate the genetic influence on longitudinal traits.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4172 Genetic variation in the FMO and GSTO gene clusters impacts arsenic metabolism in humans

Authors:

L. Tamayo[1], L. Tong[1], T. Davydiuk[2], X. Le[2], F. Jasmine[1], M. G. Kibriya[1], J. Graziano[3], H. Ahsan[1], M. V. Gabmle[3], B. L. Pierce[1]; [1]Univ. of Chicago, Chicago, IL, [2]Univ. of Alberta, Edmonton, AB, Canada, [3]Columbia Univ., New York, NY

Background: In Bangladesh, >50 million individuals are chronically exposed to inorganic arsenic (iAs) through drinking water, increasing risk for cancer and other arsenic-related diseases. Previous studies have shown that an individuals' ability to metabolize and eliminate iAs (and their risk of toxicity) is influenced by genetic variation in the *AS3MT* and *FTCD* regions.
Methods: To identify additional loci related to arsenic metabolism, we used data from Bangladeshi individuals participating in arsenic-related research studies to conduct genome-wide association analyses of the relative abundances of arsenic species/metabolites measured in both urine (n=6,540) and blood (n=976). The arsenic species measured include inorganic arsenic (iAs), mono-methylated arsenic species (MMA) and di-methylated arsenic species (DMA), with DMA representing the "end metabolite" most rapidly excreted in urine.
Results: In analyses of urine arsenic species, we identified a novel association signal in the FMO gene cluster (1q24.3), with the lead SNP residing in *FMO3* (MMA% P=4.2x10-16). In analyses of blood arsenic species, we identify an additional signal in the FMO cluster, distinct from the signal for urine arsenic species, with the lead SNP residing in *FMO4* (DMA% P=2.26x10-22). Analyses of blood arsenic species also identified a novel signal in the *GSTO1/2* region (10q25.1) with the lead SNP residing in *GSTO1* (DMA% P=5.3x10-13). FMOs are xenobiotic metabolism enzymes, but have no known role in arsenic metabolism, while *GSTO1* has a well-known role in the reduction of arsenic species, a key step in arsenic metabolism. These newly identified SNPs did not show clear association with risk for arsenic-induced skin lesions (P>0.05), the most common sign of arsenic toxicity (based on 3,448 cases and 5,207 controls).Lead SNPs in *FMO3* region are associated with the splicing of *FMO3* (sQTL) in multiple GTEx tissue types and contain a missense variant in *FMO3*. The *GSTO1* signal is also an sQTL (for *GSTO1*) in multiple tissue types and contains a missense variant. The lead SNPs for blood DMA% in the *FMO4* gene are associated with *FMO4* expression (eQTL) in multiple tissue types.
Conclusion: We identified novel genetic associations with arsenic species/metabolites measured in both urine and blood. While previously identified regions affecting arsenic metabolism (*AS3MT* and *FTCD*) are associated with arsenic species in both blood and urine (and toxicity risk), the associations reported here appear specific to blood or urine, with no detectable impact on toxicity. These findings reveal to complexities in arsenic metabolism and its genetic contributors that require further study.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4173 Genetically regulated gene-expression associates with human functional brain-network organization

Authors:

N. Sardaripour[1], N. Hoang[1], K. Schilling[1], M. Benton[2], J. Capra[3], M. Rubinov[1]; [1]Vanderbilt Univ., Nashville, TN, [2]Baylor Univ., Waco, TX, [3]Univ. of California San Francisco, San Francisco, CA

Objective: Large-scale brain networks mediate diverse sensory-motor and cognitive functions. The structure of these networks is heritable, but the specific genes involved in their formation are still unknown. Here, we advanced our understanding of the association between regional gene expression and human brain-network organization by leveraging joint-tissue transcriptome association and local-global neuroimaging modeling to integrate information from distinct sets of data: neuroimaging, genomic, and transcriptomic.
Methods: We first inferred genetically regulated gene expression from 10 cortical and subcortical regions from the European ancestry subsets of the Human Connectome Project (772 people, mean age 28.97 ± 3.58, 52% female) dataset. Separately, we computed structural and functional phenotypes for these regions, including gray-matter volume, regional myelination, as well as inter- and intra-regional coactivity. We also computed the correlation strengths of seven major cortical networks. We then built an interpretable model to predict global network phenotypes from local regional phenotypes. Finally, we used this model to mechanistically associate individual variation in network phenotypes with individual variation in regional gene expression. We used permutation testing to compute the significance of these associations.
Results: Our local-global neuroimaging modeling revealed significant associations (nominal $p<0.01$) between putamen volume and 5/7 cortical networks associated with motor control, reward processing, and learning. Our integration of these results with data on regional gene expression identified specific genes (e.g., *SLC5A9* and *CLEC12B*) significantly associated with the default and dorsal attention networks, and a separate group of genes (e.g., *STAP1*, *PROZ*, *PKD1L1*) associated with the control and limbic networks. Separately, we found that these genes exhibited high expression levels in the corresponding brain regions of the Allen Human Brain atlas.
Conclusions: Collectively, our integrative analysis allowed us to associate regional gene expression with individual variation in human brain network organization. This approach ultimately makes it possible to advance our understanding of the transcriptomic underpinnings in healthy brains and consider how this organization varies in diseased populations.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4174 GenoGAN: Synthesizing genomic data with stable deep generative networks for UK Biobank, Taiwan Biobank, and 1000 Genomes Project.

Authors:

E. Lim[1,2], A. Lim[1,3], C. Fann[1]; [1]Academia Sinica, Taipei, Taiwan, [2]Georgia Inst. of Technology, Atlanta, GA, [3]ASUS Intelligent Cloud Services (AICS), Taipei, Taiwan

Background: The limited sample size of certain rare diseases and minority populations continues to pose a significant challenge. However, instead of solely focusing on collecting more samples, an alternative approach involves generating realistic synthetic human genomic data. One method for that is through the use of Generative Adversarial Networks (GANs). These networks have shown encouraging results on specific datasets, however, often require extensive hyperparameter tuning and can struggle to converge successfully. In this regard, we introduce GenoGAN. **Methods:** GenoGAN incorporates Wasserstein generative adversarial networks with gradient penalty and Heaviside function to ensure training stability, overcoming issues like the vanishing gradient problem and mode collapse. GenoGAN takes input data with set of sequences of haplotype pairs and output user-defined n-fold of synthetic data. We evaluated GenoGAN's performance using selected sequences of phased haplotype data from chromosome 1. The dataset consisted of individuals with type 1 diabetes (T1D: n=3,698) and individuals without any reported diseases (Healthy: n=51,857) sourced from the UK Biobank. The input data included T1D associated rs6679677 and rs2476601 in *PTPN22*. We validated the results with Taiwan Biobank and the no-phenotype-info 1000 Genomes Project (1KGP). **Results:** GenoGAN generated synthetic datasets (T1D′, Healthy′) that doubled the size of the original samples, with small distance from the real data and similar minor allele frequencies (MAFs). The Jensen-Shannon divergence (JSD) between T1D and T1D′ was 0.276, and between Healthy and Healthy′ was 0.275. The cosine distances (CD) of T1D, T1D′ and Healthy, Healthy′ were 0.110 and 0.109, respectively. Notably, GenoGAN performed well even with lower frequency alleles. The minor allele frequencies (MAFs) observed at rs6679677 were T1D=0.137, T1D′=0.160, Healthy=0.101, and Healthy′=0.110. Similarly, for rs2476601, the MAFs were T1D=0.137, T1D′=0.145, Healthy=0.101, and Healthy′=0.111. However, it is worth noting that GenoGAN's performance was less favorable when the sample size was too small, e.g., with East Asian ancestry from 1KGP (n=504), resulting in higher JSD(EAS, EAS′)=0.516 and CD(EAS, EAS′)=0.385. **Conclusion:** The results demonstrate that our proposed GenoGAN can capture the complex multidimensional structure of the data distribution, thereby providing a novel method for human genomic data synthesis. We envision that by scaling up GenoGAN with a large number of GPUs, its potential for enhancing disease risk prediction for rare diseases and in minority populations can be significantly amplified.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4175 Genome wide association study (GWAS) of 110 putative fatty acids, eicosanoids, and related oxylipins in the Hispanic Community Health Study/Study of Latinos reveals shared and distinct genetic architecture

Authors:

C. Downie[1], H. M. Highland[1], A. G. Lilly[1], B. M. Welch[2], R. Kaplan[3], S. Cheng[4], B. Thyagarajan[5], M. Jain[6], C. L. Avery[1]; [1]Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, [2]Univ. of Nevada, Reno, Reno, NV, [3]Albert Einstein Coll. of Med., Bronx, NY, [4]Cedars Sinai, Smidt Heart Inst., Los Angeles, CA, [5]Univ. of Minnesota, Minneapolis, MN, [6]Sapient, San Diego, CA

Introduction: Chronic low-grade inflammation is a pathological feature of multiple cardiometabolic traits, including obesity and type 2 diabetes. Bioactive polyunsaturated fatty acid (PUFA)-derived eicosanoids and related oxylipin compounds are involved in the modulation of inflammation. To-date, no large-scale GWAS of eicosanoids has been conducted. **Methods:** Analyses were performed in n=11,584 participants with measured genetic and oxylipin data from the baseline study visit of the Hispanic Community Health Study/Study of Latinos (HCHS/SOL), a multi-center community-based cohort of Hispanic/Latinos recruited from four US communities between 2008-2011. Putative oxylipin spectral peaks were quantified from plasma samples via directed non-targeted liquid chromatography-mass spectrometry. HCHS/SOL participants were genotyped on the MEGA array and imputed to the TOPMed imputation panel. We performed GWAS of 110 putative fatty acid and oxylipin peaks using SUGEN, adjusting for age, age^2, sex, recruitment site, background group, and 10 ancestral principal components.**Results:** The spectral peaks were grouped into three broad classes: putative free fatty acids, very long chain dicarboxylic acids, and eicosanoids and related oxylipins. 61 (55.5%) of the 110 spectral peaks had at least one genome-wide significant locus (Bonferroni-adjusted $p < 5.15E-11$), and most peaks were associated with only one or two significant loci. Putative fatty acid peaks displayed primarily shared loci (e.g., *FADS*, *FUT2*); several of these loci were also associated with putative eicosanoid peaks. The *SLCO1B1* locus, which encodes the OATP1B1 uptake transporter protein, was associated with multiple eicosanoid peaks, including putative prostaglandins, leukotrienes, and putatively novel oxylipins. Other loci were only associated with one eicosanoid peak; for example, *UGT2B7,* which encodes a protein involved in the conjugation and elimination of xenobiotics and endogenous compounds, was associated with a putative diHOME peak.**Conclusions:** Our results suggest that putative fatty acids and related oxylipins have an oligogenic genetic architecture. Although some genetic loci are highly shared across similar classes of compounds, there are also some distinct genetic effects. These insights on the genetic architecture of oxylipin compounds may help inform the development of eicosanoid-targeted therapeutics.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4176 Genome-wide association identifies novel etiological insights associated with Parkinson's Disease in African and African admixed populations.

Authors:

M. Makarious[1], M. Rizig[2], S. Bandres-Ciga[1], O. Ojo[3], P. Wild Crea[1], O. Okunoye[2], Nigeria Parkinson Disease Research Network, nternational Parkinson's Disease Genomics Consortium - Africa (IPDGC Africa), Black and African American Connections to Parkinson's Disease (BLAAC PD) Study Group, the 23andMe Research Team, M. Nalls[4], C. Blauwendraat[1], H. Houlden[5], A. Singleton[6], N. Okubadejo[3], Global Parkinson's Genetics Program; [1]NIH, Bethesda, MD, [2]UCL, London, United Kingdom, [3]Univ. of Lagos, Idi Araba, Lagos State, Nigeria, [4]Natl. Inst Aging, Bethesda, MD, [5]Inst Neurology, London, United Kingdom, [6]NIA, Bethesda, MD

Aim: The central objective of this study is to perform a comprehensive genome-wide assessment of Parkinson's disease (PD) in populations of African and African admixed descent. The goal is to characterize ancestry-specific risk, genetic variations, and to identify novel mechanisms of disease etiology.Background: Understanding genetic mechanisms across diverse populations can provide unique insights into complex traits like PD. Historically, our understanding of the genetics of PD has been disproportionately based on European, Asian, and Latin American populations. This has led to a significant gap in our knowledge about the disease's genetics and clinical characteristics in underrepresented populations, particularly individuals of African and African admixed ancestries.Materials & Methods: We performed an extensive genome-wide assessment of PD in a large cohort comprising 197,918 individuals. The cohort included 1,488 PD cases and 196,430 controls, all of African and African admixed ancestry. The study analyzed a range of factors, including ancestry-specific risk, differential haplotype structure, coding and structural genetic variation, and polygenic risk profiling.Results: The analysis identified a novel common risk factor for PD at the GBA1 locus, specifically the rs3115534-G variant. This risk factor also appeared to influence the age at onset. Interestingly, this factor was found to be rare in non-African/African admixed populations. Unlike previously identified GBA1 associated disease risk variants, this newly discovered signal potentially mediates PD risk through expression quantitative trait locus (eQTL) mechanisms, suggesting a novel functional mechanism.Discussion: Our study identifies a novel African-ancestry genetic risk factor in GBA1 as a risk factor of PD in the African and African admixed populations. This discovery is in stark contrast to previous work focused on Northern European populations. It underscores the importance of understanding ancestry-specific genetic risk in complex diseases, and the need for equitable inclusion of diverse groups in PD clinical trials. Given the genetics of these underrepresented populations, their inclusion in such studies opens new avenues towards RNA-based and other therapeutic strategies aimed at reducing lifetime risk of PD. This is particularly important as the field moves towards precision medicine in PD clinical trials.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4177 Genome-wide association studies in a large Korean cohort identifies novel quantitative trait loci for 36 traits and sheds light on the genetic architecture of complex traits

Authors:

Y. Jee[1], A. Martin[2,3,4], P. Kraft[5]; [1]Harvard T.H. Chan Sch. of Publ. Hlth., Boston, MA, [2]Analytic and Translational Genetics Unit, Massachusetts Gen. Hosp., Boston, MA, [3]Stanley Ctr. for Psychiatric Res., Broad Inst. of MIT and Harvard, Cambridge, Cambridge, MA, [4]Dept. of Med., Harvard Med. Sch., Boston, MA, [5]Transdivisional Res. Program, Div. of Cancer Epidemiology and Genetics, Natl. Cancer Inst., NIH, Bethesda, MD

Genome-wide association studies (GWAS) have identified many loci associated with complex traits. However, most of these studies were conducted in populations of European ancestry, which has limited opportunities for biological discovery.

We conducted GWAS of 36 human quantitative traits in the Korean Cancer Prevention Study-II (KCPS2) Biobank (n=153,354). Our analysis discovered 2,453 associated loci (range 2-312 loci, median 45), including 922 novel loci (range 0-192 loci, median 12) that were not reported in previous GWAS related to the corresponding trait. We also identified widespread pleiotropy in which 4,506 gene regions contained variants associated with one or more traits (range 1-25 traits, median 2.3). For example, variants near *ALDH2* were associated with 25 traits, including blood pressure and alcohol-related traits. Following meta-analysis of 20 traits across KCPS2, Biobank Japan (BBJ), and Korean Genome and Epidemiology Study, we further identified 1,109 novel loci (range 1-171 loci, median 33).

We investigated the genetic architecture of 12 traits in KCPS2 and compared with previously published BBJ and UK Biobank (UKB). While the S parameters linking minor allele frequency and effect sizes were similar across the biobanks (median -0.5), the polygenicity and heritability parameters estimated in UKB were mostly higher than those in KCPS2 and BBJ. Compared to BBJ, KCPS2 showed higher heritability parameters for body mass index, height, and blood pressure but lower heritability for most of the hematological traits such as white blood cell and hemoglobin.

To identify potential causal variants, we performed fine-mapping using SuSiE. For example, we fine-mapped 25 traits associated with the region spanning *ALDH2* on chromosome 12 (± 500kb from rs671, which is known to be functionally related to alcohol metabolism). 1,476 variants in this region were fine-mapped to a total of 48 credible sets, among which 14 contain exactly one variant (range 1-568 variants, mean 43.9). The posterior inclusion probability of rs671 was greater than 90% for 7 traits including alcohol intake, gamma-glutamyl transferase, and blood pressure.

Our findings highlight how broadening the population diversity of GWAS samples can aid discovery (e.g., the minor allele of rs671 is common in East Asian but quite rare in European ancestry populations). Our results also provide insights into the genetic architecture of complex traits in East Asian populations. By increasing the sample size and ancestral diversity of GWAS samples, our analysis may help identify novel targets for prevention and treatment, thus offering equitable access to precision medicine to diverse populations.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4178 † Genome-wide association studies with Japanese National Health Insurance System.

Authors:

Y. Yamasaki[1], N. Nishiya[1], Y. Sutoh[1], S. Komaki[1], S. Minabe[1], H. Ohmomo[1], K. Tanno[1], K. Asahi[1], Y. Ishigaki[1], M. Sasaki[1], A. Shimizu[2]; [1]Iwate Tohoku Med. Megabank Organization, Iwate Med. Univ., Yahaba, Japan, [2]Iwate Tohoku Med. Megabank Organization, Iwate Med. Univ., Iwate, Japan

Large-scale genome-wide association analyses (GWAS) have contributed to the development of precise polygenic models for accurately estimating disease risks. Community cohort studies are used to estimate the effect size of a variant in the general population while minimizing bias. However, accurately defining case subjects based on the diagnosis can be challenging in these studies. To overcome these limitations, a medical data source that covers most of the target population is promising. Here, we present the results of a feasibility study of GWAS using health insurance claim data retrieved from the Japanese National Health Insurance System, which covers a broad segment of the general population.Health insurance claim data for approximately 14,000 residents of the Iwate prefecture, Japan, participating in the Tohoku Medical Megabank Project was retrieved from the Japanese National Health Insurance System. A case-control GWAS was performed using SAIGE, with age, sex, and the top 10 principal components of genotypes as covariates. Different case definitions were compared, “including”/“excluding” health insurance claim data in the same samples. The analysis focused on hypertension, diabetes, dyslipidemia, Alzheimer's disease (AD), and Parkinson's disease (PD), for which health insurance claim data is substantial.When health insurance claim data was not included in the definition, the percentages of patients with hypertension, diabetes, dyslipidemia, AD, and PD were 46.6%, 11.2%, 53.6%, 0.2%, and 0.2%, respectively. In contrast, when health insurance claim data was included in the definition, the percentages were 50.6%, 12.2%, 59.9%, 1.3%, and 0.7%, respectively. In GWAS results of hypertension, diabetes and PD, the samples including and excluding health insurance claim data showed no significant differences. However, for dyslipidemia, genes around 5p13-14 were detected as seen in previous GWAS studies of serum low-density lipoprotein cholesterol but only in the definition including health insurance claim data. Additionally, the apolipoprotein E (APOE) gene, which has been repeatedly associated with AD in previous case-control AD-GWAS studies, was only detected when using the definition with health insurance claim data for AD. APOE has never been detected in previous AD-GWAS studies based on a community cohort, even in recent large studies using UK Biobank data. Therefore, this study could also be the first replication of the association between APOE and AD in a community-based cohort of the general population. These findings indicate the validity of using health insurance claim data in GWAS on community-based cohorts.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4179 Genome-Wide Association Study (GWAS) by proxy (GWAX) of risk for Alzheimer's disease and related disorders (AD/ADRD) with optical coherence tomography (OCT) retinal thickness endophenotype in the Tohoku Medical Megabank Organization (ToMMo) Eye Study, a large Japanese cohort

Authors:

M. Taira[1,2], A. J. Saykin[3,4], N. Fuse[1,5]; [1]Dept. of Integrative Genomics, Tohoku Med. Megabank Organization, Tohoku Univ., Sendai, Japan, [2]Div. of Neurology, Dept. of NeuroSci. & Sensory Organs, Graduate of Med., Tohoku Univ., Sendai, Japan, [3]Indiana Univ. Sch. of Med., Indianapolis, IN, [4]Indiana Alzheimer's Disease Res. Ctr., Indiana Univ., Indianapolis, IN, [5]Tohoku Univ. Advanced Res. Ctr. for Innovations in Next-Generation Med., Sendai, Japan

Background: AD/ADRD is age-related and highly heritable. GWAX is an efficient research strategy for population-based studies of complex diseases, in which first-degree relatives of cases serve as surrogate phenotypes. However, recent GWAX have indicated low heritability in AD/ADRD. Retinal imaging by OCT is noninvasive, repeatable, and can detect structural changes of the retina that thin out with progressive neurodegeneration. The ToMMo Eye Study is an ophthalmology cohort consisting of 95K individuals included in the TMM Project, a prospective cohort study of 120K adults. Combining retinal thickness and genetics may highly be informative for AD/ADRD pathophysiology, but so far GWAX combined with retinal thickness has not been performed. Method: Retinal nerve fiber layer and ganglion cell layer thickness calculated on subjects' SD-OCT (Topcon OCT-2000) scans. 5,250 pseudo-cases and 42,305 pseudo-controls as discovery, and 916 pseudo-cases and 12,483 pseudo-controls as replication, were analyzed in GWAX of AD/ADRD excluding close relatives. Imputation was performed using beagle5.4. GWAX using plink2 included 65 million variants that satisfied RSQR > 0.3. 10 principal components of the population structure were used as covariates. AD/ADRD GWAX with retinal thickness for 3,071 pseudo-cases and 20,308 pseudo-controls performed as meta-analysis. Results: On hg37, in AD/ADRD GWAX, the strongest association was *APOE* (chr19: Base Position (BP) 45411941, p < 3.0e-36) as expected. This gene effect stood out via mother's dementia (p < 1.13e-30) but had no correlation via father's dementia. Six loci (1:91939115, 10:134466627, 12:6538149, 18:59219633, 11:33590240, 4:54422066) were also statistically significant (p < 5e-8) and 61 loci (p < 1e-6) were suggestive. In AD/ADRD GWAX with OCT, the strongest association was the same locus on *APOE* (p < 2.4e-21). *INPP5A* (10:134388125, p <3.3e-9) and *CDH20* (18:59219633, p < 7.5e-9) remained. *APOE* showed the strongest risk heritability. *INPP5A* may dysregulate of phosphoinositide (PI) and PI- 5-phosphatase in AD. The CDH (cadherin superfamily) genes code vascular/cell adhesion, the regulation of tissue organization, and morphogenesis. *CDH13* was already reported on AD/ADRD GWAS in a Caucasian population. Some other genes are reportedly associated with psychiatric disorders relevant to ADRD. Conclusion: We report the first AD/ADRD risk gene investigation by GWAX with retinal thickness worldwide and highlight possible roles and pathways for susceptibility loci with high odds ratio and *ApoE* effects in Japanese ethnicity, that correlated with retinal thickness in a Japanese population.

---PAGEBREAK---

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4180 Genome-wide association study of emphysema- and airway-predominant deep learning subtypes.

Authors:

D. Hill[1], K. Kim[2], M. Moll[2], M. Torop[1], A. Masoomi[1], S. Bodduluri[3], P. J. Castaldi[2], B. D. Hobbs[2], J. Dy[1], S. P. Bhatt[3], M. H. Cho[2]; [1]Northeastern Univ., Boston, MA, [2]Brigham and Women's Hosp., Boston, MA, [3]Univ. of Alabama at Birmingham, Birmingham, AL

Background: Chronic obstructive pulmonary disease (COPD) is characterized by a combination of two different structural diseases: emphysema and airway disease. The relative contribution from these two diseases can be used to categorize patients into emphysema-predominant and airway-predominant phenotypes, which is established with CT measures of emphysema and airway remodeling. While CT can differentiate between these two phenotypes, CT results are often unavailable due to expense and entail radiation exposure. In previous work, a neural network model (SpiroNet) trained on spirometry volume-time curves was effective in differentiating emphysema-predominant and airway-predominant subtypes on CT scans. Whether these phenotypes have specific genetic susceptibility, differing from summative measures of lung function and COPD, is unknown.
Methods: We used SpiroNet, a fully convolutional neural network, to predict emphysema-predominant and airway-predominant phenotypes from volume-time exhalation curves from the UK Biobank (N=352,684) and performed GWAS on each phenotype as a binary indicator (Emphysema: case 30,677, control 322,007; Airway: case 101,849, control 250,835), adjusting for age, sex, pack-years, ever smoking status, batch, and ancestry principal components using plink with Firth fallback. We considered $< 1x10^{-10}$ as genome-wide significant ($5x10^{-9}$, two traits), and loci as associations > 500kb apart. We compared results to previously published GWAS using COPD and measures of lung function (forced expiratory volume in one second, forced vital capacity, and the ratio of these two measures).
Results: We identified 83 loci between the two phenotypes (3 for emphysema and 80 for airway) at genome-wide significance. Seventy-nine of these loci were within 500kb of previously described regions associated with lung function or COPD, with the exception of 4 airway-associated variants in the HLA region (spanning HLAA and ZKSCAN4), and one airway-associated variant > 500kb away from the lead variant near HHIP). The top loci for emphysema were near EFEMP1 and CCDC91, and for airway, HHIP and NPNT.
Conclusions: Emphysema-predominant and airway-predominant phenotypes can be established in the UK Biobank by using a neural network model applied to spirometry, and are associated with different genetic loci. Deep learning models can be used to infer disease subtypes in large population-based cohorts and identify subtype-specific genetic risk variants.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4181 Genome-wide association study of fibrotic disease across organs and systems.

Authors:

E. Joof[1], G. M. Massen[2], O. C. Leavy[1], G. Parcesepe[1], I. D. Stewart[2], R. G. Jenkins[2], J. K. Quint[2], G. Aithal[3], C. J. Scotton[4], D. P. Auer[3], S. Francis[3], L. V. Wain[1], R. J. Allen[1], DEMISTFI Consortium; [1]Univ. of Leicester, Leicester, United Kingdom, [2]Imperial Coll. London, London, United Kingdom, [3]Univ. of Nottingham, Nottingham, United Kingdom, [4]Univ. of Exeter, Exeter, United Kingdom

Rationale: Fibrosis is a major factor in global mortality and can affect virtually every organ system. The incidence of fibrotic conditions in many different organ systems hinders the development of targeted therapies. A potential strategy to overcome this challenge is to target common genetic mechanisms and pathways shared across organs.
Objectives: We sought to identify shared genetic signals across different fibrotic organs and systems, and for fibrotic multimorbidity.
Methods: The study utilised samples of people with European ancestry from the UK Biobank. Fibrotic diseases, as defined by a published consensus list, were combined across 13 organs or systems (Respiratory, Liver, Bile, Cardiomyopathy, Intestinal-Pancreatic, Integumentary, Skeletal, Systemic, Reproductive, Urinary, Blood-vessel, Atherosclerosis and Diabetes). Individuals that had a fibrotic condition recorded in their hospital or mortality records were defined as cases (and all others defined as controls). For the fibrotic multimorbidity analysis, individuals with a fibrotic disease(s) in more than one organ/system were defined as cases and the rest as controls. Genome-wide association analyses were conducted assuming an additive genetic effect and conditioning on the first ten genetic principal components. LD score regression was used to estimate global genetic correlation and colocalisation analysis was used to test for shared causal genetic variants between organs and systems.
Results: We detected multiple genome-wide significant signals including 11 signals of association for fibrotic multimorbidity. Colocalisation analyses revealed at least six shared common genetic causal variants. A signal at *PTPN22* (a gene known to be associated with conditions such as Systemic Lupus Erythematosus and Rheumatoid Arthritis) overlapped between two single fibrotic organ/system GWAS and the fibrotic multimorbidity GWAS.
Conclusions: Functional follow-up of these shared signals may reveal potential causal pathways of general fibrosis and provide new insight into signalling responses shared by different fibrotic diseases which can be leveraged in the development of new therapies. We are currently expanding the analyses to other ancestry groups to identify more genetic signals that overlap between different fibrotic diseases.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4182 Genome-wide Implementation of Multi-population Joint Analysis Marginal Summary Statistics (mJAM) and its Applications in Polygenic Risk Score Models

Authors:

J. Shen[1], J. Baurley[2], J. He[1], G. King[1], C. A. Haiman[1], D. V. Conti[1]; [1]Univ. of Southern California, Los Angeles, CA, [2]BioRealm LLC, Mammoth Lakes, CA

Previously we proposed a multi-population fine-mapping approach called "mJAM" to jointly analyze marginal GWAS summary statistics. mJAM effectively fits a multi-SNP model within each population with population-specific reference LD and then performs a fixed-effect meta-analysis of the joint model while incorporating a Bayesian 'g-prior' for robust estimation. The resulting model provides the flexibility to identify noteworthy SNPs using existing feature selection approaches. We propose a scalable version of the previous region-wise mJAM implementation that can be applied to genome-wide data ("GW-mJAM"). We illustrated GW-mJAM with forward selection as the main approach for selecting index SNPs, applied GW-mJAM on the latest prostate cancer summary statistics with 4 populations, and constructed PRS using the index SNPs selected by GW-mJAM. Since mJAM model estimates the joint SNP effects at a multi-population level and does not assume population-specific weighting, resultant PRS built from GW-mJAM can be applied to individuals of any ancestry without estimating ancestry or using self-identifying population information. We evaluated GW-mJAM's PRS using incident data from the MultiEthnic Cohort Study (5 populations, 2,885 cases, 25,904 controls; Kolonel et al., 2000). GW-mJAM selected 471 index SNPs that are conditionally genome-wide significant. Among men of African ancestry in the MEC (592 cases and 1,957 controls), the AUC of GW-mJAM PRS is 0.800 (0.782 - 0.818) and the OR of 1-standard-deviation increase in PRS is 1.96 (1.76 - 2.18). Among men of European ancestry (530 cases and 5,571 controls), the AUC of GW-mJAM PRS is 0.823 (0.807 - 0.840) and the OR of 1sd PRS model is 2.25 (2.03 - 2.50). By lowering the significance threshold for index SNP selection, GW-mJAM includes additional regions and SNPs across the genome. However, evaluation performance in the MEC did not improve with more variants being included in the PRS model. For example, using a p-value threshold of 10^{-6}, GW-mJAM selected 601 index SNPs but AUC and 1sd OR showed similar decrement among all populations in MEC (AUC = 0.795 (0.778 - 0.813); 1sd OR = 1.93 (1.73 - 2.15) among men of African ancestry). In addition to GW-mJAM with forward selection, we incorporate GW-mJAM with other feature selection approaches as well as apply GW-mJAM to three other cancer types (breast, colorectal, and lung) with available multi-population summary statistics. PRS models built from GW-mJAM's results are evaluated and compared with other GW PRS approaches such as PRS-CSx (Ruan et al., 2022) to investigate the impact of trait characteristics, quality of reference panel, and methodology on test performance.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4183 Graph Attention Networks-based Predictive Model for Healthcare using EHR

Authors:

X. Huang[1], Z. Wang[1], J. De Jong[2]; [1]Yale Univ., New Haven, CT, [2]Boehringer Ingelheim Pharma GmbH & Co.KG, Biberach, Germany

The widespread digitalization of patient data via electronic health records (EHRs) has brought unprecedented opportunities for risk modeling. While numerous risk models have been developed for a wide spectrum of diseases, they often overlook the inclusion of family relationships and histories. Additionally, deep learning predictive models, although effective, suffer from a lack of interpretability. Moreover, there are limited interpretable models that utilize the prior medical ontology. To leverage these available information, we developed a Graph Attention neTworks-based predictive model for EHR (GATEHR) to simultaneously predict future disease risks for thousands of diagnosis codes using patients' EHRs, family health history and knowledge graph of medical ontology. This model incorporates family health history, which has been well accepted as a major predictor for a range of diseases, through electronically constructed family pedigrees (e-pedigrees), and knowledge graph of medical ontology, via graph attention networks (GAT). Furthermore, it is able to enhance the model interpretability by assessing the impact of disease risk factors using attention-based feature importance. We compared predictive performance of GATEHR and other existing methods, assessed the interpretability of our model. Precision, recall, F1 score, and area under the curve (AUC) were selected to measure the performance. The results showed that our graph-based neural network model outperforms the other baselines in disease risk prediction tasks. Besides, our model demonstrated the impact of family health history among patients with Nonalcoholic Fatty Liver Disease (NAFLD) and Nonalcoholic Steatohepatitis (NASH). The results emphasized the significant potential and value of integrating family health history and medical ontology into GAT for disease risk prediction. This study offers improved predictive accuracy and interpretable insights into disease risk models.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4184 graphPred: multi-ancestry polygenic risk prediction with millions of SNPs in minutes

Authors:

P. Salehi Nowbandegani[1], G. Genovese[2], L. O'Connor[1]; [1]Broad Inst. of MIT and Harvard, Cambridge, MA, [2]Broad Inst, Cambridge, MA

State-of-the-art polygenic risk score (PRS) methods employ computationally intensive Bayesian algorithms, leading to compromises that affect their prediction accuracy. Most rely on HapMap3 SNPs, representing <20% of common SNPs in European populations, and they typically analyze data from a single ancestry group, causing non-transferability. These limitations apply both to individual-level methods and to those that operate on GWAS summary association statistics.

For summary statistic-based methods, the computational bottleneck is their use of slow and memory intensive LD matrix operations. We recently introduced linkage disequilibrium graphical models (LDGMs), which accelerate LD matrix operations dramatically with no loss in accuracy. We developed a new PRS method, graphPred, which leverages LDGMs to perform exceptionally fast Bayesian inference without the computational compromises of existing methods. graphPred is a Gibbs sampler for a generalized SuSiE model (Wang et al. 2020). It uses a novel sampling scheme that greatly accelerates convergence. graphPred incorporates data from multiple ancestry groups, it jointly models almost every SNP at >1% frequency in 1000 Genomes, it learns a flexible effect-size distribution during sampling, and it can incorporate functional annotations into the prior distribution.

We evaluated the performance of graphPred in analyses of real and simulated data. In a wide range of realistic simulations, with different sample sizes and genetic architectures, it outperforms the widely used PRS-CS (Ge et al. 2019) by 16% in a realistic simulation with N=100k, and by over 20% in well-powered simulations with high sample size (N=1M) or low polygenicity; moreover, it ran ten times faster. The difference in performance was mostly eliminated when artificially restricting graphPred to HapMap3 SNPs, indicating that its ability to model a larger number of SNPs confers a significant advantage. In a preliminary analysis of three real phenotypes (height, BMI and T2D), the improvement was even larger, including in individuals of non-European ancestry. graphPred is implemented in C as a plugin for bcftools, and it runs in under 10 minutes. Our results illustrate how linkage disequilibrium graphical models enable the development of powerful new methods.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4185 Group-penalized exponential tilt model for identification of differentially methylated genes in epigenetic association studies

Authors:

H. Sun; Pusan Natl. Univ., Busan, Korea, Republic of

DNA methylation is a representative epigenetic change that occurs in our body and plays an essential role in regulating gene expression as well as in cancer progressing. Identification of differentially methylated genes between two biological conditions has been popularly conducted in epigenetic association studies. However, most of statistical methods aim to detect differences in mean methylation levels between two conditions. So, they are limited to identify differences in methylation variances which have been recently observed in cancer research. We propose a new statistical method based on a group-penalized exponential tilt model that essentially combines an exponential tilt model and group lasso. The proposed method can identify differentially methylated genes when two biological conditions are different in methylation mean only, methylation variance only, or both. It can be extended to genes with overlapping CpG sties when group lasso is replaced by overlapping group lasso. The proposed method is able to prioritize differentially methylated genes based on their selection probability, which is computed by bootstrap resampling. In our extensive simulation study, we demonstrated that the proposed method has superior selection performance, compared with the existing statistical methods developed for detection of differentially methylated genes. We also applied it to 450K DNA methylation data of The Cancer Genome Atlas Breast Invasive Carcinoma Collection (TCGA-BRCA). We were able to identify potentially cancer-related genes.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4186 † GUIDE learns interpretable latent factors from genetic association studies

Authors:

D. Lazarev[1,2,3], A. Bloemendal[1,3], G. Chau[1,3], C. Churchhouse[1,3], B. Neale[3,1]; [1]Broad Inst. of MIT and Harvard, Cambridge, MA, [2]Massachusetts Inst. of Technology, Dept. of Mathematics, Cambridge, MA, [3]Massachusetts Gen. Hosp., Boston, MA

Learning a set of modules, or latent factors, mediating the effects of a large set of SNPs on some set of complex traits can help inform disease classification and elucidate complex disease mechanisms, otherwise muddled by polygenic and pleiotropic effects. We propose a method, Genetic Unmixing by Independent DEcomposition (GUIDE), to estimate a set of statistically independent latent factors that best extract the information encoded in the underlying genetic architecture by picking a basis to express these factors that is simultaneously informed by the SNPs and traits corresponding to a set of summary statistics. In particular, GUIDE builds a three-layer network consisting of SNPs, traits, and latent factors that mediate their interactions, and picks a basis for the latent factors that maximizes the sparsity of interlayer weights, resulting in factors that are more enriched for particular traits and load onto sets of genetic variants that are more significantly associated with given diseases and disease pathways. For example, we built a GUIDE model using UK Biobank data and focused on Alzheimer's disease (AD), peripheral vascular disease, and high cholesterol, finding that a common latent factor that explains a large part of the genetic variance for all three traits specifically and significantly implicates genes in the cholesterol metabolism pathway. Moreover, the top GUIDE factor for AD—explaining 87.5% of the genetic variance of the disease—is highly enriched for the *APOE* genes, with 49 of the top 54 loci in the 19:45386467-45445517 region near *APOE* contributing 50.9% of the variance of this latent factor. By contrast, the top factor for the state-of-the-art method based on principal component analysis explains only 10.2% of the genetic variance of AD, mostly picking up a host of related genes such as *APOC*, *APOB*, *LDLR*, among others, without significant enrichment for *APOE*, and with the top 50 loci—spanning most of the chromosomes—contributing only 12.7% of this factor's variance. Moreover, GUIDE prioritizes genes, such as *RBFOX1* (chr. 16), *CSMD1* (chr. 8), and *PCSK9* (chr. 1), which have only recently been implicated in AD. GUIDE thus allows using summary statistics for the construction of parsimonious and interpretable network models that deconvolve the shared and specific effects of SNPs on complex traits.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4187 GWAS conditioned by SNP-specific shape effects identifies 11 loci underlying shape covariation of the cerebral cortex and cranial vault

Authors:

S. Goovaerts[1], M. Yuan[1], H. Hoskens[2], N. Herrick[3], M. D. Shriver[4], J. R. Shaffer[5], S. M. Weinberg[5], S. Walsh[3], P. Claes[1]; [1]KU Leuven, Leuven, Belgium, [2]Univ. of Calgary, Calgary, AB, Canada, [3]Indiana Univ. - Purdue Univ. Indianapolis, Indianapolis, IN, [4]Penn State Univ, University Park, PA, [5]Univ. of Pittsburgh, Pittsburgh, PA

Covariation in brain and cranial vault shape inevitably arises from their structural relationship and proximity during development, where close coordination between both tissues is required for “normal” craniofacial development. Several studies have suggested cell-to-cell signaling as an important driver for their integrated co-development, with the FGF, BMP, Wnt, and Hedgehog signaling pathways playing essential roles in both brain and cranial vault development. Functional studies have supported this hypothesis but are limited to candidate genes and pathways. GWAS provides an unbiased framework and could potentially implicate novel genes and mechanisms underlying brain-vault covariation. A recent study scanned the mouse genome and identified a single locus associated with global axes of shape covariation between the brain and cranial vault. Here, we propose a novel approach that differs in that it extracts a SNP-specific latent phenotypic axis for one multivariate trait that is used for conditioning on another multivariate trait under coinvestigation. Specifically, we assessed the dependency of genome-wide associations for multivariate 3D cortical and cranial vault shape derived from MRIs of 4,148 participants of recent European ancestry from the ABCD study cohort. We demonstrate an improved discovery rate over previous methods and identify genes involved in signal transduction (*BMP2*, *PTHLH*, *ABR*, and *TIAM2*) and regulation of ossification and osteoblast differentiation (*RUNX2*, *BMP2*, *PTHLH*, and *DLX5*). Among the 11 identified loci, 4 could not be identified in separate GWASs on brain and cranial vault shape using the same data and were hence a product of our novel approach. Altogether, we demonstrate a new, powerful method for identifying the genetics underlying brain-vault covariation, thereby identifying genes that support the current hypothesis that close regulation of calvarial ossification is crucial for their integrated co-development and morphological covariation.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4188 GWAS of 5,811 hepatitis B virus infection cases and 152,913 controls of Koreans with replication analysis of 171,822 of Japanese

Authors:

S-H. Jee, K. Jung, H. Kimm; Graduate Sch. of Publ. Hlth., Yonsei Univ., Seoul, Korea, Republic of

Hepatitis B virus (HBV) infection is a major causal factor in the development of liver cancer. So far, genetic epidemiological studies on HBV have been small in scale. The objective of this study was to perform genome wide association study (GWAS) analysis to identify new genetic factors related to HBV in Koreans and Japanese. Korean Cancer Research and Prevention (KCPS-II) data of 158,724 people and summary statistic of Biobank of Japan (BBJ) data of 171,822 people were obtained. Novel SNPs were found through GCTA-COJO analysis. Related genes were mapped through FUMA GWAS analysis in KCPS-II data. BBJ data were used for replication analysis. In KCPS-II data, 24,346 SNPs with $P < 10^{-8}$ were identified from GWAS analysis. Of these, 5,521 ($P < 10^{-8}$)/19,218 SNPs were replicated/identified in BBJ. Finally, 23 SNPs were identified as Novel SNPs independent of previously reported SNPs. Of these, 14 ($P < 10^{-8}$)/20 SNPs were replicated/identified in BBJ. In addition, the first 232 genes were mapped. Among them, 54 genes with a CADD score of 20 or more were identified. This is the largest HBV GWAS study conducted in Asian countries. It replicated previously reported results and found new genetic factors. Results of this study could be used as a basis for further research on prevention and treatment of HBV infection and liver cancer.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4189 GXwasR: A tool for conducting sex-aware quality control, association analysis, and testing various models of sex-dependent genetic effects in complex traits.

Authors:

B. Bose[1], L. L. Pesce[1], E. Khramtsova[2], B. Stranger[1]; [1]Northwestern Univ., Chicago, IL, [2]Janssen Pharmaceuticals, Spring House, PA

Motivation: Many human complex traits and diseases exhibit sex-differentiated characteristics, yet the causes of these differences are largely unexplored. To understand the processes underlying these differences, sex-aware genomic analyses are crucial and so is a need for computational tools that implement currently accepted state-of-the-art best practices for performing sex-aware quality control (QC) and genetic analysis.
Our Tool: We present a new software tool, GXwasR, that in addition to enabling a standard genome-wide association study (GWAS), can perform a sex-aware GWAS. Standard QC and association functions are provided, together with sex-aware QC of genotype data. It performs sex-combined and sex-stratified GWAS and X-Chromosome Wide Association Studies (XWAS). Importantly this tool has functionality for testing genetic associations between phenotype and variants on the X-chromosome (chrX). ChrX requires different statistical association models than autosomes because of, e.g., chrX inactivation (XCI), unique patterns of recombination.
Functionalities: GXwasR includes several functions grouped in six different categories (A-E) enabling a comprehensive pipeline for sex-aware genetic analysis: (A) sex-aware pre-imputation and post-imputation QC of genotype data, such as filtering variants considering differential missingness between cases and controls, handling of pseudo-autosomal regions, X-transposed regions and ampliconic regions of chrX, sex-specific filtering of variants based-on Hardy-Weinberg Equilibrium and Minor Allele Frequency etc., (B) sex-combined and sex-stratified GWAS with XWAS implementing different statistical models to consider XCI patterns for dosage compensation while controlling for multiple covariates and the interactions with those covariates, (C) sex-differential tests to evaluate the difference in effect size, trait heritability etc., (D) high level analysis enabling sex-specific gene-based tests, GWAS and XWAS meta-analyses, and generation of both standard and sex-aware polygenic risk scores and (E) several utility functions for the application of the best practices for additional quality control of genetic data necessary for the sex-aware tests.
Implementation: GXwasR is written in R, and is fully open-source with a comprehensive tutorial with simulated data and several use-case scenarios using real genotype datasets.
Impact: With this work, our goal is to make sex-aware genomic analysis more accessible to researchers by providing a theoretical background and a practical framework to perform sex-aware analysis of genotype and phenotype.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4190 Haptoglobin genotype calling and allele specific expression in AD

Authors:

E. Vazquez-Hidalgo, H. Bai, W. Bush; Case Western Reserve Univ., Cleveland, OH

Alzheimer's disease (AD) is a complex neurodegenerative disorder with multifactorial etiology. We explore the combined effects of apolipoprotein E (*APOE*) and haptoglobin (*HP*) in AD. In most cases, the greatest genetic risk factor for AD is *APOE*. *APOE* has three alleles: *APOEε2*, *APOEε3*, and *APOEε4*. Relative to *APOEε3*, *APOEε2* protects against AD but *APOEε4* increases AD risk. AD susceptibility is further modulated by dosage effect. Homozygous individuals for *APOEε4* are at higher risk for AD but those that are homozygous for *APOEε2* have greater protection against AD. The *APOE* protein, apolipoprotein E (APOE), potentially plays a role in amyloid-β (Aβ) protein removal. Increased Aβ plaques and *APOEε4* allele dosage are associated with AD. APOE's ability to clear Aβ depends on APOE's oxidation state. Oxidized APOE decreases its ability to remove Aβ. Recently, we have shown that *APOE* is modulated by haptoglobin (HP) encoded by *HP*. HP is a glycoprotein that possesses antioxidant properties. Experiments revealed that HP binds to APOE, likely protecting APOE from oxidation. Furthermore, HP promotes APOE mediated Aβ clearing. Yet, the impact of *HP* genotype on AD remains unknown. *HP* has two alleles, *HP1* and *HP2*, and consists of five or seven exons, respectively. As a result, *HP* has three common genotypes: *HP1-1*, *HP1-2*, and *HP2-2*. In European ancestry populations, *HP1-2* is the largest genotype class and exhibits considerable impact on AD risk. Expression of *HP* copy number variation (CNV) adds an additional layer of complexity to HP-APOE interaction. Therefore, expression imbalance potentially contributes to AD phenotypic variation and disease pathophysiology. Understanding the interplay between *HP* CNV and *HP* allele specific expression (ASE) in AD could provide some insight into the underlying genetic interactions that regulate AD.

Here we evaluate the impact of *HP* CNV and measure ASE to identify allelic imbalance. To determine the copy number of the *HP* CNV and other structural polymorphisms, we aligned WGS reads to custom *HP* references for deletion states that reflect *HP1* and *HP2* states, allowing identification of reads spanning intron junctions that uniquely identify *HP2* carriers. Allelic copy numbers were determined based on bi-allelic copy number for each sequence boundary. ASE was measured using RNA-seq to evaluate allele-specific read counts to exploit differential expression of *HP* isoforms in heterozygotes. We show that risk of developing AD is not only linked to *APOE* genotype, but may also be influenced by level and type of *HP*. Knowing transcriptional expression of specific alleles can help us better understand how genetic variants affect phenotypes

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4191 Haptools: a toolkit for admixture and haplotype analysis.

Authors:

A. Massarat[1], M. Lamkin[1], C. Reeve[1], A. Williams[2], M. D'Antonio[1], M. Gymrek[1]; [1]Univ. of California San Diego, La Jolla, CA, [2]23andMe, Inc., Sunnyvale, CA

Genome-wide association studies (GWAS) have promised to revolutionize our understanding of the genetic components of many diseases and facilitate personalized medical care. Yet researchers still struggle to compute disease risk equitably. GWAS predictions are often limited to the study population and perform poorly when applied to individuals of mixed descent.

In addition to variant-level analyses, phenotypic effects may also be mediated by haplotypes (combinations of variants that are inherited together) or by the local ancestry background on which a variant falls. Thus, leveraging local ancestry and haplotype information in GWAS and downstream analyses may improve the utility of genomics for individuals from diverse and recently admixed ancestries. We and others have begun to develop methods to improve risk prediction in admixed individuals by inferring haplotypes and local ancestry, but an absence of ground truth data makes it difficult to benchmark these tools. Most existing simulation, visualization and variant analysis frameworks are based on variant-level analysis and do not automatically handle these features.

We present haptools, an open-source toolkit for performing local ancestry aware and haplotype-based analysis of complex traits. Haptools supports fast simulation of admixed genomes, visualization of admixture tracks, and simulation of haplotype- and local ancestry-specific phenotype effects. Since its original publication, haptools has been expanded to simulate traits deriving from complex variants such as tandem repeats. The haptools python library powers its ability to quickly load an exceptionally wide variety of file formats into standardized data structures, including genotypes from the new PLINK2 PGEN format or from VCFs generated by five different tandem repeat callers. We also developed a new file format for haplotypes (.hap) which, compared to other formats, is more flexible but also easier to parse and faster to query.

We demonstrate the utility of our toolkit by benchmarking it against others in our field, such as admix-simu and GCTA. Using haptools, we can illustrate the decrease in power that we expect for detecting haplotype and ancestry-dependent causal effects via a GWAS. Since the release of our publication, we've also demonstrated this phenomenon in fine-mapping methods.

Overall, haptools provides a valuable set of utilities for developing and benchmarking methods for ancestry-aware analysis of complex traits by simulating genotypes and phenotypes for GWAS and polygenic risk scores. We intend for it to help improve our ability to predict disease risk for individual patients based on their genotypes.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4192 Hardy-Weinberg equilibrium test accounting for population structure and genetic relatedness in large-scale whole genome sequencing studies.

Authors:

D. Shyr, R. Dey, X. Lin; Harvard T.H. Chan Sch. of Publ. Hlth., Boston, MA

Large-scale cohort studies such as the Center for Common Disease Genomics have integrated deep whole-genome sequencing and other omics data with clinical data. Assessing the Hardy-Weinberg Equilibrium (HWE) assumption appropriately for these datasets is integral to quality control procedures; however, this is challenging due to ancestral heterogeneity and genetic relatedness of the samples. Currently, there are no existing methods that incorporate both population structure and genetic relatedness when testing for HWE. We propose a novel HWE test using the generalized estimating equation that accounts for population structure with principal components and the relationship among samples with a family-specific genotype correlation matrix. Our results demonstrate that ignoring population structure and relatedness when evaluating HWE inflates the false positive rates drastically. Compared to other methods, our approach controls for type-I error the best while maintaining high power. Our implementation is scalable and practical such that HWE tests can be performed efficiently across millions of markers and over a hundred thousand samples.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4193 † Harnessing the AI revolution: Application of Large Language Models (LLMs) for genotype refinement and imputation

Authors:

D. Pandey[1], R. Dhar[2], V. Narasimhan[1]; [1]The Univ. of Texas at Austin, Austin, TX, [2]Indian Inst. of Technology Kharagpur, Kharagpur, India

A critical technical step in the processing of next-generation sequencing data for medical or population genetics is genotype refinement or imputation, which is the process of predicting unobserved genotypes or correcting spurious genotypes in a newly sequenced individual using a reference panel of haplotypes to get the most accurate genetic sequence information for each person. Existing imputation methods are based on Hidden Markov Models (HMMs) which are a slow statistical process that must evaluate how a new sequence matches each one of the millions of references and suffers from limitations related to accuracy, computational resources, and scalability. In the field of natural language processing, the introduction of LLMs trained on extremely large datasets for similar problems of word auto correction or sentence completion have rendered HMMs largely obsolete.

Here we examined the use of LLMs in genotype refinement/imputation by training two transformer-based language models from scratch for a similar task of predicting the 101st variant given 100 preceding variants as input. We built a benchmarking dataset comprised of genotype data from chromosome 20 of the 1000 Genomes Project Phase3, and the UKBiobank whole genome sequencing dataset and evaluated the performance of our models by examining the accuracy of our model at predicting variants at different minor allele frequencies (MAF) bins. The first model, a bigram language model, had 0.205 million trainable parameters. It was trained for 500 epochs, resulting in a decrease in cross entropy loss from 0.482 to 0.071 on the validation set. The second model, GPT2, was trained from scratch along with the GPT2 Tokenizer on the same data. It had 124.4 million trainable parameters and was trained for 160,000 steps, resulting in a decrease in cross entropy loss from 2.3066 to 0.995 on the validation set. We observed that the predictive performance of both models improved as the predicted variant transitioned from rare to common within the MAF bins. Using our benchmarking data, we also examined the impact of different tokenization schemes, training dataset size, and model architecture on the model accuracy.

This study serves as a proof of principle for the application of LLMs in various population genetic tasks and will lay the foundation to develop architectures designed not for language but specifically for genomic sequence information.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4194 Harnessing transcriptomic signals for amyotrophic lateral sclerosis to identify novel drugs and enhance risk prediction

Authors:

O. Pain[1], A. Jones[1], A. Al Khleifat[1], D. Agarwal[2], D. Hramyka[3], H. Karoui[4], J. Kubica[5], D. J. Llewellyn[6], J. M. Ranson[6], A. Iacoangeli[7], A. Al-Chalabi[7]; [1]King's Coll. London, London, United Kingdom, [2]Wellcome Ctr. for Human Genetics, Oxford, United Kingdom, [3]Charité – Univ.smedizin Berlin, Berlin, Germany, [4]Imperial Coll. London, London, United Kingdom, [5]Univ. of Warsaw, Warsaw, Poland, [6]Univ. of Exeter Med. Sch., Exeter, United Kingdom, [7]King's Coll. London, London, United Kingdom

Introduction: Amyotrophic lateral sclerosis (ALS) is a fatal neurodegenerative disease. This study integrates the latest ALS genome-wide association study (GWAS) summary statistics with functional genomic annotations with the aim of providing mechanistic insights into ALS risk loci, inferring drug repurposing opportunities, and enhancing prediction of ALS risk and clinical characteristics.
Methods: Genes associated with ALS were identified using GWAS summary statistic methodology including SuSiE SNP-based fine-mapping, and transcriptome- and proteome-wide association study (TWAS/PWAS) analyses. Using several approaches, gene associations were integrated with the DrugTargetor drug-gene interaction database to identify drugs that could be repurposed for the treatment of ALS. Furthermore, ALS gene associations from TWAS were combined with observed blood expression in two external ALS case-control datasets to calculate polytranscriptomic scores and evaluate their utility for prediction of ALS risk and clinical characteristics, including site of onset, age at onset, and survival.
Results: SNP-based fine-mapping, TWAS and PWAS identified 117 genes associated with ALS, with TWAS and PWAS providing novel mechanistic insights. Drug repurposing analyses identified five drugs significantly enriched for interactions with ALS associated genes, with directional analyses highlighting α-glucosidase inhibitors may exacerbate ALS pathology. Additionally, drug class enrichment analysis showed calcium channel blockers may reduce ALS risk. Across the two observed expression target samples, ALS polytranscriptomic scores significantly predicted ALS risk (R^2 = 4%; p-value = 2.1×10^{-21}).
Conclusions: Functionally-informed analyses of ALS GWAS summary statistics identified novel mechanistic insights into ALS aetiology, highlighted several therapeutic research avenues, and enabled statistically significant prediction of ALS risk.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4195 High-dimensional causal inference from GWAS summary statistic data

Authors:

N. Machnik, S. Mahmoudi, M. R. Robinson; Inst. of Sci. and Technology Austria, Klosterneuburg, Austria

Inferring causal relationships using high-dimensional observational data is a crucial first step for understanding the ethology of common complex disease. Mendelian Randomization (MR) is currently the most commonly used approach to investigate causal relationships in data where genomic information is linked to electronic health records, but relies upon strong assumptions that are often not fulfilled. In this study, we introduce Casual Inference GWAS (CI-GWAS), a novel approach that leverages graphical models to perform causal inference using GWAS summary statistic data. CI-GWAS consists of three main steps: (i) a GPU-based algorithm for variable selection, called cuda-skeleton, to learn a sparse graph representing all causal SNP-trait and trait-trait relationships within a single-step; (ii) a modified sRFCI algorithm to infer causal directions in the inferred sparse graph, accounting for potential hidden variables and confounders; and (iii) a data-driven sDAVS algorithm to estimate the magnitude of causal effects across the learnt graph structure. CI-GWAS estimates causal SNP-trait and trait-trait relationships for 17 phenotypes and 2.2 million SNP markers in less than a day in over 400,000 UK Biobank individuals. By inferring causal relationship among 17 traits jointly we find, contrary to MR, that almost all phenotypic dependencies can be attributed to a common underlying latent cause, with no evidence for a causal relationship. The exception is a clear causal relationship between higher alcohol consumption and cardiovascular disease. Our model also provides a likely causal set of fine-mapped SNP-trait associations and reveals a lack of direct pleiotropy among all traits. Thus, we demonstrate the unique biological insight that can be obtained from high-dimensional graphical inference in large-scale summary statistic data.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4196 *HORNET* identifies a complex gene network underlying Alzheimer's disease risk in diverse populations

Authors:

N. Lorincz-Comi[1], Y. Yang[1], X. Zhu[2]; [1]Case Western Reserve Univ., Cleveland, OH, [2]Case Western Reserve Univ., Shaker Heights, OH

Background
The ABCA7 gene is highly associated with both early- and late-onset Alzheimer's disease (AD) in diverse populations. In vivo models suggest that ABCA7 is involved in the clearance of amyloid beta plaques, a hallmark feature of AD pathology, suggesting it has a causal role in modifying AD risk. However, the 500Kb window around ABCA7 contains 25 other genes, making evaluation of its causal role challenging because of complex regulatory mechanisms. Here, we searched the 1Mb region around ABCA7 to identify genes with direct causal effects on AD risk after adjusting for inter-relationships between expression patterns of surrounding genes.
Methods
We developed the HORNET software to estimate direct causal effects of gene expression using multivariable Mendelian Randomization (MR) and GWAS summary statistics. We applied HORNET to gene expression data in blood from European (EUR; n=31k) and African American (AA; n=757) populations separately. AD GWAS data were from Jansen et al. (n=455k). Population-specific regulatory networks were constructed using graphical lasso.
Results
25 genes in a 1Mb window around ABCA7 had cis-eQTLs (P<5E-8) in EUR and 23 in AA. The regulatory network of gene expressions was marginally denser in AA than EUR (P=0.051). For example, each gene on average either regulates or is regulated by 10 other genes in EUR and 16 in AA. The expression of these genes explained 95.7% of the local heritability of AD in EUR and 73.5% in AA, suggesting well-specified causal models and the appropriate tissue.
The local genetic correlation between AD risk and ABCA7 expression was 0.93 in EUR and 0.90 in AA and its total unmediated causal effect on AD was large (AA odds ratio=1.33, EUR odds ratio=2.40; both P<1E-100). However, after considering neighboring genes, ABCA7 had no direct causal effect on AD risk in EUR (P=0.24) or AA (P=1.00). Our regulatory network suggests that ABCA7 expression indirectly confers AD risk by regulating nearby genes such as KISS1R and ARID3A (both P<0.05 in EUR and AA). Genetic correlations between ABCA7 and these two genes ranged in absolute value from 0.71-0.76 in EUR and 0.48-0.96 AA, suggesting strong regulatory dependencies.
Conclusions
This was the first application of HORNET, which has statistical and computational advantages over existing alternatives that will be discussed. Using ABCA7, we demonstrated that evaluation of a gene for causality must consider surrounding genes in the target gene's network. We could only identify genes with direct causal effects on AD risk by considering the broader regulatory context. We plan to validate these findings using data from African, Asian, and European individuals in the UK Biobank.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4197 Human age- and sex-specific molecular phenomena inferred by leveraging public transcriptome data with machine learning.

Authors:

K. Johnson, A. Krishnan; Univ. of Colorado Anschutz Med. Campus, Aurora, CO

Age and sex are historically understudied factors in biomedical studies even though many complex traits and diseases vary by these factors in their incidence and presentation. As a result, there are massive gaps in our understanding of genes and molecular mechanisms that underlie sex- and age-associated physiology and disease. Hundreds of thousands of publicly-available human transcriptomes capturing gene expression profiles of tissues across the body and subject to various biomedical and clinical factors present an invaluable, yet untapped, opportunity for bridging these gaps. Here, we present a computational framework that leverages these data to infer genome-wide molecular signatures specific to sex and age groups. As the vast majority of these expression profiles lack age and sex labels, the core idea of our framework is to use the measured expression data to predict missing age/sex metadata and derive the signatures from the predictive models. We first manually curated ~30,000 primary samples associated with age and sex information and profiled using microarray and RNA-seq. To the best of our knowledge, this dataset is the largest collection of expression data labeled with age and sex. We used this dataset to infer sex-biased genes within each of eleven age groups along the human lifespan. We also trained machine learning (ML) models to predict these age groups from gene expression values separately within females and males. Dataset-level cross validation shows that these ML classifiers are able to discriminate between age groups in a biologically meaningful way in each sex across technologies. Further, these predictive models capture sex-stratified age-group ‘gene signatures’, i.e., the strength and the direction of importance of over 18,000 genes across the genome, specific to each age group in each sex. Enrichment analysis of these gene signatures with prior gene annotations helped in identifying relevant age- and sex-associated multi-tissue and pan-body molecular phenomena (e.g., general immune response, inflammation, metabolism, hormone response). We have made our results — thousands of gene, pathway, trait, and disease signatures — and curated expression data, the largest public dataset of its kind, available for reanalysis or application to a broad range of disorders with age or sex biases. A Shiny app we developed makes these results and data easy to query and visualize for the community to use for hypothesis generation and novel biological discovery. Overall, we have presented a path for effectively leveraging massive public omics data collections to broadly investigate the genomic basis of age- and sex-differences in physiology and disease.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4198 IBDMap: leveraging IBD mapping for validating and uncovering binary trait associations at biobank scale

Authors:

G. Evans[1], J. Baker[1], A. Petty[2], L. Petty[1], H-H. Chen[1], R. Bohlender[3], C. Huff[3], J. Below[2]; [1]Vanderbilt Univ., Nashville, TN, [2]Vanderbilt Univ Med Ctr., Nashville, TN, [3]Univ. of Texas MD Anderson Cancer Ctr., Houston, TX

The recent increase in available biobank genetic datasets has created numerous opportunities for novel disease gene discovery. Identity by descent (IBD) mapping is an established approach for gene discovery that evaluates segment sharing among cases and controls to identify genomic regions harboring rare, disease-causing variation undetectable through traditional genome wide association studies. However, IBD mapping in biobank scale data is a substantial computational challenge. To this end, we developed IBDMap, a multithreaded, scalable C++ tool which employs a map/reduce parallelized permutation strategy to determine IBD sharing enrichment by comparing genome wide sharing rates among cases and controls. To demonstrate feasibility and viability for end users employing data inputs of all sizes, we developed a robust, phenome-wide analytical pipeline that (1) extracts biobank scale phenotype data, (2) executes IBDMap in reasonable runtime with a highly parallelized approach, and (3) maps known associations from both OMIM and the NHGRI-EBI GWAS Catalog to genome-wide significant breakpoint regions in the IBDMap results. We used iLASH to identify consensus IBD segments for 67,128 individuals of European descent present in Vanderbilt University's biobank, BioVU who had been previously genotyped on the Multi-Ethnic Genotyping Array (MEGAEX).
We then applied IBDMap and IBDReduce across 1,641 unique categorical phenotypes (delineated by PheWAS Catalog phecodes) across our entire cohort. Our phenome-wide analysis identified genome-wide significance in over 100 unique signal locations spanning over 50 unique phenotypes. As an example, we highlight three associations related to phenotypes affecting height, including (1) congenital spine deformities (phecode 754), overlapping the gene POLR3GL where locus 1q21.1 has been previously connected to short stature, (2) malnutrition (phecode 260.3), overlapping gene FAM110C and variant rs300695 implicated in affecting height, and (3) spinal enthesopathy (phecode 715.3), matching multiple genes also implicated in affecting height. We demonstrate the utility of applying established IBD mapping statistical method(s) at scale in a computationally tractable infrastructure and timeframe. We are continuing to refine our analytical pipeline and aim to uncover more detailed associations - both novel and previously validated - across all significant signals from our phenome-wide analysis.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4199 Identification of genetic effects on seasonal variation in disease risk based on temporal patterns from 1,952 diseases in over 7 million individuals.

Authors:

S. Rognvaldsson[1], P. Vartiainen[1,2], E. Viippola[1], M. Pirinen[1,3,4], S. Jones[1], H. M. Ollila[1,5,6,7], A. Ganna[1,8]; [1]Inst. for Molecular Med. Finland (FIMM), Helsinki Inst. of Life Sci., Univ. of Helsinki, Helsinki, Finland, [2]Pediatric Res. Ctr., Helsinki Univ. Hosp. and Univ. of Helsinki, Helsinki, Finland, [3]Dept. of Publ. Hlth., Faculty of Med., Univ. of Helsinki, Helsinki, Finland, [4]Dept. of Mathematics and Statistics, Univ. of Helsinki, Helsinki, Finland, [5]Broad Inst. of MIT and Harvard, Cambridge, MA, [6]Ctr. for Genomic Med., Massachusetts Gen. Hosp., Boston, MA, [7]Anesthesia, Critical Care, and Pain Med., Massachusetts Gen. Hosp. and Harvard Med. Sch., Boston, MA, [8]Analytic and Translational Genetics Unit, Massachusetts Gen. Hosp., Boston, MA

Seasonal variation has affected human societies throughout history, shaping various aspects of life including agriculture, migration patterns and culture. This influence is observed in the occurrences of diseases such as viral and bacterial infections, cardiovascular disease and mental disorders. While there are a multitude of factors influencing the timing of disease diagnoses, genetic factors have not been extensively explored. The aim of this project was to elucidate genetic variation that contributes to seasonal disease risk. First, we derived a statistical approach to study seasonality patterns for 1,952 disease endpoints in the entire Finnish population (n=7,166,416), which is subjected to marked seasonal changes due to Finland's location in the northern hemisphere. A significant seasonality pattern was observed in 54% of the diseases with infectious, mental health and respiratory disease categories showing the most extreme seasonality. A total of 69 diseases with significant seasonal disease patterns were selected for genetic analyses in the FinnGen study population (n=473,897). For each disease we tested both a dichotomous phenotype representing whether a person was diagnosed in a period of high/low disease prevalence based on the nationwide seasonality pattern and a continuous phenotype representing the magnitude of the seasonality pattern at the diagnosis time. While we discovered that the majority of variants that associate with diseases do not show seasonal association, we identified 14 genome-wide significant loci associating with seasonality, including a top-sQTL, rs41273830[T], in *ITGB8* for major depression and a stop-gain variant, rs601338[A], in *FUT2* for intestinal infections. Both the *ITGB8* and *FUT2* variants were associated with increased disease risk in periods of low prevalence. Further replication studies are on-going in collaboration with other nordic countries to assess how robust seasonality affecting variants are across populations. Our approach introduces a new aspect to genetic research and enables identifying new disease variants whose effects are currently obscured by seasonal variation.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4200 Identification of optimal values, or sweet spots, for 69 blood metabolites, and genetic loci associated with deviation from optima.

Authors:

O. Vishnyakova[1,2], J. Min[2], X. Song[3,2], K. Rockwood[4], L. Elliott[2], A. Brooks-Wilson[1,2]; [1]Genome Sci. Ctr., BC Cancer, Vancouver, BC, Canada, [2]Simon Fraser Univ., Burnaby, BC, Canada, [3]Surrey Mem. Hosp., Fraser Hlth.Authority, Surrey, BC, Canada, [4]Div. of Geriatric Med., Dalhousie Univ., Halifax, NS, Canada

Homeostasis allows the maintenance of stable physiological conditions and body chemistry. Previous observations on a group of exceptionally healthy 'Super Seniors' show a lower variance of multiple physiological measures, including telomere length and hematology measures, compared to a less healthy group matched for age, indicating that healthier individuals have values closer to an optimal 'sweet spot'. We tested the hypothesis that lower variance in healthier people is an indicator of health-relevant phenotypes using metabolomics data from a larger community sample, the Canadian Longitudinal Study on Aging (CLSA) Comprehensive Cohort. The findings validated the proposed approach. We restricted analysis to the largest ethnic group (European). After sample and data quality control, data used included measures of 989 metabolites measured in a subset of 8,764 CLSA participants aged 45 to 85 years. Health deficit was defined through five instruments, including a deficit accumulation frailty index (FI), a count of major diseases, a count of all disorders, physical and cognitive functioning.

Our two-step sweet spot discovery pipeline revealed heteroskedasticity with lower variance for the healthiest adults for 151 (15%) metabolites. We identified optimal values for 69 of them by testing for the changes of the effect on the health deficit using segmented regression.

We examined the sweet spots in two ways. First, we constructed a per-participant representation of homeostatic dysregulation by summing the absolute difference of each metabolite from its corresponding sweet spot, across 69 metabolites. This cumulative deviation from sweet spot (CDSS) biomarker was strongly associated with FI on a test set (p-value=$3.9x10^{-21}$). Survival analysis using Cox regression model with age, sex and CDSS as predictors revealed that CDSS is associated with increased risk of mortality (HR=1.48; 95%CI [1.38;1.59]).

Second, we conducted genome-wide association analyses on the absolute difference of the metabolite measures to their corresponding sweet spots, for each of the 69 metabolites with sweet spots. We identified 20 distinct {metabolite, sex, genetic variant} combinations that were not present in a GWAS of metabolite concentrations in CLSA. Findings of biological interest include {1-oleoyl-GPC (18:1), male, rs148368178 (associated with *CERK*, which is involved in calcium homeostasis)} and {lithocholate sulfate, female, rs4149056 (*SLCO1B1*, associated with statin-induced myopathy)}. This work identifies a new paradigm for identification of biomarkers relevant to health, using analysis of variation between healthier and less healthy groups.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4201 Identification of potential novel genes associated with clonal hematopoiesis using exome sequencing data

Authors:

Y. Han, G. McInnes, M. Luth, D. Jakubosky, B. Cajes, S. Bruse, O. Gottesman; Empirico Inc., San Diego, CA

Clonal hematopoiesis of indeterminate potential (CHIP) is a state characterized by the presence of somatic mutations in blood cells and has been associated with the pathogenesis of hematological malignancies as well as risk for several non-malignant diseases. Large-scale biobanks comprising sequencing data linked to rich phenotypic information, such as the UK Biobank, have enabled a greater understanding of the role of CHIP in disease. A greater understanding of the driver genes associated with the development of CHIP could enable targeted screening and early intervention strategies to manage potential risks, and may lead to the discovery of potential novel therapeutic targets for CHIP-associated diseases. While several driver genes such as DNMT3A, TET2, ASXL1, JAK2, and SRSF2 have been well-studied, much remains to be discovered. In this study, we conducted a comprehensive investigation of the phenotypic associations of genetic variants in known CHIP driver genes and sought to identify novel CHIP driver genes. We first performed phenome-wide association studies for burdens of genetic variants in known CHIP driver genes using exome sequencing and phenotypic data from the UK Biobank. We then employed a novel genome-wide clustering method to discern patterns of associations across these genes, and identified consistent clusters of associations across a range of phenotypes including hematological phenotypes, inflammation, and cancer, in particular hematological malignancies. Building upon these findings, we employed our clustering method to identify new CHIP driver genes that exhibit similar patterns of associations to known CHIP driver genes. We identified several novel genes with clusters of associations that match those observed in known driver genes and with plausible biological mechanisms for influencing CHIP. This study demonstrates the utility of large-scale biobanks with exome sequencing data paired with phenotypic information for both the identification of CHIP and expanding our understanding of the role of CHIP in disease. Further, this study highlights the potential of our pattern-matching approach for discovering potential new CHIP driver genes that may contribute to early risk detection and the development of intervention strategies for hematological malignancies and other CHIP-associated diseases. This research has been conducted using the UK Biobank Resource under Application Number 34229.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4202 Identifying associations of rare noncoding variants with autism through integration of gene expression, sequence and sex information

Authors:

R. Li, J. Ernst; Univ. of California, Los Angeles, Los Angeles, CA

The growth of whole-genome sequencing (WGS) data has facilitated genome-wide identification of rare noncoding variants. However, elucidating these variants' associations with complex diseases remains challenging.
To better understand these associations we first revisit a previous report of significant brain-related noncoding association signals of autism spectrum disorder (ASD) detected from *de novo* variants in the Simons Simplex Collection (SSC) WGS data when using a deep-learning-based pathogenicity score. We first demonstrate that local GC content is sufficient to capture association signals in variants near brain-expressed genes similar to those previously reported based on deep learning. Additionally, we show that this local GC content signal is specific to male probands with a female sibling (minimum p-value in brain tissue=$1.3x10^{-4}$) compared to male probands with a male sibling (minimum p-value in brain tissue=0.31). We further show among male probands and female siblings the signal is specific to variants upstream of their assigned TSS (minimum p-value in brain tissue=$3.3x10^{-6}$ for upstream variants vs. 0.057 for downstream variants).
Based on these findings, we developed an approach, k-mer-based gene expression neighborhood test (KGNT), to more systematically consider gene expression and sequence information for testing for association signals. KGNT first organizes variants into "neighborhoods" based on their assigned genes and pairwise gene expression correlations determined based on a large compendium of expression data. Then for each neighborhood KGNT tests for associations between the variants' k-mer distributions and phenotype. Applying KGNT to the SSC *de novo* variants upstream of TSS in male proband-female sibling families, we identified ASD association signals with 6-mers ($p=1.7x10^{-9}$ for the top neighborhood), surpassing the association using GC content ($p=1.2x10^{-7}$ for the top neighborhood). From the top neighborhoods, we extracted specific 6-mers driving the proband-sibling difference beyond GC content. Genes associated with top neighborhoods showed the strongest enrichment for synapse and some other brain-related gene ontology terms. In addition, we examined the chromatin state assignments of the variants in the top neighborhoods and observed sibling variants preferentially in quiescent states.
In summary, we identified using local GC content an ASD association signal from *de novo* noncoding variants in male probands with female siblings upstream of brain-expressed genes, which we were able to further refine and enhance with KGNT.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4203 Identifying common genetic susceptibility underlying co-morbid phenotypes using binomial regression.

Authors:

P. Panja[1,2], S. Bhattacharjee[2]; [1]Regional Ctr. for Biotechnology, Faridabad, India, [2]Natl. Inst. of BioMed. Genomics, Kalyani, India

The models suggested in multiphen and bamp provide an alternative to study population based genetic association with multivariate phenotypes by exploring the dependence of genotype on phenotype instead of the naturally arising dependence of phenotype on genotype. However, these approaches test the null hypothesis of no association with any of the constituent traits versus the alternative hypothesis of association with at least one of the constituent traits of the multivariate phenotype vector. Thus, such tests do not provide evidence of pleiotropy or common genetic factors underlying all the traits constituting the multivariate phenotype, which might be correlated. With respect to a pair of co-morbid phenotypes (both binary, a combination of binary and quantitative or both quantitative), we aim to modify the proposed bamp (binomial regression based association of multivariate phenotypes) approach to test the null hypothesis of no association with at least one of the phenotypes versus the alternative hypothesis with both the phenotypes. Identifying the pleiotropy of two diseases is important to decipher the underlying molecular events of two correlated phenotypes. Likelihood ratio test (lrt) will be used for this purpose. Proposed method can also differentiate between mediated pleiotropy and horizontal pleiotropy. Extensive simulations would be done for different values of correlation coefficients between the two comorbid traits to identify the power of the test and to compare the power of the same model which has used multiple testing correction instead of considering a multivariate phenotype vector. p { line-height: 115%; margin-bottom: 0.25cm; background: transparent }p { line-height: 115%; margin-bottom: 0.25cm; background: transparent }

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4204 Identifying demographic, socioeconomic, behavioral, and environmental factors contributing to disparities in the prevalence of sepsis

Authors:

A. Biji[1,2], M. Stapylton[2], S. Cullina[1,2], G. M. Belbin[3], E. E. Kenny[2,4,5], S. Asgari[5,2,4]; [1]Graduate Sch. of BioMed. Sci., Icahn Sch. of Med. at Mount Sinai, New York, NY, [2]Inst. for Genomic Hlth., Icahn Sch. of Med. at Mount Sinai, New York, NY, [3]Gencove Inc., New York, NY, [4]Dept. of Med., Icahn Sch. of Med. at Mount Sinai, New York, NY, [5]Dept. of Genetics and Genomics, Icahn Sch. of Med. at Mount Sinai, New York, NY

Background: Sepsis is the body's extreme reaction to an infection. It is the second cause of in-hospital deaths and the most expensive inpatient medical condition in the US, with the highest incidence among African American/Africans (AA) and Hispanic/Latin Americans (HL). Thus, identifying potential factors that cause racial/ethnic disparities in sepsis prevalence is a public health priority. **Methods**: We investigated the incidence of sepsis in the *BioMe* biobank (N = 50292, ages: 18-105, median (m) = 52.32). Sepsis cases (N = 871 including 245 AA and 353 HL) were assigned any of the following ICD10 codes: A40, A41, R65.20, and R65.21, and controls (N = 49421) did not have any of these codes. We used a logistic model with sepsis as the outcome and sociodemographic covariates as predictors after accounting for age, sex, and 10 genotyping principal components. **Results**: Compared to self-reported European Americans (EA), we found a significantly higher sepsis burden in self-reported AA (odds ratio (OR) = 2.45 [95% CI 2.02-2.98], p-value (p) = 4.9e-19) and HL (1.98 [1.65-2.38], p = 2.7e-13). Self-reported ancestry groups are too broad and do not capture differences in sub-populations. So, we assigned population groups based on recent, shared genetic ancestry (identity-by-decent (IBD), Belbin G et al, Cell 2019) and observed a similar trend of significantly higher sepsis burden in AA (1.98 [1.58-2.49]) and HL (1.9 [1.53-2.36]) IBD communities than in Europeans. Puerto Ricans (OR 2.13) and Dominicans (OR 1.63) showed different sepsis burdens, a distinction that is missed when only using self-reported ancestry. We tested 55 sociodemographic factors, 24 of which were significantly ($p < 9.1e-4$) associated with sepsis, with kidney disease, age, and diabetes being the top three risk factors. We observed that the median age of sepsis diagnosis was significantly lower in AA (m = 58.6, p = 1.4e-05) and HL (m = 57.8, p = 1.4e-4) compared to EA (m = 64). Within age groups, sepsis risk was significantly higher in younger (< 50) AA (2.95 [1.92-4.53]) and HL (3.04 [2.04-4.54]) compared to same-age EA. This is notable because while sepsis in older adults may not be preventable due to age-related comorbidities, it can be preventable in younger patients. We are currently exploring the relative contribution of all significantly associated sociodemographic factors to racial/ethnic disparities in sepsis risk. **Conclusion**: By leveraging biobanks we can replicate known epidemiological trends of racial/ethnic disparities in sepsis and explore sociodemographic risk factors in diverse populations, enabling the development of targeted preventive measures for sepsis.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4205 Identifying risk genes from copy number variations by statistical fine mapping

Authors:

B. Jiang[1], M. Qiao[1], P. Carbonetto[1], K. Tayeb[1], W-P. Lee[2], X. He[1], G. Wang[3]; [1]Univ. of Chicago, Chicago, IL, [2]Univ. of Pennsylvania, Philadelphia, PA, [3]Columbia Univ., Fort Lee, NJ

Copy number variations (CNVs) are important sources of genetic variations. Many CNVs have been associated with a number of human diseases and phenotypes. CNVs, however, often span multiple genes, making it difficult to identify causal genes underlying the associations. Existing gene association analysis methods cannot distinguish between casual genes and nearby non-causal genes that are similarly disrupted by CNVs. In this study, we developed a statistical method, borrowing ideas from statistical fine-mapping in genome-wide association analysis (GWAS), to identify causal genes from CNVs. Our method, CNV-Mapper, is based on a regression model that jointly analyzes all genes in a region and infer the effects of CNV disruption of these genes on the phenotype. By using a sparse prior, CNV-Mapper encourages the selection of one or few causal genes that drive the CNV association, and reports Posterior Inclusion Probabilities (PIPs), which are probabilities that any genes are causal given the data. Thus CNV-Mapper not only reports potential causal genes, but quantifies the uncertainty of these predictions. CNV-Mapper is able to handle both quantitative traits, and case-control data. To evaluate the performance of CNV-Mapper, we conducted realistic simulations by using parameters learned from real data and real breakpoints of CNVs. Our findings demonstrate that PIPs from CNV-Mapper are calibrated, meaning it effectively controls the False Positive rates at the specified levels. Compared with existing methods, CNV-Mapper is significantly better at classifying risk genes and non-risk genes. We are applying CNV-Mapper to data obtained from the Alzheimer's Disease Sequencing Project (ADSP), and will release CNV-Mapper as an R package. In summary, CNV-Mapper provides a much-needed statistical tool that translates CNV associations into knowledge of putative risk genes underlying human diseases.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4207 Impact of GWAS meta-analysis heterogeneity on polygenic prediction accuracy.

Authors:

Y. Wang, C-T. Liu; Boston Univ. Sch. of Publ. Hlth., Boston, MA

Introduction. A polygenic risk score (PRS) represents an individual's genetic predisposition for a particular trait or disease based on the cumulative effects of multiple genetic variants. Various PRS methods have leveraged genome-wide association study (GWAS) summary statistics that often derived from GWAS meta-analyses in a large-scale consortium to maximize statistical power. However, the impact of heterogeneity in effect sizes among cohorts on PRS predictive accuracy remains unclear. **Methods.** We conducted simulations to investigate the implication of heterogeneity on PRS performance. We simulated 20 independent cohorts (N=10,000 each) for each ancestry (AFR, EUR), using haplotypes from the 1000 Genomes Project Phase 3 reference panel. Heterogeneity was introduced for true causal effect size by varying genetic correlation levels across cohorts (rg = 1, 0.9, 0.6, 0.3). We performed GWAS for each cohort and meta-analyzed ten distinct cohorts with various proportions of AFR cohorts. We applied clumping and thresholding (C+T) to the meta-GWAS summary statistics and evaluated the predictive performance with the remaining AFR samples. We next re-constructed the PRS by excluding heterogeneous ($I^2 > 75$) variants. **Results.** The simulation study illustrated that the number of heterogeneous variants increased as genetic correlation decreased across cohorts. Generally, the PRS was more predictive with higher proportions of target ancestry-matched cohorts and larger cross-cohort genetic correlations. When rg=1, excluding heterogeneous variants substantially improved prediction accuracy for PRS built from multi-ancestry meta-GWAS, but not for single-ancestry meta-GWAS. However, when rg < 1, removing the heterogeneous variants reduced prediction accuracy for both single and multi-ancestry meta-GWAS. We will conduct real data analysis with the UK Biobank data and the summary statistics (without_UKB) from the Global Lipids Genetic Consortium (GLGC) meta-GWAS results. **Summary.** In conclusion, our study provides insights into the impact of between-study heterogeneity on polygenic prediction and highlights the potential to enhance prediction accuracy by incorporating heterogeneity measurements. This underlies the importance of sharing heterogeneity measurements with GWAS meta-analysis summary statistics for large-scale consortia.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4208 Improved Estimation of Functional Enrichment in SNP Heritability Using Feasible Generalized Least Squares

Authors:

Z. Xiong, T-Q. Thach, D. Zhang, P. Sham; The Univ. of Hong Kong, Hong Kong, China

Functional enrichment results typically implicate tissue or cell-type specific signaling or biological pathways in disease pathogenesis and as therapeutic targets. We propose generalized Linkage disequilibrium score regression (g-LDSC), a novel method that requires only genome-wide association studies (GWAS) summary level data to estimate functional enrichment. The method adopts the same assumptions and regression model formulation as stratified Linkage disequilibrium score regression (s-LDSC). While s-LDSC only partially utilizes LD information, our method utilizes the whole LD matrix which accounts for possible correlated error structure via a feasible generalized least squares estimation. We demonstrate through simulation studies under various scenarios, that g-LDSC provide less biased and more precise estimates of functional enrichment than s-LDSC, regardless of model misspecification. In an application to GWAS summary statistics of 15 traits from the UK Biobank, estimates of functional enrichment using g-LDSC were lower, and more realistic, than those obtained from s-LDSC. In addition, g-LDSC detected more significantly enriched functional annotations among 24 functional annotations for the 15 traits than s-LDSC (118 versus 51).

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4209 Improvements in Disease Risk Predictions Resulting from a Novel System of Local Ancestry Informed Polygenic Risk Scores

Authors:

S. Kulm[1], G. Busby[1], A. Bolli[1], J. Kintzle[2], G. Botta[1]; [1]Allelica, New York, NY, [2]Allelica, New York City, NY

Polygenic risk scores (PRSs) can identify individuals of high risk for disease who may benefit from targeted medical care. Nearly all PRSs are evaluated upon the same ancestry group of which the original variant effect sizes were estimated. This procedure fails individuals of admixed ancestry as they are not of a single, clear ancestry group. We hypothesizes that by inferring the local ancestry of an individual we could improve the risk predictions of admixed individuals.

To test this hypothesis we first used the FLARE tool, with the 1000 Genomes as reference, to infer the ancestry at single variant resolution for 8,294 individuals of admixed ancestry from the UK Biobank. To each continental ancestry specific segment we applied 150 difference panels of variants effect sizes designed to predict the risk of coronary artery disease. Under ten fold cross validation we created a ensemble PRS for each ancestry segment using the training fold and compared its performance in the testing fold to the single panel that was found to work best across the entire region of the genome.

Logistic regression models that adjusted for age and the top four genetic principal components, containing the sum of the ancestry-specific PRSs generated a lower brier score (0.05730 SE = 0.00256) than the PRS created from a global ancestry approach (0.05744 SE = 0.00252). Furthermore, the PRSs generated from just the majority specific ancestry segment, led to lower brier scores than the single global PRS when applied to groups whose ancestry were over 80% African, European or South Asian.

While improvements can be made that limit the use of discrete ancestry groups, this procedure provides one of the first large-scale examples of how local ancestry information can be practically combined with the traditional PRS creation process. Ongoing work shows that the improvement identified for coronary artery disease also extends to breast and prostate cancer. As a result, the predictive benefits of PRSs may now be more equitably applied to a wider share of individuals.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4210 Improving Diagnostic Specificity of Polygenic Risk Scores: A Methodological Approach Combining Large Biobank-Scale Data and Target Phenotype Cohorts with Limited Sample Sizes.

Authors:

Y-C. Lin[1], D. Pal[2], L. Strug[1]; [1]Univ. of Toronto, Toronto, ON, Canada, [2]King S Coll. London, London, United Kingdom

Polygenic risk scores (PRS) are increasingly being developed on large biobank-scale data to improve the prediction of genetic susceptibility to complex traits, which relies on large sample sizes. However, biobank-scale data can lack phenotypic specificity, leading to PRS being defined in cohorts with increased phenotypic heterogeneity and suffering from biased effect size estimates. The resulting PRS are thus not specific for the clinical samples where their application might be the most beneficial. For instance, cohorts with Idiopathic Generalized Epilepsy (IGE) have been used to construct PRS for Juvenile Myoclonic Epilepsy (JME), a sub-type which would benefit from the availability of specific biomarkers. IGE-PRS is correlated with many epilepsy subtypes (e.g., AUC; Rolandic Epilepsy (RE): 0.75; Electrical Status Epilepticus in Sleep (ESES): 0.63) including JME (AUC: 0.69), and therefore cannot offer diagnostic specificity which would be helpful in the clinic. Here we ask whether one can refine a PRS such as the IGE-PRS to improve specificity for sub-types such as JME, particularly when we only have access to (significantly smaller) genomic data from the JME clinical samples? We propose a framework that incorporates both the effect sizes from biobank scale GWAS of related phenotypes and the statistical evidence (e.g., p-values) obtained from GWAS on the specific phenotype of interest (e.g., JME) as penalized weights. We apply the methodology to (1) the IGE-JME example and (2) a second example that uses a large-scale attention deficit hyperactivity disorder (ADHD) GWAS to improve a smaller PRS of impulsivity (a component trait of ADHD) in JME measured by the Barratt Impulsivity Scale (BIS). We show that the derived JME-specific PRS is no longer correlated with the epilepsy subtypes (AUC; RE: 0.56; ESES: 0.53) but remains strongly correlated with JME (AUC: 0.75). We further show that for the BIS-PRS there is a three-fold improvement in predictive capacity compared to using the ADHD cohort alone (R^2; ADHD-PRS: 2.2%; BIS-PRS: 7.8%). Our framework represents a holistic approach to combine and learn from limited sample sized clinical cohorts and GWAS summary statistics derived from large biobank-scale data, to improve PRS predictive capacity for the phenotype of interest. We propose one such methodological approach to combine these data for improved predictive capacity. Although large biobank-scale data has provided unique opportunities to generate replicable PRS, improved applicability in the clinic can still be derived from the incorporation of clinically refined samples to improve external validity.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4211 Improving genetic analyses of cardiovascular traits via Multimodal REpresentation learning for Genetic Discovery on Low-dimensional Embeddings (M-REGLE)

Authors:

Y. Zhou[1], J. Cosentino[2], C. McLean[1], F. Hormozdiari[1], B. Behsaz[1]; [1]Google, Cambridge, MA, [2]Google, San Francisco, CA

EHR systems containing multiple high-dimensional clinical data (HDCD) modalities (e.g., ECG, PPG, MRI) for each individual provide a unique opportunity in clinical diagnosis and genetic studies of complex traits as multiple modalities relevant to a single organ system (e.g., circulatory system) may encode complementary information. We propose a novel deep learning method (M-REGLE) to discover genetic associations on multiple complementary HDCD and showcase its effectiveness for cardiovascular modalities. In M-REGLE, we jointly learn a lower representation (i.e., latent factors) of multimodal HDCD using a convolutional autoencoder (CAE) and then perform GWAS on each latent factor to study the genetics of the underlying system. To validate the advantage of M-REGLE's multimodal learning, we compare it to single modal learning where representations are learnt on each data modality separately and downstream analyses are performed on the summation of all representations.
We demonstrate the utility of M-REGLE on cardiovascular traits in UK Biobank by jointly modeling 12 leads of ECG and PPG waveforms. We trained multiple CAEs: i) trained on each ECG lead individually (SCAE-12E), ii) trained jointly on 12 leads (MCAE-12E), iii) trained individually on lead I ECG and PPG (SCAE-EP), and iv) trained jointly on lead I ECG and PPG (MCAE-EP). As baselines for comparison, we compute PCA on the same four input data settings. Multimodal learning produced better latent representations than single modal (MCAE-12E reconstruction MSE = 7.3e-5 vs SCAE-12E = 2.62e-4, MCAE-EP = 2.1e-4 vs SCAE-EP = 2.4e-4) and CAE outperformed PCA (MPCA-12E = 7.7e-5, MPCA-EP = 3.7e-4).
To show that M-REGLE improves statistical power, we performed GWAS on uncorrelated latent factors and then combined their GWAS. MCAE-12E detected 275 genome-wide significant (GWS) hits (20 more than SCAE-12E), 243 of which have been reported for ECG-related traits in the GWAS catalog. We then calculated the $E[X^2]$ statistics on all GWAS catalog variants for ECG-related phenotypes. We observed a significant power increase for MCAE-12E compared to the second best method (SCAE-12E) ($E[X^2]$= 130.6±2.4 vs 89.2±1.8; P<5.0e-38). GWAS performed on MCAE-EP showed a similar pattern: MCAE-EP detected the largest number of GWS hits, rediscovered the most known ECG-related hits, and improved GWAS power (46.2±1.0 vs 39.7±0.8 for SCAE-EP). CAE GWAS show significant cardiovascular functional enrichments (56-66 unique terms) and improve enrichment power over PCA (SCAE-12E over SPCA-12E, P=1e-10). We believe M-REGLE will become a standard method for using multimodal HDCD for GWAS and downstream analyses.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4212 Improving the understanding of psychiatric disorders using multivariate analysis of brain phenotypes

Authors:

A. Auvergne[1], N. Traut[1], L. Troubat[1], L. Henches[1], H. Julienne[1], A. Frouin[1], C. Boetto[1], S. Kazem[2], R. Toro[1], H. Aschard[1]; [1]Inst. Pasteur, Paris, France, [2]CHU Sainte-Justine, Montreal, QC, Canada

There are increasing evidences of genetic correlations between psychiatric disorders and brain magnetic resonance imaging (MRI) phenotypes. However, deciphering the joint genetic architecture of these outcomes has proven challenging, and new approaches are needed for inferring potential genetic structure underlying those phenotypes. Here, we demonstrate how multivariate analysis can reveal links between brain phenotypes and psychiatric disorders missed by univariate approaches.

We first conducted univariate and multivariate genome-wide association studies (GWAS) for eight MRI-derived brain volume phenotypes in 20K UK Biobank participants, using JASS, a robust multitrait analysis pipeline we recently developed. We then clustered these variants based on their multitrait association with MRI phenotypes using an optimized k-medoids approach along an innovative data-driven algorithm for selecting the number of clusters. We conducted enrichment analysis to assess whether and how these various (univariate, multitrait, and cluster) can distinguish disease-associated and non-disease-associated variants from six psychiatric disorders: bipolarity, attention-deficit/hyperactivity disorder (ADHD), autisms, schizophrenia, obsessive-compulsive disorder and major depressive disorder.

Univariate GWAS display only negligible genetic correlation with psychiatric disorders and top variants did not show any enrichment for association with psychiatric disorders. Conversely, the multitrait GWAS, which identified 49 significant loci including 27 not detected by the univariate analysis, show significant enrichment at top variants with both ADHD and schizophrenia ($P<10^{-7}$). The cluster-based analysis also detected two clusters of variants with enrichment for association with ADHD (P = 3.04e-9) and schizophrenia (P = 6.3e-10) and also a concordant direction of signals. By construction, those associated clusters correspond to linear combination of the volumetric phenotypes studied, and we argue that they can be interpreted as latent brain phenotypes. Altogether, this approach can demonstrate how multitrait analysis can be used to infer genetically driven neuroanatomical latent structure associated with psychiatric disorders.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4213 ImputeTrans: a transformer-based deep-learning tool for genotype imputation

Authors:

J. Fu[1], U. An[1], J. Choi[1], A. Ament[1], E. Eskin[2], S. Sankararaman[3]; [1]Univ. of California, Los Angeles, Los Angeles, CA, [2]Univ. of California Los Angeles, Los Angeles, CA, [3]UCLA, Los Angeles, CA

Genotype imputation is a computational strategy that predicts unobserved or missing genotype values to approximate whole-genome sequencing data. It is a foundational tool for fine mapping and identification of causal variants, which further facilitates meta-analysis that combines results from multiple studies, and increases the power of genome-wide association studies. Genotype imputation reduces expense and efforts for large-scale studies such as genome-wide association studies (GWAS) without losing much accuracy.
Currently, most state-of-the-art methods in this area employ a Hidden Markov Models (HMM) framework, while some others apply positional Burrows-Wheeler-Transform (pBWT). However, standard statistical methods may suffer in performance when working with regions involving complex linkage disequilibrium regions. In recent years, deep learning methods has proved its capability and suitability in imputation, and it has been widely-used in many areas such as imputation tasks including image imputation, as well as natural language processing such as sequence translation tasks. Recently, deep learning methods that incorporate autoencoders and transformer have also been developed for genotype imputation, which prove a relatively fast and accurate performance compared to some standard statistical methods, and requires no further access to a reference panel. However, these methods has not gone thorough comparison with other state-of-the-art methods.
Therefore, we present a fast and scalable method, ImputeTrans, which adapts a transformer-based deep learning method as a framework to impute genotypes. We evaluate this method comprehensively against current state-of-the-art methods in different scenarios. We show that our method achieves a robust and accurate performance and scalability, and gain improvements in certain aspects compared to standard methods.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4214 † Incorporating annotation stratified genetic covariance across complex traits to improve polygenic risk prediction

Authors:

C. Zhang[1], H. Zhao[2], G. Zhou[3]; [1]Yale Univ., new haven, CT, [2]Yale Univ. Sch. of Publ. Hlth., New Haven, CT, [3]Yale Univ., New Haven, CT

In recent years, the rapid development of novel polygenic risk score (PRS) methods have led to the widespread utilization of PRS in disease prevention, monitoring, and treatment. However, the accuracy of genetic risk prediction remains moderate. Currently, most PRS were built based on the summary statistics from the target trait, while many traits exhibit some degree of shared genetic architecture or pleiotropy dependent on the functional annotation. Appropriate leverage of pleiotropy and functional annotation can potentially improve the performance of PRS. In this study, we present PleioSDPR, a novel statistical method designed to leverage pleiotropy by integrating genome-wide association study summary statistics from various complex traits. PleioSDPR characterizes the joint distribution of effect sizes across different traits to be null for all traits, only causal in one trait, and causal in multiple traits with different functional and non-functional genetic covariance. Through extensive simulations and real trait applications, we demonstrate that PleioSDPR significantly improves prediction performance compared with PRScsx, SDPRx, and also weighted combinations of PRSs derived from LDpred2 and PRScs. For example, by incorporating information from schizophrenia (N = 130,644) or Neuroticism (N = 63,661), PleioSDPR effectively improves the prediction accuracy of bipolar disease(N = 353,899) and major depressive disorder (N = 142,646), respectively. Moreover, our findings demonstrate that traits exhibiting high genetic correlations and low overlapping sample sizes help more to the improvement of prediction accuracy of the target disease. Overall, our study highlights the potential of PleioSDPR to enhance the accuracy of genetic risk prediction by leveraging pleiotropy and considering a broader spectrum of traits and diseases. These findings contribute to the understanding of polygenic risk prediction and underscore the importance of incorporating pleiotropic information for improved utilization in disease prevention and treatment strategies.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4215 Incorporating external risk information with the Cox model under population heterogeneity: Applications to trans-ancestry polygenic hazard scores.

Authors:

D. Wang[1], W. Ye[1], J. Zhu[1], G. Xu[1], W. Tang[2], M. Zawistowski[1], L. Fritsche[1], K. He[1]; [1]Univ. of Michigan, Ann Arbor, MI, [2]Harvard Univ., Boston, MA

Background: Polygenic hazard scores (PHS) offer valuable insights for survival risk discrimination, yet these models, predominantly developed for European ancestry (EUR) individuals, encounter substantial limitations when applied to non-EUR populations. The inherent heterogeneity between external EUR-based PHS and internal non-EUR individual-level data can introduce substantial bias. To address this challenge, we propose a novel method designed to effectively integrate survival risk information across diverse populations, thereby enhancing the prediction accuracy of PHS models for non-EUR cohorts. **Methods:** Here, we develop a Kullback-Leibler (KL)-based Cox model (CoxKL) to incorporate external risk scores derived from published EUR-based PHS models with internal non-EUR genotype data with time-to-event outcomes. To adjust for population heterogeneity, partial-likelihood-based KL information is utilized to measure the discrepancy between the external risk information and the internal data. We also establish the asymptotic properties of the CoxKL estimator. **Results:** Simulation studies show that the proposed CoxKL method drastically improves both estimation efficiency and prediction accuracy. We applied the Cox method to develop trans-ancestry PHS models for prostate cancer and breast cancer in African ancestry (AFR) individuals. By integrating EUR-based PHS with internal AFR genotype data, the trans-ancestry PHS models by CoxKL method yielded considerable improvements on the risk discrimination of prostate cancer (C-Index: 0.68 [CoxKL PHS]; 0.54 [EUR-based PHS]; 0.49 [AFR-based PHS]) and breast cancer (C-Index: 0.83 [CoxKL PHS]; 0.53 [EUR-based PHS]; 0.50 [AFR-based PHS]). **Conclusion:** Our proposed CoxKL method serves as an accurate, efficient, and stable data integration tool, with an important application in developing trans-ancestry PHS models and enhancing survival risk discrimination in non-EUR populations.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4216 Incorporating information from relevant endophenotypes improves polygenic risk score prediction for asthma.

Authors:

E. Kharitonova[1], F. Ockerman[1], Q. Sun[1], L. Zhou[1], H. Cao[2], L. Raffield[1], A. P. Reiner[3,4], N. Cox[5], S. Kelada[1], R. Tao[5], Y. Li[1]; [1]Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, [2]Florida State Univ., Tallahassee, FL, [3]Univ. of Washington, Seattle, WA, [4]Fred Hutchinson Cancer Res., Seattle, WA, [5]Vanderbilt Univ Med Ctr., Nashville, TN

Current polygenic risk scores (PRS) for complex traits suffer from poor predictive performance. This is partially caused by the inability of studies to identify all genetic variants associated with a disease due to the limited number of cases in study populations. Since inflammation and immune response are known to contribute to a variety of health conditions, we hypothesized that endophenotypes could be used to improve PRS models. We focused on asthma as a test case as it is a common, chronic lung disease that is characterized in part by inflammation that typically features eosinophil and lymphocyte recruitment to the airways. As such, we reasoned that incorporating genetic variants into the asthma PRS model that have been identified to be associated with well-powered quantitative blood cell traits, i.e., endophenotypes, even if they have not yet been shown to be associated with asthma should improve the prediction power of the PRS model.

Here, we developed a novel PRS method which incorporates genetic information from endophenotypes. First, genetic variants associated with the clinical endpoint of interest and those associated with relevant endophenotypes are identified. Next, the selected variants are used to create a polygenic risk score through penalized regression, by weighting the variants based on whether they are associated with the clinical endpoint, the endophenotypes, or both. Unlike previous studies that developed polygenic risk scores based on the PRSs of many associated traits, our method integrates genetic variants from the endophenotype summary statistics directly into the PRS for the clinical endpoint without generating separate risk scores for each endophenotype.

We applied the novel polygenic risk score method to European individuals in UK Biobank to generate a PRS for asthma that incorporates genetic information from selected relevant blood cell traits (eosinophil count and lymphocyte percentage). We found that the proposed method significantly improves the prediction of asthma compared to most current PRS methods, including those designed for multiple traits (such as PRSice2, LDpred2, penalized regression, multi-PRS elastic net, and wMT-SBLUP) in a held-out test set of European individuals from UK Biobank.

In conclusion, we found that relevant endophenotypes can be used to improve polygenic risk score performance for asthma. However, due to the heterogeneity of asthma phenotypes, it is expected that this method's improvement in risk score prediction of asthma will vary in different cohorts based on the distribution of asthma subtypes in the populations of interest.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4217 Incorporating local ancestry information to predict genetically driven CpG methylation from SNP genotype data

Authors:

Y. Cheng, H. Li, A. Justice, K. Xu, H. Zhao; Yale Univ., New Haven, CT

To integrate results from genome-wide association studies (GWAS) with gene expression data from eQTL studies, transcriptome-wide association study (TWAS) has been developed, which aims to impute gene expression from SNP genotype data, and investigate its association with phenotypes. This framework has been widely used to identify the role of gene expressions in complex traits and to illustrate the path from genotype to expression, and then to phenotypes. It is also of interest to investigate this path with gene expression replaced by DNA methylation (DNAm), which plays an important regulatory role in complex phenotypes. Fryett et al. has extended the TWAS framework and proposed a methylome-wide association study (MWAS) model to impute DNAm from genotypes. However, their training and testing DNAm data are measured using array-based platforms, which profile only a small proportion of CpG sites in the human methylome. Also, the datasets used in their imputation model are based on a British ancestry population, which has limited utility to be applied to non-European populations. To address these gaps, we utilized Methylation Capture Sequencing (MC-seq) data in an admixed cohort (African Americans) to establish an MWAS framework. Our model was fitted by elastic-net regularized linear model, in which we incorporated the ancestral heterogeneity across genomic regions (local ancestry (LA)) to improve the DNAm prediction. Incorporating LA in our model enabled us to detect and control different genetic effects of each SNP by ancestry groups and allowed the model to be applied to cohorts with African, European, or admixed African and European ancestral backgrounds. Our MWAS framework was established using genomic data from the Veterans Aging Cohort Study (VACS) with a total of 2,244 participants. Among them, 377 samples self-reported as African Americans had DNAm data profiled by MC-seq, and this subset was used to train models. The remaining samples without MC-seq data were used as an application set to test the association between imputed methylation and phenotypes of interest. Using the tenfold cross-validation R^2 (the square of the correlation between predicted and observed methylation) in the training set, and R^2 in an independent testing set, we compared our model with the conventional model without LA information and demonstrated that our model improved the prediction R^2 (paired t-test p-value = 1.77E-139 in the training set, 5.69E-14 in the testing set). These findings highlight the importance of incorporating LA in MWAS frameworks for an admixed population.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4218 Incorporating short tandem repeats (STRs) improves genetic predictions of gene expression.

Authors:

M. Lamkin[1], Y. Li[1], H. Ziaei-jam[1], M. Gymrek[2]; [1]Univ. of California, San Diego, San Diego, CA, [2]Univ California San Diego, La Jolla, CA

Transcriptome wide association scan (TWAS) utilizes the imputation of gene expression from genotype datasets in order to identify expression-trait associations, which can help identify potential causal genes and cell types underlying complex traits. Multiple methods for predicting gene expression from genotypes exist, but all limit feature selection to common single nucleotide polymorphisms (SNPs) and thus may not capture variation in expression driven by other variant classes. Short tandem repeats (STRs), which consist of consecutively repeated units of 1-6bp are among the most polymorphic regions of the human genome. Recent studies have identified thousands of STRs for which variation in repeat length across individuals is associated with complex traits including gene expression. Here, we extend TWAS models to incorporate both linear and non-linear effects of STRs.
To train and test our linear models, we focused on 359 European individuals from the 1000Genomes Project, for which we have previously generated genome-wide STR genotypes and for which expression data for lymphoblastoid cell lines has been generated by the Geuvadis Project. Prior to training, we filtered and quantile normalized expression data and regressed out PEER factors, top 3 ancestry principal components, and sex. We used 5-fold cross validation to train and evaluate four different linear models, linear regression, Lasso regression, Ridge regression, and ElasticNet regression. For each linear model tested, we generated three feature sets: only SNPs, SNPs + linear STR effects, and SNPs + both linear and non-linear (quadratic) STR effects. These models were evaluated with an R squared score and compared to the prior state of the art models generated by TWAS / FUSION. Overall, when tested on twenty genes we saw that the STR models outperformed SNP models in six of the twenty. When testing on genes with current known eSTRs such as TIMM10 and CSTB, we saw an increase in R squared from the SNP Lasso model to the SNPs + linear STRs Lasso model by 0.028 and 0.038 respectively. We also saw an increase in the R squared of 0.038 and 0.054 when compared to the SNPs + both linear and nonlinear STRs Lasso model. Given these results, we've seen that incorporating STRs into the gene expression imputation models has increased predictive power of these models. Overall, our results suggest STRs explain additional heritability of gene expression and result in improved genetic predictors of gene expression.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4219 Increasing the Power of the Sequence Kernel Association Tests (SKAT) with Common Variant eQTLs

Authors:

Y. Yang[1], W. Bush[2]; [1]Case Western Reserve Univ., Cleveland, OH, [2]Case Western Reserve Univ, Cleveland, OH

Rare alleles arise from mutations on an existing haplotype background. However, current gene-based tests for rare-variants generally consider the impact of low-frequency coding variants as an independent effect from the more commonly occurring regulatory variants that surround them. In this work, we propose to increase the statistical power of kernel-based rare-variant association tests by accounting for the surrounding cis-regulatory variant's effects on gene expression. This method incorporates common variant cis-eQTLs around the gene or SNP-set of interest as a fixed effect by estimating their impact on gene expression. We estimate gene expression changes using PrediXcan association models as a covariate. SKAT is then used to test for the residual random effect of coding rare-variants. We perform comprehensive simulation studies that demonstrate the biological conditions under which our method can substantially increase power, while maintaining the type I error in the absence of association. SKAT is known to perform best when a high proportion of test variants have a causal effect, and when these effects are in the same direction. Our proposed method improves the power of SKAT where a low proportion(<10%) of tested variants have small mixed effect sizes (OR<1.0), more closely resembling real world variant effects. We also simulate a range of effect sizes for gene expression, to provide insight into the relationship between eQTLs and rare variants. Power analyses were performed with both exact p-values computed via the Davies method and with p-values derived empirically through permutation tests.We applied our approach to real data from the Alzheimer's Disease Sequencing Project for the amyloid precursor protein (APP) gene as a proof of principle. Our general approach can be extended to all current set-based rare variant association tests whose performances are also likely dependent on the impact of surrounding regulatory variation. This approach can then be extended to jointly consider multiple contributing genes in the same pathway, providing further information to detect rare variant effects.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4220 Inferring susceptibility of multiple populations to TCDD using estimates of its interaction with aryl hydrocarbon receptor (*AHR*) variants.

Authors:

U. Chimeh, M. C. Rogers, D. L. Aylor; North Carolina State Univ., Raleigh, NC

Current genetic literature mostly represents people of European ancestry and, consequently, does not adequately account for variants that are uncommon in European populations. Despite this deficit, we propose that information about trait variation in understudied populations can be gleaned from the variants that they share with studied populations. Here, we use the frequencies of single nucleotide polymorphisms (SNPs) in a well-studied nuclear receptor, aryl hydrocarbon receptor (AhR), to predict population risk. AhR is essential in the metabolism, and thus, the toxicity of compounds such as 2, 3, 7, 8-tetrachlorodibenzo-p-dioxin (TCDD) and polycyclic aromatic hydrocarbons (PAHs). The highest recorded TCDD exposure happened in 1976 in Seveso, Italy, from an industrial plant explosion. TCCD exposure from this accident differentially decreased offspring birthweight in individuals with different variants of 6 *AHR* SNPs.

We predict birthweight in multiple populations if they were similarly exposed to TCDD. To do this, we integrate the Seveso estimates of *AHR*-TCDD interaction with *AHR* variant frequencies in these populations. AHR variant frequencies were obtained from a public repository of human variation established by The 1000 Genomes Project. We identified haplotypes from the 6 *AHR* SNPs and used the European haplotypes as a proxy for those in Seveso. Most haplotypes (96%) in the 1000 Genomes pool were shared between Europe and at least one other population. Based on the distribution of risk variants, we predicted that average birthweight would decrease the most in the admixed American population and the least in Europe after TCDD exposure. The same was true for the increase in the proportion of underweight births (birthweights below 2500g). These interactions between *AHR* and TCDD have yet to be replicated in other populations; however, the risk allele of one SNP, rrs2282885, has been associated with increased susceptibility to PAHs in coke-oven workers from China. Our results suggest that toxicant exposure will have disparate effects on populations with different risk variant frequencies. This sets a precedent for predicting population susceptibility to toxicants and other adverse exposures using the frequencies of functional risk variants.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4221 † Integrating EHR-based biobanks and GWAS summary statistics to predict the progression of autoimmune diseases from pre-clinical outcomes.

Authors:

C. Wang, H. Markus, L. Carrel, G. Foulke, N. Olsen, B. Jiang, D. Liu; Penn State Coll. of Med., Hershey, PA

Many autoimmune diseases have a preclinical stage where early symptoms start to appear, but full-blown disease has not developed. Developing biomarkers to inform disease progression from pre-clinical stage will be critical for slowing down progression and mitigating symptoms. With electronic health record (EHR)-based biobanks, it has become feasible to identify individuals with pre-clinical disease status by lab tests and calculate polygenic risk scores (PRS) for disease progressions. Compared to case control (CC) studies, EHR-based biobanks have much fewer number of cases. To combine the large sample sizes of CC and detailed phenotypes in biobanks, we propose a novel method called Genetic Prediction Score (GPS) to predict disease progressions from pre-clinical stage. GPS incorporates PRS weights for CC as prior via a penalty term that penalizes deviations of the parameter estimates from the prior. Using data-driven validations, GPS forces parameter estimates to be similar to the CC weights when the prior information helps improve prediction accuracy and trains the model using the biobank data alone if prior does not help. We conducted extensive simulations considering different sample sizes of patient cohorts, genetic correlations between progression and CC phenotypes, and trait polygenicity. We compared GPS with different PRS construction strategies including applying PRS methods (LDpred2, Lassosum, and PRS-CS) to biobank data only, stacking CC and biobank only risk scores, using MTAG to analyze CC and progression traits and generate PRS using MTAG results, and TL-PRS, a transfer learning-based method incorporating pre-trained CC weights. In all scenarios, GPS consistently achieved the highest or comparable prediction accuracy. The improvement was particularly significant leading to 2x to 4x improvements in R^2 when genetic correlation between CC and progression trait was low and when the biobank had a limited sample size. For applications, we analyze progression phenotypes of systemic lupus erythematosus (SLE) from antinuclear antibody positive and rheumatoid arthritis (RA) from rheumatoid factor positive individuals in the BioVU biobank, and then validate progression prediction in All of US biobank. GPS demonstrates the strongest agreement between PRS deciles and observed progression prevalence among individuals within each respective decile. Compared to the second-best method TL-PRS, the estimates are 0.749 vs 0.370 for RA and 0.508 vs 0.466 for SLE. In summary, GPS is a useful tool to predict disease progression from pre-clinical outcomes and will play a key role in extending utility of PRS in the era of precision medicine.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4222 † Integrating genetic data in Trial Emulation: Exploring the potential of Polygenic Risk Scores.

Authors:

J. German[1,2], A. Ganna[1], A. Philippakis[2]; [1]Inst. for Molecular Med. Finland, Helsinki, Finland, [2]Eric and Wendy Schmidt Ctr., Broad Inst. of MIT and Harvard, Cambridge, MA

Using genetic information to identify individuals that would most benefit from pharmacological intervention is a central goal of pharmacogenomic and precision medicine. However, naively approaches in large-scale biobank data is subjected to confounding, which limits interpretability. Trial emulation approaches are the state-of-the art methodology to ask causal questions using real world data, and they thus represent a promising venue to test the impact of genetics on differential treatment response, when randomized control trials (RCTs) are not available or too small. In this study, we aim to investigate the utilization of polygenic risk scores (PRS) in trial emulation designs using 429,191 individuals from FinnGen, a biobank-based study with comprehensively longitudinal drug purchases information from national registers. We selected and emulated four trials using the same methodology as the RCT Duplicate project, which is the largest RCT replication initiative. Two RCTs (EMPAREG, TECOS) focused on T2D patients and two on atrial fibrillation (ARISTOTELE, ROCKET). We achieved estimate agreement in 3 out of 4 trials when considering the overlap of confidence intervals (95% CI) between the original RCT and our trial emulation. First, we assessed how genetics can help understanding confounding effects present in observational data. In the EMPAREG trial we show that using a trial emulation approach (including propensity score matching) results in reduced differences in a PRS for T2D between the treatment and control cohorts compared to a naïve approach (standardized difference = 0.0042, P-value = 0.9089 and standardized difference = 0.0875, P-value = 8.75e-10 for trial emulation and naïve approach, respectively). Second, within the successfully emulated trials, we investigated the potential interaction between PRS for the trial outcome and the drug, aiming to identify subgroups with differential treatment response. No significant interaction was observed, indicating that PRS did not identify individuals with better treatment response. In conclusion, our work highlights the importance of trial emulation when studying genetic impact on differential treatment response in biobank studies, as genetic studies are not immune to traditional epidemiological bias. While the idea of integrating genetics into trial emulation is promising, overcoming the challenges of statistical power, even at the scale of half a million individuals remains a critical area of improvement.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4223 Integration of polygenic risk scores with clinical factors improves 10-year risk prediction of coronary artery disease

Authors:

P. Tshiaba, D. Ratman, J. Sun, T. Tunstall, R. Maier, P. Shah, M. Rabinowitz, A. Kumar, K. Im; MyOme, Inc, Menlo Park, CA

Background: Polygenic scores (PRS) have the potential to improve the accuracy of clinical risk tools and identify individuals at elevated risk of coronary artery disease (CAD) not captured by risk models based on traditional risk factors. However, generalizability across cohorts and populations remains a major challenge for PRS models.
Methods: We trained ancestry-specific ensemble models by combining internally developed with publicly available PRS models for CAD and type 2 diabetes. The cross-ancestry PRS (caPRS) is a weighted sum of ancestry-specific ensembles. Then, we combined the caPRS with the Atherosclerotic Cardiovascular Disease - Pooled Cohort Equation (ASCVD-PCE) to derive the integrated risk score (IRS), an absolute 10-year risk of developing CAD. We compared the performance of IRS against ASCVD-PCE overall and for each self-reported ethnicity in the 3 validation cohorts. Hazard ratios (HRs) were estimated with Cox proportional hazards models adjusted for age and sex to examine the association between an SD change in each predictor (caPRS, IRS or ASCVD-PCE) and 10-year CAD incidence. Model performance includes discrimination assessed with the Harell's C index, calibration comparing observed vs expected event probabilities, and net reclassification index using a 20% high risk threshold.
Results: The caPRS was significantly associated with 10-year CAD incidence across all validation cohorts, including non-European ethnicities, with HRs ranging from 1.35 (95% CI: 1.16-1-56) in Africans up to 1.82 (95% CI: 1.43-2.32) in South Asians. The IRS showed improved discrimination compared to the ASCVD-PCE in all validation cohorts and ethnicities tested. The largest gain in performance was observed in South Asian individuals with a 6% increase in C-index. The IRS was better calibrated than the ASCVD-PCE across all validation cohorts. For individuals classified as borderline/intermediate ASCVD-PCE risk (5-20%), the IRS reclassified between 6% and 10% of individuals into high-risk category with the relative risk of 10-year CAD incidence ranging from 3.22 (95% CI: 2.90-3.58, UK Biobank) to 3.67 (95% CI: 2.93-4.59, Atherosclerosis Risk in Communities). Identifying up to 16.3 additional CAD cases per 1K individuals in the borderline/intermediate ASCVD-PCE group.
Conclusions: Adding a PRS to a commonly used clinical risk tool more accurately identified people at high risk of developing CAD across diverse populations. The integrated risk score efficiently reclassified individuals at borderline/intermediate clinical risk to high-risk demonstrating potential utility to guide preventive interventions, such as statin therapy initiation.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4224 † Integrative analysis of genetic studies and gene co-expression patterns highlights shared and distinct transcriptional mechanisms associated with asthma, COPD and asthma-COPD overlap (ACO) syndrome.

Authors:

M. Pividori[1], J. Joo[2], B. E. Himes[2]; [1]Univ. of Colorado, Aurora, CO, [2]Univ. of Pennsylvania, Philadelphia, PA

Asthma, chronic obstructive pulmonary disease (COPD), and asthma-COPD overlap (ACO) syndrome are complex respiratory disorders with some overlapping clinical features. Among the many genetic loci associated with asthma and COPD, some were shared with ACO, and a few potential ACO-specific loci have been observed. However, most of these studies have focused on individual variants and ignored context-specific gene-gene interactions.We performed GWAS using the UK Biobank (UKB), where we determined asthma-only status (n = 19,217) using ICD codes and self-reported doctor diagnosis, COPD-only status (n = 13,055) using spirometric evidence of moderate-to-severe airflow limitation by the modified GOLD criteria, and ACO status (n = 7,035) as those with asthma and COPD. Controls (n = 162,645) had none of these conditions. We then performed transcription-wide association studies (TWAS) for asthma-only, COPD-only, and ACO, and integrated them with gene co-expression modules learned from recount2, a large expression compendium. We used MultiXcan for TWAS to compute gene-trait associations. We defined 987 gene modules extracted with the Pathway-level Information Extractor matrix factorization approach on recount2, where a module represents a group of genes with similar expression profiles across the same contexts (such as tissues or cell types). We then computed gene module associations with asthma-only, COPD-only, and ACO by testing whether genes that were part of a module were more strongly associated than genes that were not. For associated modules, we conducted a phenome-wide analysis using additional gene module associations for more than 4,000 traits computed in UKB, and 309 phecodes in the Electronic Medical Records and Genomics (eMERGE) network phase III.We identified 24 significantly associated gene modules: 13 associations with asthma-only, 11 with COPD-only, and 6 with ACO. One of these modules, which was expressed in lung adenocarcinoma samples and was associated with smoking, neutrophils and rheumatoid arthritis in UKB and emphysema in eMERGE, was specifically associated with asthma-only and COPD-only. Four modules, which were expressed in different populations of tonsil innate lymphoid cells (ILCs) and associated with atopy, eosinophils and Crohn's disease in UKB, were specifically associated with asthma-only and ACO. Another module, which was expressed in lung and tonsil ILCs and associated with smoking in the UKB and atherosclerosis in eMERGE, was specifically associated with COPD and ACO.Our study highlights shared and distinct, context-specific gene modules that can help elucidate the genetic underpinnings of asthma, COPD, and ACO.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4225 Integrative multi-omics approach for improving causal gene identification

Authors:

C. Wu[1], A. King[2]; [1]UT MD Anderson Cancer Ctr., Houston, TX, [2]Florida State Univ., Tallahassee, FL

Transcriptome-wide association studies (TWAS) have become popular methods in exploring gene-trait associations, integrating gene expression datasets and genome-wide association studies (GWAS) to identify likely causal genes, thereby enhancing interpretability. Nonetheless, traditional TWAS approaches primarily focus on gene expression, making it challenging to prioritize causal genes due to linkage disequilibrium and correlations among gene expressions. Several other regulatory mechanisms, such as DNA methylation and splicing, substantially influence gene expression, contributing to the genetic underpinnings of complex traits and diseases. These mechanisms potentially provide orthogonal information useful for prioritizing causal genes. Here, we introduce a novel integrative omics model that aggregates gene expression, DNA methylation and splicing data to identify associated genes for our traits of interest. Specifically, we rely on gene expression, methylation and splicing prediction models in blood tissue based on data obtained from the eQTLGen Consortium, BIOS Consortium, and PredictDB group, respectively. These omics prediction models are then combined using an aggregated Cauchy combination test. By analyzing genome-wide association study (GWAS) summary statistics for 24 complex traits, we show that our integrated method, which leverages these complementary omics biomarkers, achieves higher statistical power, and improves the accuracy of likely causal gene identification in blood tissues than individual omics methods. Under Bonferroni correction, our innovative method identified 7,432 significant gene-trait pairs, outperforming TWAS (SUMMIT), MWAS, and splicing TWAS, which identified 3,478, 4,816, and 1,383 pairs, respectively. Moreover, our method demonstrated an AUC of 0.796 in identifying likely causal genes, reflecting a significant improvement over SUMMIT ($\Delta = 0.043$, p-value = 5.9x10-4) and MWAS ($\Delta = 0.024$, p-value = 9.3x10-4). Finally, we apply our integrated model to a lung cancer GWAS dataset, demonstrating the integrated models improved identification of prioritized genes for lung cancer risk.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4226 Intergenerational Socioeconomic Status and Genotype-by-Environment Interaction Impacting Metabolic Syndrome Traits: A Feasibility Study

Authors:

X. Mao[1], E. MANUSOV[2], V. Diego[1], E. Abrego[1], M. Almeida[1], J. Blangero[3], S. Williams-Blangero[1]; [1]Univ. of Texas Rio Grande Valley, Brownsville, TX, [2]Univ. of Texas Rio Grande Valley Rio Grande Valley, Brownsville, TX, [3]Univ. of Texas Rio Grande Valley Sch. of Med., Brownsville, TX

Background: Understanding the impact of intergenerational socioeconomic status (SES) and genotype interactions on metabolic syndrome severity among Mexican Americans is crucial. While previous studies have examined gene-environment interactions (GXE) in relation to health outcomes, limited research has explored the role of intergenerational SES, specifically parental education, in this context.
Objective: This feasibility study aims to elucidate the mechanisms underlying GXE interactions with respect to intergenerational SES factors. We hypothesize that parental education directly influences variations in offspring phenotypes, including BMI, lipids, and blood pressure (metabolic syndrome).
Methods: We employed Mendelian Randomization analysis to examine the causal effect of education on metabolic syndrome. The genetic instrumental variable effect (GIV-Wald ratio) was given by the ratio of the regression coefficient of the child's outcome on the GIV to the regression coefficient of the parent's exposure variable on the GIV. We calculated the Wald Ratio (W) using the Mahalanobis distance of gene expression variables related to adaptive immunity, innate immunity, and hemostasis systems, along with three parental exposure variables (education, income, and socioeconomic index), for waist circumference, blood pressure (SBP/DBP), triglycerides (TG), high-density lipoprotein (HDL), fasting glucose (FG), and fasting insulin (FI).
Results: The results of the Mendelian Randomization analysis and Mahalanobis distance calculations will provide insights into the causal relationship between parental education and offspring metabolic syndrome phenotypes.
Discussion: Parental SES has a direct impact on offspring outcomes through socioeconomic effects, and GXE effects of SES on metabolic syndrome phenotypes have been established. Investigating the intergenerational effects of SES through genetic variation is a novel approach. Utilizing the Wald-ratio from Mahalanobis distance as an instrument in the Mendelian Randomization as a quasi-randomized trial to investigate GXE effects represents another innovative aspect of this study.
Conclusion: This feasibility study utilizes a novel approach to analyze the interaction of genetic variation and phenotypes, employing a causal relationship model. The findings from this study will contribute to our understanding of intergenerational socioeconomic status effects on metabolic syndrome and inform future research in this area.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4227 Interplay between *EIF2AK4* and *HLA-DRA* genes is associated with human survival to the oldest old age: New findings from the UK Biobank data

Authors:

A. Yashin, D. Wu, H. Duan, R. Holmes, O. Bagley, A. Yashkin, I. Akushevich, K. Arbeev, S. Ukraintseva; Duke Univ., Durham, NC

Introduction: Experimental studies show that the *GCN2/EIF2AK4* gene, which is a major sensor of amino acid deprivation in the integrated stress response pathway, is involved in the regulation of aging and survival. The *HLA-DRA* gene belongs to the human leukocyte antigen gene family (class II region) which encodes molecules involved in response of adaptive immunity to infections which play important role in organism's survival. It remains unclear whether interplay of these genes is associated with human survival. The objective of this study is to test the presence of such connection using human data. **Method.** We estimated associations of interactions between SNPs in *EIF2AK4* and *HLA-DRA* genes in UK Biobank participants, with survival of their mothers (114,573-cases and 213,618-controls) and fathers (66,206-cases and 288,281-controls) using a logistic regression model with interaction. Education, sex, smoking status, and the first five genetic principal components were included as observed covariates. The primary outcome was the parent survival to age 85. Linkage Disequilibrium (LD) testing and the SNP-clumping procedure were used to reduce the number of tests in SNPxSNP interaction analysis of survival. **Results.** We found three statistically significant associations of interacting SNPxSNP pairs with mothers' survival (one SNP, rs35602605, in *EIF2AK4*, and three -- rs3129887, rs2239805, rs115317719 -- in *HLA-DRA*) with β coefficients ranging from 0.0508 to 0.0523, with p-values ranging from 2.37E-04 to 2.78E-04, and with the Bonferroni correction threshold 4.13E-04. The associations of interactions between SNP-pairs with fathers' survival were nominally significant and consistent with the results for mothers' survival. They involve the same three *HLA-DRA* SNPs interacting with rs2412463 and rs11635537 SNPs in *EIF2AK4*. The LD between these SNPs and rs35602605 is characterized by D'=1.0, R^2=0.10 for rs11635537, and D'=1.0, R^2=0.09 for rs2412463 SNPs, respectively. **Conclusion.** The observed associations between SNPxSNP interactions within the *EIF2AK4* and *HLA-DRA* genes and survival beyond age 85+ show consistency of mechanisms across biologic sex indicating that genetic machinery regulating survival in the part dealing with nutritional stress and immunity have overlapping components in males and females. These results also show that hypotheses about the role of genetic interactions in complex traits derived from experimental data can be tested using human data. More analyses will allow for a better understanding of the role of interplay between multiple genes in multifactorial regulation of human aging, health, and lifespan related traits

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4228 Interpretable machine learning models for pre-screening cardiomyopathies.

Authors:

M. Kasapi, K. McGurk, T. M. D. Ebbels, J. M. Posma, J. Ware; Imperial Coll. London, London, United Kingdom

Cardiomyopathies (CM) intervene with the normal function of the heart muscle and are the most common cause of sudden death in young adults. The two most common inherited CMs are Hypertrophic Cardiomyopathy (HCM) and Dilated Cardiomyopathy (DCM). Rare variants account for less than 50% of cases and recent GWAS have shown that SNP heritability has a strong polygenic influence on the phenotype which can account for the rest of the cases. As we move towards an era where genome sequencing becomes part of routine care, it is imperative to use these personalized data types to streamline patient care and contribute to the better diagnosis of these diseases. We used the UK Biobank, alongside an internal HCM cohort, to train models that can pre-screen for HCM and DCM via prediction-classification modelling. Models are trained on basic clinical and demographic features with HCM/DCM-specific genomic information. This includes rare variants in genes associated with HCM and DCM, which are curated based on evidence for disease, protein consequence, and minor allele frequencies, among other filters. We compared 4 distinct machine learning methods: Support Vector Classifiers (SVC) with both linear and non-linear kernels, and 2 ensemble methods, Random Forests (RF) and Gradient-Boosting Trees (GBT). The classifiers were trained 100 distinct times, where each time random training and test splits (80:20 ratio) were done ab initio (total sample sizes: HCM: 1040, DCM: 921, Controls: 7944). Splits of sets were tracked to ensure that the same train-test split was assigned across different models, to allow for robust modelling. Data samples were kept moderately imbalanced to reflect the prevalence of HCM and DCM in the general population, while ensuring good sample size for training. Performance was assessed by precision and recall. For HCM, the SVCs showed the highest recall, 0.79±0.026 and 0.77±0.025 for linear and non-linear kernels, respectively. The ensemble methods showed the highest precision with 0.94±0.017 for RF and 0.93±0.019 for GBT, whereas the SVCs showed slightly lower values (linear: 0.78±0.028, non-linear: 0.85±0.02). In most cases we observed that non-HCM controls are correctly classified. For DCM, the highest recall is also with SVCs (linear 0.73±0.03, non-linear:0.71±0.03), however the precision in these models is below 0.5. We observe the opposite trend in ensemble methods with high precision (RF: 0.85±0.01, GBT: 0.84±0.01) and recalls below 0.5. Our findings indicate that integrating an SVC with an ensemble method yields the best performance, representing a promising step in the development of an interpretable ML pipeline for pre-screening cardiomyopathies.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4229 Interrogating MR instruments for biological plausibility.

Authors:

C. Fawcett, C. P. Nelson, V. Codd; Univ. of Leicester, Leicester, United Kingdom

Mendelian Randomisation (MR) is a method of investigating the causal relationship between an exposure and outcome, using genetic variants as unconfounded instruments and proxies to examine this relationship. However, many instruments are generated using all SNPs previously identified to be of statistical significance in a GWAS. This means that these instruments fail to examine the important nuance which may be revealed by further biological consideration of which SNPs are included in an instrument. A 130 SNP telomere length instrument was generated by Codd et al. (2021) using genome-wide data. By examining these 130 SNPs and identifying their potential roles, we can categorise these by established and predicted biological relevance. These categories allow for new biologically-relevant instruments to be constructed, with varying levels of strength and evidence of involvement in telomere biology. This allowed for the creation of five tiers of MR instrument, with differing levels of biological relevance in telomere length. Using this five-tiered system, we reinvestigate over 100 diseases in UK Biobank to investigate the role of telomere biology on previously established causal associations. Re-examining previous MR findings with this tiered system can provide new insights into the mechanisms underlying the relationship between telomere length and disease. By categorising a genome-wide instrument into biologically relevant tiers, it is possible to uncover more information about important factors linking exposures to outcomes. This is of particular importance as MR is used routinely to establish causal relationships with genome-wide instruments which do not consider the role of included SNPs.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4230 Introducing an optimized, automated pipeline for phasing, local ancestry inference, and Tractor GWAS on admixed cohorts

Authors:

N. Shah[1], A. X. Maihofer[2], T. Tan[1], J. Mauer[1,3], C. Zai[4], M. Santoro[5], C. Nievergelt[6], E. Atkinson[1]; [1]Baylor Coll. of Med., Houston, TX, [2]UCSD, La Jolla, CA, [3]Univ.e Federal de São Paulo, São Paulo, Brazil, [4]Ctr. for Addiction and Mental Hlth., Toronto, ON, Canada, [5]Federal Univ. of Sao Paulo, Sao Paulo, Brazil, [6]Univ California San Diego, La Jolla, CA

Admixed individuals have been traditionally excluded from association analyses, contributing to disparities and Eurocentric biases in post-association analyses such as polygenic risk scores (PRS) and fine-mapping efforts. While recent efforts to sequence admixed individuals globally have shown promise in addressing this issue, limited methods exist to fully overcome it. A key challenge is the presence of randomly distributed stretches with variable lengths from different ancestral backgrounds, with individuals having 0, 1, or 2 copies of specific ancestral segments. This variability can lead to false positive and false negative results, particularly for ancestry-specific risk loci. Admixed populations may also exhibit different patterns of linkage disequilibrium (LD), complicating LD-based methods and compromising association analysis accuracy. To overcome false-positive results and produce ancestry-specific summary statistics, we propose the *Tractor* GWAS method, leveraging local ancestry for association analysis in admixed cohorts. Unlike traditional GWAS, the *Tractor* GWAS method involves multiple steps: phasing, local ancestry inference, and subsequent regression analysis with local ancestries. These steps require some preprocessing and nuance which may pose a barrier to adoption. To overcome this barrier, we have developed a scalable and user-friendly pipeline for conducting *Tractor* GWAS in collaboration with the PGC-PTSD Ancestry Working group. We have identified appropriate tools/methods for each step, curated tailored reference panels for 2-way and 3-way admixed individuals based on the recently released TGP/HGDP joint-call dataset, and containerized the workflow in Docker for easy accessibility to users. Our pipeline uses QC'd PLINK data files as input for statistical phasing, which are then processed for local ancestry inference. These local ancestry calls are subsequently employed for association analyses using the *Tractor* method. By adopting our approach, researchers can achieve harmonized analysis and facilitate cross-cohort comparisons, thereby identifying ancestry-specific loci beyond well studied homogeneous populations. We aim to promote wider adoption of the *Tractor* GWAS method by offering a more user-friendly approach, particularly benefiting groups with limited bioinformatics expertise. Through our efforts, we strive to overcome challenges associated with admixed individual analyses, eliminate biases, and emphasize the significance of the *Tractor* GWAS method in advancing genetic research.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4231 Investigating the Genetic Association between Type 2 Diabetes and Dementia.

Authors:

Y. Jobbi[1], N. ElHouda Barhoumi[2], E. Ahmed Fouad ElHady[3], A. Olaitan I.[4]; [1]Aminu Kano Teaching Hosp., Kano, nigeria, Kano, Nigeria, [2]Univ. of Med., Tunis, Tunisia, Tunis, Tunisia, [3]Faculty of Med., Ain Shams Univ., Cairo, Egypt, Cairo, Egypt, [4]Dept. of Computer Sci., Univ. of Ibadan, Ibadan, Oyo State, Nigeria, Ibadan, Nigeria

Dementia is a chronic, progressively degenerative brain syndrome. It is one of the major causes of disability in older populations. Type 2 diabetes mellitus (T2D) is a well-known risk factor associated with dementia. Older patients with T2D have a higher risk of about 60% of dementia when compared to those individuals without the disease. This study aims to identify novel genetic biomarkers that link dementia with type 2 diabetes in humans using Insilco analysis. NCBI, GEO2R, NIH Microarray expression dataset GSE161355 (Affymetrix GPL570 platform, the Affymetrix Human Genome U133 Plus 2.0 Array) from the brain temporal cortex neurovascular units of T2D and matched control samples. The findings from this study were downloaded, and further filtrations were performed using R- language, and limma to remove probe sets without corresponding gene symbols. Genes with multiple probes were filtered to retain features with maximum variance using the WGCNA collapseRows function, and significant Differentially Expressed Genes (DEGs) were identified using an unadjusted p-value < 0.05 and |log fold change (logFC)| > 1. The upregulated and downregulated genes were determined by the positive and negative values of the logFC values, and the data visualization of differential analysis (cases vs controls) was designed by EnhancedVolcano package for volcano plots using R- language. The Gene Ontology (GO) analysis indicated that genes from diabetes were mainly concerned with macromolecule biosynthesis, oxidation of organic compounds etc. The results of the cellular component were significantly mentioned in mitochondria, and respiration chain system, while the results of KEGG analysis were significantly enriched in tumor necrotic factor (TNF) signaling pathway, cell cycle, and inflammatory pathways. The understanding of this study revealed that cytochrome oxidase subunits 5A, 6C, and 7C are the potential biomarkers in T2D and dementia in the mitochondrial electron transport chain, and their upregulations lead to oxidative phosphorylation. In the present study, we performed bioinformatics analysis to investigate the link between dementia and T2D. We found out the shared relationship is based on mitochondrial dysfunction and altered metabolism in the related body cells. Cytochrome oxidase was identified as a potential biomarker of these diseases. Our findings provided new insights and directions to the understanding of dementia and T2D and can be used to understand better the pathology of these diseases and develop better treatment options. Nevertheless, these findings need further validation using animal models and related clinical investigations.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4232 Investigating the impact of different copy number variant region definitions on copy number variation association analysis with whole genome sequencing data: Lessons learned from the Alzheimer's Disease Sequencing Project (ADSP).

Authors:

A. Tucci[1], H. Wang[2], Y. Si[1], Y. Cheng[1], L-S. Wang[2], G. Schellenberger[2], J-Y. Tzeng[3], W-P. Lee[2]; [1]North Carolina State Univ., Raleigh, NC, [2]Univ. of Pennsylvania, Philadelphia, PA, [3]North Carolina State Univ, Raleigh, NC

Background: The heritability of Alzheimer's disease conferred through single nucleotide variants (SNVs) identified through genome wide association studies accounts for less than half of the total estimated heritability. Copy number variations (CNVs) within the genome are an alternative form of genetic variation that could explain some of the remainder of these AD heritability estimates. A critical early step in CNV association analysis is to define a "CNV region" (CNVR) to use as the locus unit. Selecting a specific definition for CNVRs is a non-trivial task that affects the analysis and result interpretation. In addition, misaligned CNV breakpoints that arise in whole genome sequencing (WGS) datasets with large samples can further complicate the impact of these CNVR definitions. We focus on elucidating how different CNVR definitions can influence a CNV association analysis of AD.
Methods: Using WGS data from the Alzheimer's Disease Sequencing Project (ADSP) R3 dataset of 16,905 samples, we called CNVs with Manta and Smoove and consolidated using the joint-genotyping algorithm GraphTyper2 to generate a CNV callset. Three different approaches (as defined by CNVRuler) were used to define CNVRs: density trimming, reciprocal overlap, and fragment-based. For the density trimming-based approach, two sets of CNVRs were created using density thresholds of 0.1 and 0.5. Each CNVR set was then used to perform an AD association analysis via logistic regression. The abundance of CNVRs defined and the regions identified as statistically significant were used to compare CNVR definitions to help construct guidelines for future such analyses.
Results and Discussion: Focusing on regions of genes previously identified as associated with AD status through variation in copy number, we observe significant differences between the three definitions of CNVRs. For CNVRs defined using density trimming, adjustment of the density threshold from 0.1 to 0.5 did not substantially impact the number of CNVRs defined, and neither identified a significantly associated region. In contrast, reciprocal overlap and fragment-based approaches defined more granular regions, resulting in improved sensitivity for identifying regions potentially associated with AD. The fragment-based method can provide the most detailed resolution, but is sensitive to misaligned breakpoints and may be better coupled with joint-genotyping methods that attempt to resolve this, like GraphTyper2. Ultimately, reciprocal overlap remained more sensitive than density trimming without being as susceptible to generating spurious CNVRs that represent sequencing noise rather than true genetic events.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4233 Investigating the role of rare microRNA-associated germline variants in the epileptic encephalopathies

Authors:

I. Fashina[1,2], Genomics England Research Consortium, S. J. Furney[1], G. L. Cavalleri[1]; [1]RCSI, Dublin, Ireland, [2]SFI Ctr. for Res. Training in Genomics Data Sci., Galway, Ireland

The epileptic encephalopathies (EE) represent rare and devastating forms of epilepsy with a largely monogenic genetic architecture. Although a genetic cause is identifiable in around 40% of cases, the majority screen negative on genomic testing. Rare damaging variants have been shown to be overrepresented in EE cohorts compared to controls. However, these studies have largely been limited to coding variants from whole exome sequencing (WES) data. Therefore, the role of rare variants within non-coding elements of the genome is understudied in these epilepsies. MicroRNAs are small ~22nt non-coding RNAs which act as master gene regulators, by binding to the 3'UTR of their mRNA targets. microRNAs have been observed to be dysregulated in studies of epilepsy patient samples, animal models as well as through in-vitro experiments. Using whole genome sequencing (WGS) data, we set out to test the hypothesis that rare microRNA-associated variation are contributing to genetic risk for the EEs.
We analysed whole genome sequencing data (WGS) in 1836 individuals (348 EE cases and 1,488 chronic kidney disease controls) accessed through the rare disease programme of the Genomics England 100,000 Genomes Project. Using SAIGE-GENE v0.44 in the Aggregate Variant Testing (AVT v3.1) pipeline of the Genomics England Research Environment (GEL RE), we assessed the enrichment of rare damaging variants across 4 functional domains. These domains are within coding sequences (CDS), conserved 3'UTRs and predicted microRNA-binding 3'UTRs of 540 confirmed epilepsy genes from the GEL panel application. We also tested enrichment of rare variants in microRNA-encoding genes.
We will present results of enrichment for rare damaging variants across functional domains, in coding sequences, predicted 3'UTR microRNA-binding sites and microRNA-encoding genes.
Ours is the first study to use WGS data to test the enrichment of rare damaging variants outside of coding sequences, particularly with regards to microRNA regulation in the epilepsies. We will extend this analysis to larger epilepsy cohorts.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4234 *Is inherited muscle strength a proxy for intrinsic capacity to resist age related pathologies and mortality?*

Authors:

E. Sillanpää[1], P. Herranen[1], K. Koivunen[1], T. Rantanen[1], K. Tiainen[1], U. Kujala[1], S. Ripatti[2], T. Palviainen[3], J. Kaprio[3], FinnGen, K. Waller[1], L. Joensuu[1]; [1]Univ. of Jyväskylä, Jyväskylä, Finland, [2]Inst. for Molecular Med. Finland FIMM, Helsinki, Finland, [3]Inst. for Molecular Med. Finland, Helsinki, Finland

Background: Low muscle strength has been associated with lower risk for diseases and premature mortality. We hypothesize that muscle strength genotype may reflect individual's intrinsic capacity to resist age-related pathologies, as well as predict better survival after acute diseases and longer lifespan. These hypotheses, and whether the potential association of muscle strength genotype and mortality can be modified with long-term adherence to physical activity (PA), were tested in Finnish population-based datasets.Methods: A polygenic score for hand grip strength (PRS HGS, ~1M SNPs) was constructed using Pan-UK Biobank (N=418,776) as a base data and SBayesR methodology. Association tests were conducted in the FinnGen cohort (n=342,443, age 40-108-y). For PRS HGS validation (n=429) and to test if physically active lifestyle can modify the association of PRS HGS and mortality (n=5,275) we used the Finnish Twin Cohort (FTC). Disease diagnosis, death dates and causes (all-cause and cardiovascular [CVD]) were obtained from the national digital registers. PA was assessed three times during years 1975-1990 using validated questionnaires. Linear and Cox regression models were utilized.Results: PRS HGS explained 6.1% of the variation HGS and 5.4% in knee extension strength in FTC. In FinnGen, a higher PRS HGS predicted a lower body mass index (β=−0.112 kg/m^2, P=1.69x10^{-11}) in women but not in men (β=0.004 kg/m^2, P=0.768). A higher PRS HGS associated with reduced risk for selected cardiometabolic diseases 3-6%, chronic pulmonary diseases 6%, musculoskeletal and connective tissue diseases 2-10%, depression 5%, and vascular dementia 7%. Only in women a higher PRS HGS associated with decreased risk for any dementia 6%, and for Alzheimer's disease 4%. Participants with a higher PRS HGS had 4 % decreased risk for cardiovascular and 3% for all-cause mortality. PRS HGS was not associated with better survival after adverse acute health events. In FTC a total of 1,330 deaths occurred (433 CVD related) during the mean 17.5-y follow-up. We found a significant sex*PRS HGS interaction (p=0.024, p=0.032) in all-cause and CVD mortality prediction, respectively, with higher PRS HGS associated with decreased mortality risk in men (HR 0.93 and HR 0.88) Associations persisted after adjusting for PA. No PA*PRS HGS interactions were found. In FTC PRS HGS did not predict mortality in women. Conclusion: The genotype that supports higher muscle strength protects against future health adversities but might not reflect the recovery potential after severe adversity. Low inherited muscle strength may be a predictor of CVD mortality in men regardless of their adulthood physical activity.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4235 Is linearity too strong of an assumption?

Authors:

R. Devito[1], D. Gulati[1], M. Gymrek[2]; [1]Univ. of California San Diego, San Diego, CA, [2]Univ California San Diego, La Jolla, CA

State of the art polygenic scores (PGS) assume additive and linear models of how variants and covariates influence a phenotype. While this has benefits in terms of ease of implementation and interpretability, it can lead to model-mismatch effects when the assumption is broken. Linear models' inability to capture these nonlinear interactions negatively impacts their performance both in predicting the phenotype and interpreting which variants are causal. Evaluating the extent of this mismatch is difficult, as we lack both ground truth understanding of the genetic architectures that lead to different phenotypes and genomics tools for fitting nonlinear models to biobank scale WES data.

To address the first point, we introduce a new open source phenotype simulation tool, CITRUS, where complex phenotypes are defined from genotypes using a user-defined graphical model. These networks allow modeling of arbitrary cis- and trans-acting epistatic interactions, haplotype effects, and other linear and nonlinear dynamics numerically. CITRUS provides a ground truth measure of importance for individual variants and pathways (interactions between individual variants) using approximated Shapley values, which can be used to benchmark model interpretation methods. We use these simulations to demonstrate that model-mismatch leads to worse predictive and variant identification performance as the nonlinearity of the phenotype increases for linear models when compared with nonlinear models.

To address the second challenge, we introduce a new PGS using tree-based methods which have demonstrated superior performance in traditional machine learning tasks. Prior studies applying nonlinear models to polygenic scores and causal variant identification have made sacrifices in either the types of nonlinearity they could capture or requiring a feature selection step that removed real signal from the model input, as they could not scale to the input feature size (number of variants) in traditional PGS. We address these shortcomings with an autoML pipeline for fitting tree-based nonlinear models (random forests, XGBoost, and LightGBM) for PGS. The pipeline is designed for scaling to biobank sized data in cloud and HPC settings and provides feature importance and feature interaction scores using Tree SHAP. We include additional libraries for preprocessing genotype, phenotype, and covariate data at scale. We compare phenotype prediction performance on UK Biobank data between hyperparameter optimized linear and nonlinear models, compare variants identified as associated, and for the tree-based models look for examples of pairwise nonlinear variant-variant interactions.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4236 isGWAS: ultra-high-throughput, scalable and equitable inference of genetic associations

Authors:

Z. Kuncheva[1,2], C. N. Foley[1,2], R. Marioni[3], H. Runz[4], B. Sun[4]; [1]Optima Partners Ltd, Edinburgh, United Kingdom, [2]Bayes Ctr., Edinburgh, United Kingdom, [3]Univ. of Edinburgh, Edinburgh, United Kingdom, [4]Biogen Inc, Cambridge, MA

The accelerating accumulation of genetic data and samples in large-scale deeply phenotyped cohorts poses significant computational, financial and data privacy bottlenecks for massive-scale genome-wide association analyses (GWAS). These limitations hinder collaboration and accessibility, particularly to less financially flexible research groups. Here, we present in-silico GWAS (isGWAS), a new algorithm and tool which can bypass the need to access or hold individual-level genetic data, requiring instead sample-level sufficient statistics to rapidly estimate regression parameters in case-control GWAS. Our approach offers several key advances over current gold-standard approaches: (1) massively improved computational efficiency and unique scalability for sequencing scale data across any sample size (e.g., GWA analyses of ~11m variants is completed in 4 minutes achieving ~1000-fold improvement relative to current approaches); (2) a new re-sampling approach extrapolating locus-level GWA results to larger sample sizes, facilitating future cohort and scenario planning; (3) the method requires only four input features per genetic variant: the allele frequencies in cases and controls and the number of cases and controls. These sufficient statistics satisfy wide-ranging data privacy concerns facilitating integration of case and/or control data from distinct studies. isGWAS makes feasible routine testing of alternative definitions of case status with no significant additional financial cost or carbon footprint. We showcase concordance with existing methods and the impact of isGWAS using biobank and disease consortium cohorts.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4237 Joint analysis of multiple phenotypes for extremely unbalanced case-control association studies using multi-layer network

Authors:

H. Xie[1,2], X. Cao[1], S. Zhang[1], Q. Sha[1]; [1]Michigan Technological Univ., HOUGHTON, MI, [2]Medpace Inc., Cincinnati, OH

Genome-wide association studies (GWAS) are an essential tool for analyzing associations between phenotypes and single nucleotide polymorphisms (SNPs). Most of the binary phenotypes in large biobanks are extremely unbalanced, which leads to inflated type I error rates for many widely used association tests for joint analysis of multiple phenotypes. In this article, we first propose a novel method to construct a Multi-Layer Network (MLN) using individuals with at least one case status among all phenotypes. Then, we introduce a computationally efficient community detection method to group phenotypes into disjoint clusters based on the MLN. Finally, we propose a novel approach, MLN with Omnibus (MLN-O), to jointly analyze the association between phenotypes and an SNP. MLN-O uses the score test to test the association of each merged phenotype in a cluster and an SNP, then uses the Omnibus test to obtain an overall test statistic to test the association between all phenotypes and an SNP. We conduct extensive simulation studies to reveal that the proposed approach can control type I error rates and is more powerful than some existing methods. Meanwhile, we apply the proposed method to a real data set in the UK Biobank. Using phenotypes in Chapter XIII (Diseases of the musculoskeletal system and connective tissue) in the UK Biobank, we find that MLN-O identifies more significant SNPs than other methods we compare with.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4238 Joint fine-mapping of single variants and gene-based tests from exome sequencing and genotype imputation

Authors:

Y. Zou, J. Mbatchou, L. Gurski, M. Suciu, Mayo Clinic Project Generation, M. Ferreira, J. Marchini; Regeneron Genetics Ctr., Tarrytown, NY

Fine-mapping analysis has traditionally focused on imputed common variants due to the low imputation accuracy for the rare variants. The identified causal variants are therefore usually non-coding variants, which can be difficult to map to nearby potential causal genes. To identify the nearby causal coding rare variants or genes, one common strategy is to analyze sequenced rare variants or gene-based burden tests conditional upon nearby common variant associations. One disadvantage of this approach is that it does not fully explore the space of causal models, so may lose power to detect the true causal variants when the Linkage Disequilibrium (LD) structure between single variants and gene-burden masks is complex.

To overcome the issue, we present a novel pipeline that enables joint fine-mapping of imputed common variants, sequenced rare variants, and gene-based burden masks. The pipeline takes summary statistics (z scores), genotype files with single variants (imputed and sequenced variants), and gene masks definition files. We incorporated functionality to REGENIE to compute the LD between single variants and gene masks, where the gene masks are dynamically generated according to mask definitions and annotations. We use SuSiE-RSS to identify putative causal variants or genes. In simulation studies using genotype data from an RGC cohort (Project Generation), our joint fine-mapping approach identifies the coding or gene causal signals, without losing power in identifying non-coding causal variants using 95% credible sets (CSs). To further illustrate the efficacy of the approach, we test the pipeline on empirical data using a quantitative trait, eosinophil count data from the same RGC cohort Project Generation. We identified 121 independent causal signals (95% CSs), out of which 114 contain single variants only (90 CSs with imputed common variants only, 8 CSs with sequenced rare variants only). In addition to the single variant signals, we are able to identify 7 putative causal genes (eg, IL33, SH2B3, SRSF2, S1PR4), some of which have been functionally validated to play a role in lymphocyte and eosinophil development or allergic traits. Overall, our joint fine-mapping pipeline thoroughly explores the entire range of potential causal variants, allowing for comprehensive identification of rare variants and genes.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4239 † Joint modeling of genomic and exposomic data for variance decomposition analysis to improve heritability estimation of complex human phenotypes

Authors:

F. Akhtari[1], X. Tong[2], A. Burkholder[1], J. S. House[1], D. C. Fargo[1], C. S. Schmitt[1], J. E. Hall[1], A. A. Motsinger-Reif[1]; [1]NIEHS, Durham, NC, [2]Univ. of Kentucky, Lexington, KY

The etiologies of common complex diseases involve both genetic and environmental factors and their interactions. According to the World Health Organization, approximately 23% of all deaths and 26% of deaths among children under five years of age are due to modifiable environmental factors. These include factors such as air pollution, water and sanitation, hazardous chemicals, and occupational exposures. Calculating heritability is vital for understanding disease etiology and helps anchor expectations in gene mapping. While SNP-based heritability estimation for common complex human phenotypes using genotype or whole genome sequencing data is common, few studies have estimated the environmental component of phenotypic variance for complex diseases due to insufficient exposure data to evaluate the contribution of environmental factors. Comprehensive data collection on various environmental exposures is needed to better understand the environmental component in disease etiology. The Personalized Environment and Genes Study (PEGS) study is a diverse North Carolina-based cohort with extensive health and exposure data and whole genome sequencing (WGS) data from individuals of varying age, race, education, and socioeconomic status. To improve heritability estimation, and to better explain the phenotypic variance in complex traits, we integrated genomic and exposomic data from PEGS and modeled the phenotypic variance as explained by the genome (G), exposome (E) and their interaction (GxE). We used genome-wide complex trait analysis (GCTA) to partition the complex trait variation for type 2 diabetes, height and body mass index (BMI) in PEGS to estimate the trait variance explained by variance in the genome, exposome and their interaction. Analogous to genomic relationship matrices (GRMs), we computed environmental kinship matrices using the extensive exposure data from the PEGS surveys. We show that the exposome explains and contributes to a significant amount of the phenotypic variation through its additive and interaction effects for both BMI and type 2 diabetes, while the phenotypic variation in height is mostly explained by variation in the genome. For type 2 diabetes, 50.9% of the phenotypic variance is explained by the genome, 2.8% by the exposome and 18.3% by their interaction. For BMI, the genome, exposome and their interaction explain 18.9%, 22.2% and 11.4% of the phenotypic variation respectively. These results show the importance of the exposome in dissecting phenotypic variance to understand the etiology of complex human traits. For future work, we will expand this variance decomposition analysis for additional phenotypes in PEGS.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4241 Killer cell immunoglobulin-like receptor (KIR) genetic variation in the natural history of type 1 diabetes

Authors:

A. Romero[1], L. Leaton[1], S. Onengut-Gumuscu[2], K. Waugh[3], S. D. Slack[1], N. R. Pollock[1], M. Rewers[3], J. Norris[4], P. Norman[1], R. K. Johnson[1]; [1]Univ. of Colorado Anschutz Med. Campus, Aurora, CO, [2]Univ. of Virginia, Charlottesville, VA, [3]Barbara Davis Ctr. for Diabetes, Aurora, CO, [4]Colorado Sch. of Publ. Hlth., Aurora, CO

HLA genetic variation confers the strongest risk for type 1 diabetes (T1D); however, this risk differs by stage of disease (initiation of preclinical islet autoimmunity, IA, versus progression to T1D) and the underlying causal mechanisms remain unclear. KIR physically interacts with HLA molecules to modulate innate and adaptive immune responses. Studies investigating *KIR* in T1D risk have been restricted to the T1D endpoint. We investigated the association of *KIR* genetic variation and the combination of *KIR-HLA* variation in the risk of T1D, IA, and the progression from IA to T1D in the prospective Diabetes Autoimmunity Study in the Young (DAISY) cohort. DAISY follows 2,547 high-risk children for the development of IA and T1D. Genome-wide SNP data were generated using a customized Illumina exome array for subjects selected in 2015 for multi-omics nested case-control study, including 89 T1D cases and 88 T1D controls, and 207 IA cases and 203 IA controls. By 2023, 234 IA cases had developed, of which 112 (48%) progressed to T1D at an average of 5.5 years post-IA. We imputed *KIR* with KIR*IMP and *HLA* with the Michigan HLA multiethnic reference panel. Genes were categorized into present vs. absent for analysis. We tested *KIR* genes and *KIR-HLA* combinations for association with IA and T1D risk using logistic regression adjusted for the high-risk HLA-DR3/4 haplogenotype and genetic PC1. Associations for T1D progression were tested using Cox regression with additional adjustment for age at IA seroconversion. KIR imputation accuracy ranged from 88% to 99%. No *KIR* genes or *KIR-HLA* combinations were associated with IA or T1D risk. After Bonferroni correction, six *KIR* genes were protective for T1D progression (P<0.00192), including *KIR2DS5* (OR=0.49, 95%CI: 0.31-0.76), *KIR2DS1* (OR=0.56, 95%CI: 0.39-0.81), *KIR3DS1* (OR=0.56, 95%CI: 0.39-0.81), *KIR2DP1* (OR=0.58, 95%CI: 0.44-0.76), *KIR2DL5* (OR=0.59, 95%CI: 0.43-0.79), and *KIR2DL1* (OR=0.60, 95%CI: 0.45-0.79). Two *KIR* genes increased risk of T1D progression, including *KIR2DS4* (OR=1.78, 95%CI: 1.24-2.53), and *KIR3DL1* (OR=1.78, 95%CI: 1.24-2.53). The effect of *KIR2DL1* on T1D progression differed by its *HLA-C2* ligand (interaction P=0.031). In the absence of *HLA-C2*, *KIR2DL1* was protective for the progression to T1D (OR=0.43, 95%CI: 0. 28-0.64), but was not associated with T1D progression in the presence of *HLA-C2*. Genetic variation of both inhibiting and activating *KIR* genes affect the risk of progressing from IA to T1D among children at high genetic risk. These effects differ by presence of HLA ligands and by stage of disease progression. Further elucidation of *KIR-HLA* allelic variation in T1D risk is warranted.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4242 Laipy: artificial intelligence for next-generation robust local ancestry inference.

Authors:

A. Jaras[1], D. Mas Montserrat[1], C. Lopez[1], B. Oriol Sabat[2], A. Ioannidis[1]; [1]Stanford Univ., Stanford, CA, [2]Univ. of California, Los Angeles, CA

The decreasing cost of sequencing technologies is leading to the growth of biobanks composed of numerous high-resolution whole genome sequences, which combined with new algorithmic techniques, are enabling personalized data-driven precision medicine. However, this new paradigm is benefiting mostly European-descent individuals, due to a large imbalance of available data, which includes mostly European populations, and due to the difficulty of processing genomic data from individuals of admixed populations with current predictive models. Local ancestry inference (LAI) methods, also referred to as ancestry deconvolution, provide high-resolution ancestry classification predictions, assigning an ancestry label to every genomic position or at windowed regions of the genome. These techniques are used in admixture mapping studies, historical and migration studies, genome-wide association studies, polygenic risk scores, and other predictive methods. Recent LAI methods including G-Nomix, LAI-Net and SALAI-Net, provide fast and accurate predictions, however, they are not built for easy use in clinical genetics by non-specialists. In this work, we introduce “laipy” a python framework that combines, extends, and unifies all 3 methods, while providing algorithmic and architecture improvements, better input/output and visualization capabilities, an easy-to-use phasing error correction system, and quality control methods based on uncertainty quantification techniques. The framework is easy to install, allows to interchange the updated versions of G-Nomix, LAI-Net and SALAI-Net seamlessly, and provides state-of-the-art speeds and accuracies. We run our framework within the whole UK Biobank and include a detailed benchmark with different number of populations, admixture times, and genome sequence resolution demonstrating the accuracy and effectiveness of the introduced framework.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4243 Large gene prioritization benchmark enables supervised learning of effector genes from GWAS loci

Authors:

C. Cole[1], P. Mohammadi-Shemirani[2], T. Lau[1], D. Merico[3], T. Masud[1]; [1]Deep Genomics, Toronto, ON, Canada, [2]Deep Genomics Inc., Toronto, ON, Canada, [3]Vevo Therapeutics, San Francisco, CA

Correctly prioritizing causal genes which mediate the effect of genome-wide association (GWA) loci on a phenotype remains a bottleneck for efficient target identification from complex trait association studies. Biases in the selection of genes used to benchmark existing approaches limit their generalizability between traits with different genetic architectures. In addition, many new methods have been proposed which claim superior accuracy over and above the nearest gene but their relative strengths have not been compared or combined systematically. In this study, we harmonize three large locus-to-gene (L2G) datasets and compare seven existing methods for gene prioritization before creating a supervised learning model to combine functional annotations with existing gene prioritization strategies. First, we create a high quality set of fine-mapped variants linked to effector genes following three different methods from Weeks et al 2021 (n = 715 pairs), Mountjoy et al 2021 (n = 433 pairs), and Forgetta et al 2022 (n = 415 pairs) using imputed UK Biobank summary statistics to perform functional fine-mapping through PolyFun with the BaselineLF-v.2.2 to estimate priors and SuSIE+FINEMAP to estimate credible sets. Crucially, by balancing the biases implied with these three different methods, we create a more representative set of L2G pairs with respect to complex trait architectures than any dataset alone. We used this set of 1564 L2G pairs to compare the precision and recall of seven individual gene prioritization strategies (distance by three metrics, polygenic priority scores (PoPS), robustness estimates for PoPS, combined SNP to Gene (cS2G), transcriptome-wide association studies, eQTL colocalization, and summary-based Mendelian randomization through HEIDI) and find that no method individually exceeds the performance of genomic distance. Using our harmonized benchmark, we create a supervised combination of genomic features, functional annotations, and existing gene prioritization approaches that outperforms any strategy individually on a held-out test set (AUROC = 0.7446 versus 0.6686 for best performing strategy of distance by any genic feature) or simple combination of features (PoPS + distance AUROC = 0.7192). These results demonstrate that combining strategies lends significant power over any individual approach for identifying effector genes. Moreover, as the community develops additional strategies, it will be critical to build representative datasets that serve to systematically compare gene prioritization strategies and better capture their generalizability to complex trait biology.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4244 Large-scale admixture mapping unravels novel genetic insights into human disease.

Authors:

M-C. López Cano[1], S. Moreno Grau[1,2], D. Mas Montserrat[1], J. Abante[1,3,4], G. Venkataraman[1], M. Rivas[1], A. Ioannidis[1,5]; [1]Dept. of BioMed. Data Sci., Stanford Univ. Sch. of Med., Stanford, CA, [2]Galatea Bio, Inc., Hialeah, FL, [3]Dept. of BioMed. Sci., Facultat de Med. i Ciències de la Salut, Barcelona, Spain, [4]Dept. of Mathematics and Computer Sci., Univ.t de Barcelona, Barcelona, Spain, [5]Inst. for Computational and Mathematical Engineering, Stanford, CA

The vast majority of genetic studies are conducted in samples from European-descent populations, missing out on the full diversity of genetic variation and disease risk explainability. Admixture mapping - an association analysis between local ancestry inference (LAI) results and phenotypes of interest - proves to be a powerful alternative to common methodologies like Genome Wide Association Studies (GWAS) in diverse and admixed populations with smaller sample sizes. In this work we provide a novel software package that allows researchers to easily conduct admixture mapping analysis and present a large-scale admixture mapping study on a set of unrelated individuals from the UK Biobank, revealing new population-specific associations.

Local and global ancestry inference is estimated using RFMix with African, African Hunter Gatherer, East Asian, South Asian, Oceanian (Australo-Papuan), European, West Asian, and Indigenous American population groups. We focus on 109 phecodes of medical interest, ICD10, algorithmically defined outcomes, cancer registry and family history. The association analysis is carried out per ancestry and condition, and multiple hypothesis testing correction is applied.

We detect multiple genetic loci significantly associated with numerous human diseases among various populations. While some of these regions have already been reported (i.e., well-known regions for seborrhoeic keratosis and lipidemias), many include new findings pertaining to breast cancer (BC), menopausal disorders, skin disease, and heart attack. We specifically highlight a protective association between Indigenous American ancestry and BC in chromosomal region 8p22. Our finding is in agreement with the low incidence of BC in Latinas with respect to others, and supports the rationale of protective variants acting in Latinas. Overall, our approach identifies novel genomic regions associated with clinically relevant phenotypes, and deepens the understanding of previously identified associations.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4245 Large-scale GWAS of strabismus identifies several novel risk loci and provides support for a link between maternal smoking and childhood eye disease

Authors:

S. MacGregor[1], W. He[1], P. van der Most[2], J. Ong[3], L-D. Hwang[4], L. Mauring[5], K. ÕUnap[6], B. Barry[7], E. Abner[6], N. Martin[8], D. Plotnikov[9], Y. Wu[1], P. Gharahkhani[1], H. Snieder[2], T. Palumaa[6], J. Guggenheim[10], D. Mackey[11], E. Engle[12], H. Choquet[13]; [1]QIMR Berghofer Med. Res. Inst., Brisbane, Australia, [2]Univ. of Groningen, Groningen, Netherlands, [3]QIMR Berghofer Med. Res. Inst., Herston, Australia, [4]The Univ. of Queensland, St Lucia, Australia, [5]Univ. of Tartu, Tarty, Estonia, [6]Univ. of Tartu, Tartu, Estonia, [7]Boston Children's Hosp., Boston, MA, [8]Queensland Inst Med Res, Brisbane, Australia, [9]Kazan State Med. Univ., Kazan, Russian Federation, [10]Cardiff Univ., Cardiff, United Kingdom, [11]Lions Eye Inst., Nedlands, Australia, [12]Boston Children s Hosp., Boston, MA, [13]Kaiser Permanente - Div. of Res., Oakland, CA

Strabismus (misalignment of the eyes) is a common pediatric ophthalmic condition with genetic and non-genetic risk factors contributing to its etiology. There are many manifestations of strabismus, including esotropia (exotropia) where there is an inward (outward) turn of one or both eyes. Due to the heterogeneity of strabismus and sample size limitations, previous genome-wide association studies encountered difficulties in identifying strabismus risk variants. Here, we conducted large-scale meta-analyses using 11 strabismus GWAS of European ancestry from 7 cohorts to discover new genetic variants for strabismus and its sub-types. We identified 4 genome-wide significant independent loci (*NPLOC4-TSPAN10-PDE6G-FAAP100*, *COL6A1*, *ZNF701* and *CHRNA3*) in the meta-analysis of broadly defined strabismus (20,464 cases and 954,921 controls). We also identified 5 (*UTS2*, *CHRNA3*, *DYNLRB2*, *NPLOC4-TSPAN10-PDE6G-FAAP100* and *MAD1L1*) independent loci by narrowing down the strabismus definition to esotropia (5,963 cases and 588,794 controls) and exotropia (3,998 cases and 583,468 controls). Some of these have been previously associated with ocular traits such as refractive error (e.g. *TSPAN10*, *MAD1L1*) and myopia (e.g. *COL6A1*) or cigarette smoking behaviour (*CHRNA3*). We also performed a cross-tissue transcriptome-wide association study, implicating 7 genes for esotropia (*ADAMTS7*, *C17orf70*, *CACNA2D2*, *COL6A2*, *PDE6G*, *PURG*, *SLC16A3*). In summary, we successfully replicated the previously reported strabismus locus near *NPLOC4-TSPAN10-PDE6G-FAAP100* and found a further 6 novel loci, providing new insights into strabismus biology. Finally, since previous observational studies have reported an association between maternal smoking during pregnancy and strabismus in offspring, here we conducted Mendelian randomization to help assess if there is a causal relationship; our genetic data provided evidence supportive of causality.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4246 LDPredx is a scalable, theoretically-justified method for cross-ancestry polygenic risk prediction

Authors:

M. Tudball[1], T. Lu[1], S. Bhatnagar[1], V. Forgetta[1], B. Richards[2]; [1]5 Prime Sci., Montreal, QC, Canada, [2]McGill Univ., Montreal, QC, Canada

Current polygenic risk scores (PRS) do not perform well in non-European ancestries, greatly limiting their clinical utility. This has spurred the development of new cross-ancestry methods which estimate PRS in one ancestry using data from multiple ancestries. However, many of these sophisticated methods perform similarly to, or even underperform, a simple linear combination of single-ancestry PRS (a "meta-score").
We make two contributions to this problem. Firstly, we propose a new method based on GWAS summary statistics called LDPredx, which is a strict generalisation of single-ancestry LDPred-inf to multiple ancestries. LDPredx is based on the cross-ancestry infinitesimal model. We show that this established population genetics model gives the LDPredx posterior an exact closed form, which provides computational tractability and generalisability to many ancestries, and sidesteps the need for external training data. We show in UK Biobank that LDPredx exhibits favourable performance and dramatically reduced computation time against other cross-ancestry methods. Our software is compatible with Hail 0.2, thereby facilitating scalable computing and improved access to data resources.
Secondly, we utilise LDPredx's foundation in population genetics theory to provide the first formal, mathematical description of the factors driving cross-ancestry portability. We first derive a formula for the LDPredx posterior, showing that it depends on the number of markers in the PRS, cross-ancestry genetic correlation and ancestry-specific trait heritabilities, sample sizes and linkage disequilibrium structure. We use this formula to probe the conditions under which cross-ancestry PRS will be the most effective. Furthermore, we shed light on the meta-score's surprising effectiveness by proving that it is the best linear approximation to LDPredx. When there is no linkage disequilibrium among markers in the PRS, we prove that LDPredx and the meta-score are, in fact, equivalent. Together, these results prove that LDPredx will outperform the meta-score whenever the markers in the PRS are linked. We verify all of our mathematical results using simulations and real data in UK Biobank.
LDPredx is a powerful tool both for polygenic risk prediction and illuminating the theoretical underpinnings of cross-ancestry portability.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4247 Learning portable polygenic risk score models with mixtures of pre-trained experts to improve accuracy across the continuum of ancestry

Authors:

S. Zabad, S. Gravel, Y. Li; McGill Univ., Montreal, QC, Canada

Polygenic risk scores (PRS) provide an estimate of the additive genetic component of complex traits and diseases. In recent years, there has been growing interest in incorporating polygenic scores into clinical practice, drug development pipelines, and personalized medicine. A large and fast-growing number of PRS estimates have been generated from diverse cohorts and made publicly available. One challenge for realizing the potential of current PRS models, however, is their poor portability to out-of-sample individuals and well-documented disparity in prediction accuracy across ancestries. To address this problem, recent PRS methods devised approaches to infer ancestry-specific variant effect size estimates, coupled with joint priors that encourage concordance. While this class of methods have been shown to significantly improve prediction accuracy for underrepresented populations, they do not completely bridge the gap. Moreover, cross-ancestry PRS methods operate on the assumption of discrete and uniform populations, which tends to work poorly for small or admixed subpopulations and does not capture the full continuum of ancestry. This problem is further compounded by the fact that recent analyses have shown limited PRS portability across other cohort variables, such as age, sex, and socioeconomic status.

To mitigate the PRS portability problem, we propose a Mixture-of-Experts (MoE) modeling framework, which can accommodate heterogeneity of effect sizes and automatically specialize individual experts on various partitions of the data. Specifically, MoEs consist of an ensemble of K PRS models whose outputs are combined on a per-sample basis using a flexible “gating” model. The gating model takes as input a list of covariates for each sample, such as Principal Components, age, and sex, and outputs probabilistic weights for combining the predictions of the PRS models in the ensemble.

To showcase the utility of this framework, we illustrate how it can be used to efficiently combine predictions from pre-trained PRS models from the PGS Catalog. In a 5-fold cross-validation analysis in the UK Biobank and CARTaGENE datasets, we show that MoEs significantly improve prediction accuracy over the best individual model, with mean improvements of up to 12% in underrepresented populations and admixed samples. It also produces scores that perform more consistently across a wide range of sociodemographic profiles. These benefits are achieved without imposing any assumptions or arbitrary subdivisions on the data at training time, which allows the model to automatically detect sources of heterogeneity and deploy the right model for each individual.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4248 Learning the environmental architecture of cardiometabolic disease at biobank scale.

Authors:

S. Tangirala, C. Patel; Harvard Med. Sch., Boston, MA

A key challenge in capturing the architecture of complex traits involves decoupling genetic from nongenetic variation. In nongenetic or environmental research, the scale in identifying and reproducing associations and predictions has not been achieved relative to genetic investigations. Biobank-scale analyses used to produce robust genetic findings can be repurposed to discover exposure-disease connections. We conducted a Exposure-Wide Association Study (ExWAS) to systematically identify associations across 362 nongenetic and behavioral factors for each of the 3 binary disease phenotypes (e.g., type 2 diabetes, hypertension, coronary artery disease [CAD]) using a sample of 472,240 White European (WE) participants from the UK Biobank (UKB). Furthermore, we highlight different ways of triangulating the identified associations. Genetics can help identify factors at scale. Thus, we computed causal estimates between each non-genetic factor-disease phenotype pair bi-directionally to also account for potential reverse causality (utilizing bi-directional Two-Sample Mendelian Randomization [MR] in FinnGen and UKB cohorts). We also perform ExWAS on biological mediators of complex traits such as BMI, HbA1c, Glucose, and 249 cardiometabolic biomarkers (e.g., lipoprotein lipids, amino acids, glycolysis metabolites, etc.). Among the FDR < 0.05 significant associations (341 unique exposures) identified for biological mediators, we find that the interquartile range of the absolute value of beta estimates is [0.0237, 0.074] with a mean of 0.0587. For CAD, we identified 172 FDR < 0.05 significant non-genetic factors (IQR of odds ratios (OR): [0.71,1.51], mean OR: 1.11), for hypertension we identified 248 ([0.71, 1.34], 1.07), and for type 2 diabetes we identified 224 ([0.85,1.27], 1.12). Overall, among the FDR-significant non-genetic factor associations, we observed concordance in direction between associations computed from the observational analysis and MR analysis, however only for a subset of the significant factors did we find evidence for potential causal relationship(s). For example, we found having a college or university degree with hypertension (observational OR = 0.75, observational beta-estimate = -0.287, observational FDR = 7.61×10^{-121}, MR beta estimate = -1, MR p-value = 1.71×10^{-18}) to have a more significant protective effect than mainstays such as never smoking (observational OR = 0.81, observational beta-estimate = -0.21, observational FDR = 4.19×10^{-90}, MR beta estimate = -0.42, MR p-value = 0.016). We present an integrative ExWAS and MR approach for systematically prioritizing nongenetic associations at biobank scale.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4249 Leveraging AI to improve variant prioritization algorithms and scale whole genome analysis

Authors:

S. White[1], T. Bowers[2], S. Hofherr[1], E. Frise[1], M. Reese[1]; [1]Fabric Genomics, Oakland, CA, [2]Broad Inst., Boston, MA

Over the past eighteen months, the genomics industry introduced genome sequencing with interpretation as low as USD $1,000. As genome prices continue to decline, the healthcare and insurance industry will incorporate sequencing into their standard practices of care. Labs are increasingly moving towards genome backbone sequencing and a “sequence once, interpret many times” paradigm. While sequencing costs have declined, the interpretation costs remain high due to the need for individuals highly skilled in clinical genetics to manually interpret and review each case. The expert interpretation component is more costly and takes longer than the sequencing itself. It can take a highly trained clinical genetics professional 20 minutes to 4 hours to review evidence and classify a single variant. In order to scale clinical diagnostics, labs must utilize automation and AI software as an adjunct to their clinical interpretation services. Fabric Genomics has pioneered AI-driven scalable genomic interpretation with a suite of ground-breaking proprietary and patented algorithms. Fabric’s newest software release, FE3, leverages state of the art AI including variant prioritization algorithms and automated ACMG variant classification, resulting in fewer variants to manually review and faster curation and reporting of those variants. The platform enables multiple concurrent or sequential independent analyses be performed on a genome backbone. The latest software supports build 38, MANE transcript, and additional annotations. New workflows and interfaces include auto approval for negative panel cases, upfront display of the number of classified variants for panel cases, automated ACMG classification for WES/WGS, and mitochondrial analysis. These advancements in AI algorithms will provide genomics labs better options for scalability and the opportunity to fully realize the possibilities and benefits of whole genome sequencing as a backbone.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4250 Leveraging and Partitioning of Polygenic Risk Scores Identifies Proteomic Networks Underlying Cancer Risks

Authors:

D. Dutta[1], J. Zhang[2], J. Coresh[3], A. Battle[4], E. Platz[4], N. Chatterjee[4]; [1]Natl. Cancer Inst., Rockville, MD, [2]Johns Hopkins Bloomberg Sch. of Publ. Hlth., Baltimore, MD, [3]Johns Hopkins Univ, Baltimore, MD, [4]Johns Hopkins Univ., Baltimore, MD

Background: Genome-wide association studies (GWAS) have identified numerous common susceptibility variants across various cancers. Understanding the impact of these variants on intermediate molecular phenotypes, such as gene expression and protein levels, can help prioritize disease-relevant target genes and proteins, and facilitate partitioning of polygenic scores (PGS) according to distinct networks reflecting converging effects of underlying SNPs. **Methods**: We investigated the associations of plasma protein levels with cancer-related SNPs for 21 cancers, curated from the PGScatalog, with levels of 4,955 plasma measured by SomaScan, for 9,083 participants (7,213 European Americans and 1,871 African Americans) at visit 2 in the Atherosclerosis Risk in Communities (ARIC) study. We adopt two complementary strategies to detect convergence of multiple association signals of cancer-related variants with plasma protein levels: (A) pQTS: association of the PGS for a particular cancer with plasma protein levels (B) ARCHIE: a sparse canonical correlation-based method to identify subsets or partitions of cancer related variants, constituting the PRS, associated to selected set(s) of proteins. **Results**: Across all cancers, we identified 90 associated proteins using pQTS of which 53 were guided purely by trans-associations between cancer-related SNPs and proteins. By partitioning the cancer-related SNPs, ARCHIE identified 19 significantly associated protein networks encompassing a broader set of 245 proteins and replicating a substantial fraction of trans-associations identified by pQTS (71.7%). We found that the proteins identified by pQTS and/or ARCHIE are enriched for relevant biological processes and cell types as well as cancer drivers and have concurrent somatic evidence of being associated with the respective cancers. For example, using the SNPs associated to cervical carcinoma, we identify a set of 80 proteins, primarily enriched in various immune response pathways and inflammatory responses. Additionally, we identify proteins that are associated with multiple related cancers indicating potential pleiotropic activity. **Conclusion**: Our analysis leverages known GWAS associations for cancers to identify protein networks underlying cancer risks and accordingly partition or subset polygenic scores into mechanistic components. As detailed molecular data of relevant tissues, cell-types and developmental stage becomes increasingly available, similar approaches will prove to be important tools for identifying downstream molecular targets for GWAS variants and improving interpretation and research application of PGS.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4251 Leveraging functional annotations to identify interactions between rare-variant gene sets and environmental factors

Authors:

S. McCabe[1], Z. Li[1], X. Li[2], H. Zhou[3], D. Christiani[4], X. Lin[5]; [1]Indiana Univ. Sch. of Med., Indianapolis, IN, [2]Harvard TH Chan Sch. of Publ. Hlth., Boston, MA, [3]Harvard, Boston, MA, [4]Harvard Med. Sch., Boston, MA, [5]Harvard T.H. Chan Sch Pub Hlth, Boston, MA

Analyses of gene environment interactions can provide insights into how the association of environmental risk factors, such as smoking or the concentration of air pollutants, on health-related outcomes can vary based on the genetic profiles in an individual. While analyses of common variants in this setting have been studied extensively, rare variants have been shown to explain a larger proportion of heritability and are often more deleterious. However, identification of interactions between rare-variant gene sets and environmental factors is more challenging. Though the development of large biobank datasets can provide hundreds of thousands of samples which can help in the detection of rare-variant effects and improve statistical power, these analyses are often still underpowered due to the rarity of the variants. We propose STAAR-GEI, a method that incorporates multi-faceted variant functional annotations to improve the statistical power of testing interactions between rare-variant gene sets and environmental factors. STAAR-GEI constructs an omnibus test that combines a SKAT, Burden, and ACAT framework using dynamic multiple variant functional annotation weights. STAAR-GEI also uses functional annotations to define the rare-variant sets to be tested and can efficiently handle population structure and sample relatedness through a generalized linear mixed model (GLMM) framework. In simulations, we found incorporating functional annotations with STAAR-GEI improved statistical power of testing rare-variant gene set by environment interactions compared to the existing variant-environment interaction tests. We apply this method to an analysis of the UK Biobank dataset to identify rare-variant gene sets which are interacting with several environmental factors on diseases and traits.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4252 † Leveraging genetic correlation and multiple population datasets to improve genetic prediction across populations

Authors:

L. Xu, G. Zhou, G. Leying, H. Zhao; Yale Univ., New Haven, CT

The disparity in genetic risk prediction accuracy between European and non-European populations poses a critical challenge of health disparity. Previous studies have highlighted the necessity to include datasets from multiple populations, account for genetic correlations across different populations, and consider suitable shrinkage distribution for polygenic risk prediction models. However, no previous study has comprehensively integrated all these elements into a single predictive model. To bridge this gap, we introduce a new method, JointPRS, which simultaneously models multiple populations. JointPRS improves over the existing methods by leveraging datasets from diverse populations, instead of solely relying on auxiliary European groups. Moreover, our model incorporates an automated estimation of chromosome-wise cross-population genetic correlation, outperforming methods that do not use such information. To validate our method, we chose a commonly used method, PRS-CSx, for comparison. We first assessed the prediction performance of JointPRS through simulations. We employed the spike and slab distribution for three populations (European, EUR; East Asian, EAS; and African, AFR) with specific correlation as the true model, and then varied the percentage of causal SNPs (p) from small to large (p=0.005, p=0.05, p=0.5). The average relative improvement of JointPRS over PRS-CSx measured in R-squared percentage across different correlation strengths was 0.21% for EAS and 0.33% for AFR when p=0.005; 7.75% for EAS and 3.6% for AFR when p=0.05; and 17.9% for EAS and 7.76% for AFR when p=0.5, implying the effectiveness of our prediction model. We then applied JointPRS to predict six complex traits - height (HGT), body max index (BMI), HDL cholesterol, LDL cholesterol, total cholesterol (TC), and triglycerides (TG) - in the UK Biobank dataset. Compared to PRS-CSx, JointPRS significantly improved the prediction accuracy for the EAS and AFR populations. The average relative improvement of JointPRS over PRS-CSx measured in R-squared percentage is 10.7% in EAS and 16.9% in AFR for HGT; 23.1% in EAS and 23.8% in AFR for BMI; 9.65% in EAS and 5.18% in AFR for HDL; 14.8% in EAS and no improvement in AFR for LDL; 21.4% in EAS and 7.16% in AFR for TC; and 7.45% in EAS and 3.57% in AFR for TG. These results demonstrate that JointPRS can enhance genetic risk prediction results in non-European populations over PRS-CSx.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4253 Leveraging heritability and genetic correlation to maximize GWAS power in EHR-connected biobanks

Authors:

N. Tatonetti[1], M. Zietz[2]; [1]Cedars-Sinai Med. Ctr., Los Angeles, CA, [2]Columbia Univ., New York, NY

Complex diseases are a significant global health challenge, necessitating rigorous and expansive study into their genetics. Observational data, such as Electronic Health Records (EHR), offer numerous advantages in the study of complex disease genetics. These include their large scale, cost-effectiveness, information on many different conditions, and future scalability with the widespread adoption of EHRs. Observational data, however, are challenging for research due to noise and confounding. EHR data reflect factors including the healthcare process, and access to care, as well as broader societal effects like systemic biases. In this study, we introduce a novel phenotyping method designed to strengthen and purify the genetic signal in EHR-based genetics research. Our approach quickly refines a linear phenotype by optimizing its heritability and genetic correlation, iteratively becoming closer and closer to the genetic component of the complex trait of interest. We validate the effectiveness of this approach by demonstrating its ability to predict polygenic scores, to enhance the power of genome-wide association studies (GWAS), and to improve heritability estimates in data from the UK Biobank. Our phenotyping does not rely on individual-level data, requiring as inputs only population-matched estimates of heritability, genetic correlation, and phenotypic correlation. Once the phenotype has been trained, our approach can be used to estimate the individual-level genetic component of the trait of interest. This purified genetic signal can then be used for EHR studies concerning complex traits and diseases. We demonstrate the utility by showing systematic improvement in GWAS data when using noisy and biased EHR-based phenotypes. This advancement helps to bolster the adoption and utility of observational and EHR data in genetic studies, accelerating our understanding of complex diseases.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4254 Leveraging local ancestry and cross-ancestry genetic architecture to improve the genetic prediction of complex traits in admixed populations

Authors:

G. Zhou[1], I. Yolou[1], H. Zhao[2]; [1]Yale Univ., New Haven, CT, [2]Yale Univ. Sch. of Publ. Hlth., New Haven, CT

Polygenic risk score (PRS) has been increasingly popular due to its ability to identify individuals with high disease risk for more effective population screening, diagnosis, and monitoring. However, the broader application of PRS is hindered by the limited transferability of PRS developed in Europeans to non-European populations. While many statistical methods have been developed to improve the performance of PRS in non-EUR population, most of them focused on the discrete genetic ancestry clusters and did not consider the admixed individuals. Admixed individuals pose a unique challenge for calculation of PRS due to the complexity of local ancestry and cross-ancestry effect sizes. Here we present a statistical method called SDPRX_admix for calculating PRS in admixed individuals. SDPR_admix characterizes the joint distribution of the effect sizes of a variant with two ancestries to be both zero, ancestry specific or correlated. We compared the performance of SDPR_admix with the conventional method BSLMM which assumes no heterogeneity of effect sizes by local ancestry. Extensive simulation revealed that SDPR_admix outperformed BSLMM in scenarios assuming heterogeneity of effect sizes based on local ancestry, while achieving similar performance to BSLMM in the scenario assuming no such heterogeneity. We analyzed 5 complex traits in individuals of African and European admixed ancestries where we trained the PRS model from the PAGE study (N = 13K) and tested the performance in UK Biobank (N = 4K). SDPR_admix outperformed BSLMM in all 5 traits as measured by Pearson R2 (Height: 0.015 vs 0.006; BMI: 0.0021 vs 0.0015; White blood cell counts: 0.02 vs 0.0018; Platelet counts: 0.0026 vs 0.0012; Hemoglobin concentration 0.001 vs 0.0006), though the predictive performance was low due to the small sample size of the training data. Our results suggested the benefits of incorporating local ancestry and modeling cross-ancestry genetic architecture to calculate PRS for admixed individuals.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4255 Leveraging multimodal neuroimaging data to identify novel genetic pathways to Alzheimer's disease

Authors:

Y. Tian, J. Park; Univ. of Toronto, Toronto, ON, Canada

Although genome-wide association studies (GWASs) have identified a number of genetic risk factors for Alzheimer's disease (AD), GWASs do not provide a comprehensive understanding of how the genetically-regulated structural and functional brain pathways drive AD progression, which is essential for characterizing the genetic mechanism of the disease and the development of AD targeted therapeutics. There is substantial evidence in the study of brain disorders that brain magnetic resonance imaging (MRI) measures, including multiple MRI modalities, such as structural, functional, and diffusion MRI, are useful biomarkers for AD and are modulated by genetic factors. However, due to both the high dimensionality and complexity of multi-modal MRI data, e.g., correlations of Imaging Derived Phenotypes (IDP) within modalities, current methods are unable to fully utilize brain imaging data. We propose a three-step mediation method for analyzing pathways of gene-AD effects by incorporating high-dimensional and multimodal brain MRI measures as mediators in GWAS. First, to reduce high dimensionality of MRI we apply principal component analysis (PCA) to reduce MRI dimensionality. Next, we estimate the genetic effects on reduced-dimensional MRI features using regressions. Lastly, we test for the existence of causal genetically regulated effects of genes on AD with an adaptive association test. This method has been extended to rely exclusively on publicly available GWAS summary statistics accommodating data from diverse studies, and external reference panels. Simulation results validated the robust control of type I error rates for the proposed method, pertaining to both continuous and binary traits. We applied the proposed method to UK Biobank (UKB) and International Genomics of Alzheimer's Project (IGAP) data and identified 12 genes affecting AD through structural MRI, 33 genes via diffusion MRI, and 1 gene from functional MRI. The driving motivation behind this study was the vast amount of MRI data in the UK Biobank and the power loss accompanying the analysis of numerous individual IDPs with current methods. Consequently, we devised an association test to evaluate the genetically regulated effects of specific MRI modalities on AD progression. This approach facilitates the identification of novel genetic pathways to AD, offering considerable potential for disease understanding and therapeutic development.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4256 Linkreg: Inferring cis-regulatory element to gene links using diverse epigenomic annotations across cell types.

Authors:

Q. Zeng[1], G. Xiang[2], R. Hardison[1], X. Zhu[1], Q. Li[1]; [1]Pennsylvania State Univ., University Park, PA, [2]Camp4 Therapeutics, Boston, MA

Cis-regulatory elements (CREs) are crucial in regulating gene expression, but understanding the intricate links between CREs and their target genes on a genome-wide scale remains challenging. Current methods for linking candidate CREs (cCREs) to genes often rely on simple distance thresholds or consider only a limited set of histone marks (e.g., H3K27ac), overlooking the broad impact of diverse functional marks on gene regulation. Although the epigenetic regulatory potential (eRP) score integrates epigenomic annotations that are derived from a comprehensive set of epigenomic features in multiple cell types by a computational model (e.g., ChromHMM), it lacks statistical rigor and interpretation. To address these limitations, we introduce Linkreg, a novel Bayesian model that infers cCRE-gene links by leveraging gene expression levels across cell types and epigenomic annotations in a principled statistical framework. Linkreg is built upon the scientific foundations established in eRP score, which proposes that the regulatory effects of a cCRE on gene expression are determined solely by the cCRE's epigenomic state profile, and that the effects of multiple cCREs on a gene are additive. By simultaneously selecting relevant cCREs and estimating the regulatory effects associated with different epigenomic states, Linkreg provides a more accurate and interpretable characterization of the gene regulatory network. Linkreg takes gene expression levels, cCREs, and their corresponding epigenomic annotations in various cell types as input. By using a multinomial latent formulation, it generates a posterior probability to assess if a specific cCRE regulates a particular gene. A scalable variational EM algorithm has been developed for efficient estimation in large-scale genomic analyses. Through comprehensive simulation studies based on real data, we have demonstrated the superior performance of Linkreg compared to distance and eRP score methods, achieving substantially higher precision-recall AUC values (≥ 0.2 higher) and better control over the false discovery rate. When applied to the mouse VISION dataset comprising 12 cell types in hematopoiesis, including 13418 genes and more than 200000 cCREs, Linkreg effectively identifies experimentally validated cCREs. Moreover, in the human EpiMap dataset comprising 304 cell types in diverse tissue samples, including 18306 genes and more than 1 million cCREs, Linkreg successfully identifies thousands of cCRE-gene links validated by chromatin conformation experiments or CRISPR screening. A user-friendly Python package for Linkreg has been developed.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4257 Localized Multi-Trait Model: Predicting Disease Risks, Identifying Variant Associations, and Mapping Trait Networks

Authors:

C. Lee[1], A. Khan[1], S. Ma[2], I. Ionita-Laza[1]; [1]Columbia Univ., New York, NY, [2]Jiao Tong Univ., Shanghai, China

Despite the impactful success of traditional genomic analyses such as Genome-Wide Association Studies (GWAS) and Polygenic Scores (PGS), they present certain limitations. Notably, binary trait overlooks phenotypic heterogeneity, causing information loss and obstructing advances in personalized medicine. Additionally, the case-control imbalance hampers GWAS's efficacy. While logistic regression enhances the power of selection-biased samples, it could be better for risk prediction due to the potential misinterpretation of effect sizes. We suggest the Liability Threshold-based Phenotypic Integration (LTPI). Leveraging the liability threshold model and multi-trait phenotypes from Electronic Health Records (EHR), the LTPI aims to circumvent the challenges above. It could unveil shared genetic factors between seemingly unrelated diseases, opening avenues for universal treatment strategies. We focus on utilizing comprehensive disease histories in EHR, highly available data. We conducted an LTPI analysis using UK Biobank and eMERGE, leveraging global and regional genetic covariance estimates among traits. Initially, we assumed a homogeneous genetic structure at each locus and gathered disease risk estimates for chronic kidney disease (CKD) and coronary artery disease (CAD). The LTPI scores ($LTPI_{Global}$), aggregating EHR phenotypes, showed enhanced prediction accuracy compared to binary PheCode phenotypes. However, subsequent GWAS using $LTPI_{Global}$ exposed an increased False Positive Rate (FPR), suggesting potential bias in our initial assumption. Our GWAS analysis yielded three association patterns: 1) True Positives (TP) influenced by both target and non-targets. 2) TP, guided solely by the target. 3) False Positives (FP) steered only by non-targets. The model's robustness against FPR was noteworthy when no associations were present to any included traits. However, we discovered that all TPs could face bias when local genetic covariance is current within a subgroup. Our findings led to a new hypothesis that each locus has a distinct genetic structure, which can be explored using recent techniques providing local genetic covariance estimates. To validate our model, we conducted simulations assuming local covariance. The results demonstrated that LTPI could generate locus-specific risk ($LTPI_{Local}$) suitable for GWAS, with superior predictive power than $LTPI_{Global}$, and maintains robustness even in FP scenarios (3). By the time of the meeting, we aim to present GWAS results on kidney disease and anxiety disorder traits using $LTPI_{Local}$, underscoring the model's applicability and potential in pleiotropic association mapping.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4258 LOCATER: Enhancing GWAS with genealogical relationship inference

Authors:

X. Wang[1], R. Christ[1], L. Aslett[2], D. Steinsaltz[3], I. Hall[1]; [1]Yale Univ., New Haven, CT, [2]Durham Univ., Durham, United Kingdom, [3]Oxford Univ., Oxford, United Kingdom

Genome-wide association studies (GWASs) have advanced our understanding of the genetic basis of phenotypes, yet a significant portion of phenotypic variance remains unexplained by identified genetic variants. Recently there has been renewed interest in ancestry-based methods to help address this gap by capturing loci with associations driven by allelic heterogeneity and difficult-to-tag structural variants. To complement this class of methods, we have developed LOCATER, an association testing workflow for inferred ancestral structures. While LOCATER can be adapted to work with any ancestry inference engine, in our current work we run LOCATER on top of probabilistic ancestral structures inferred using the R package kalis, a highly optimized implementation of the Li and Stephens model. LOCATER combines quadratic form and sequential testing approaches aimed at combining signals across weak-effect variants.

In experiments with simulated multiethnic datasets, we demonstrate that LOCATER can rescue signals driven by multiple causal variants that are missed by single-marker tests. On real sequencing data from Finnish individuals, LOCATER successfully replicates known loci associated with multiple lipid traits. All genome-wide significant signals identified by single-marker tests were also detected by LOCATER. In this pilot analysis, LOCATER demonstrated a stronger association signal in cases involving multiple unlinked causal variants. We further present work from ongoing analyses of a new multi-ancestry cohort, aiming to identify novel signals related to lipid traits and early-onset coronary artery disease.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4259 Longitudinal dynamics of lysophosphatidylcholines in relation to mortality, aging and *APOE*: Findings from Long Life Family Study.

Authors:

K. Arbeev[1], O. Bagley[1], A. Kulminski[1], E. Stallard[1], S. Cosentino[2], A. Yashin[1], S. Ukraintseva[1]; [1]Duke Univ., Durham, NC, [2]Columbia Univ. Med. Ctr., New York, NY

Recent studies found associations of lower levels of lysophosphatidylcholines (LPC) with adverse health outcomes and increased mortality risk. However, longitudinal dynamics of LPC in relation to mortality and aging-related outcomes as well as genetic underpinnings of such relationships remain understudied. Such analyses require specific methodologies for joint analyses of longitudinal and time-to-event outcomes that could also allow embedding different aging-related characteristics in the analytic framework. We applied such approaches, joint models (JM) and stochastic process models (SPM), to study longitudinal dynamics of LPC in relation to mortality and aging-related outcomes and the known genetic risk factor for those outcomes (*APOE4*) in the Long Life Family Study (LLFS). In total, we analyzed 23 LPC variants, with 3,614 measurements of each in 2,464 LLFS participants, and 1,330 deaths among the participants. JM analyses confirmed associations of lower levels of 18 LPC variants with increased mortality risk in the model adjusted for baseline age, sex, country, education, smoking and four genetic principal components (PCs), with the largest effect size observed for LPC 15:0/0:0 (hazard ratio for one-unit decrease=1.346, 95% CI (1.214, 1.493)). In the model adjusted for the *APOE* e4 status (carrier/non-carrier), we revealed association of the e4 allele with lower values of LPC 15:0/0:0 (regression coefficient: -0.114, p=0.015). SPM applications decomposed the observed associations of LPC 15:0/0:0 with mortality into several components representing different aging-related characteristics such as: 1) declines in robustness (p=2.1e-12) and resilience (p=0.003), related to deviations of the metabolite from optimal levels (those minimizing mortality risk at respective ages), 2) age-related decline in mean allostatic ("equilibrium") trajectories of the metabolite (p<1e-16), and 3) the persistent gap between the optimal and equilibrium trajectories that widens with age (p<1e-16). We also found that such characteristics are different in carriers and non-carriers of *APOE* e4 allele (p<1e-16). Various sensitivity analyses confirmed robustness of results to different aspects of computational procedures (e.g., analysis of a randomly selected person from a family; inclusion of additional covariates such as medication use and different numbers of PCs; additional quality control procedures). Results of our study support LPC as a major biomarker of human aging and related decline in robustness/resilience, and call for further exploration of genetic and other factors underlying age-dynamics of LPC in relation to mortality and diseases.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4260 Low Coverage Meta-Imputation Improves Imputation Accuracy for Ancient DNA.

Authors:

K. Kumar[1], S. Zollner[2]; [1]Univ. of Michigan, Ann Arbor, MI, [2]Univ Michigan, Ann Arbor, MI

Imputation is an essential, cost effective tool that increases power for downstream analyses such as GWAS or fine mapping. Traditionally, imputation has been developed for SNP array data. Low coverage sequencing is a growing alternative to SNP arrays that is both cost effective and unbiased for ancestry, which makes it better at detecting rare variants for non-European populations. New imputation algorithms such as GLIMPSE have been developed to impute low coverage samples. However, the imputation accuracy of GLIMPSE, like other algorithms, depends on the ancestry and sample size of the reference panel. For non-European ancestry populations, where we have fewer samples, low coverage imputation may benefit from a “meta” approach of combining results from smaller ancestry specific panels and larger European ancestry panels. Currently, methods for such meta imputation are designed for SNP arrays, and, therefore, do not take into account genotype uncertainty that is characteristic of low coverage samples or the difficulty in pre-phasing such samples.We design a new meta-imputation method that combines estimates from multiple reference panels using dynamically estimated weights tailored for each individual and genomic region. We use a diploid Li and Stephens Hidden Markov Model (HMM) to account for the difficulty of pre-phasing. Our method, unlike existing meta-imputation methods, accounts for genotype uncertainty by integrating over all possible genotypes in the emission probability of our HMM, weighted by genotype likelihoods. Our method outputs allelic dosages along with meta-imputed genotype dosages for each individual and marker.We meta-impute ancient DNA samples that are first imputed by GLIMPSE 2.0.1 using both African ancestry and European ancestry reference panels from 1000 Genomes. Our meta-imputed ancient African samples yield a higher R2 than imputing with either reference panel alone and performs comparably to “mega” imputation with a single combined African and European ancestry reference panel. Increasing the accuracy of imputation through meta imputation allows for better performance in downstream analyses, increasing the power for genomic discoveries, which will benefit non European populations.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4261 Low-coverage sequencing imputation with 150,119 UK Biobank reference samples

Authors:

O. Delaneau[1], S. Rubinacci[2], R. Hofmeister[3], B. Mota[1]; [1]Univ. of Lausanne, Lausanne-Dorigny, Switzerland, [2]Brigham and Women's Hosp., Boston, MA, [3]Univ. of Lausanne, Ecublens, Switzerland

Low-coverage whole genome sequencing (lc-WGS) followed by genotype imputation is a cost-effective technology that can enhance genome-wide analysis and access to personalized medicine. Recent advances show that a shift from SNP arrays to lc-WGS is beneficial, although with prohibitive computational costs for large reference panels. Therefore, sequencing projects like the recent UK Biobank (UKB) WGS, bring new possibilities, such as improved reference panels, but also new computational challenges.

We present a new lc-WGS imputation method, GLIMPSE2, designed to handle millions of reference samples with an accuracy inaccessible to other methods at sustainable costs. Using 150,119 UKB samples as a reference panel, we obtain a reduction of running time of 100-1000 times compared to existing methods, achieving imputation of a genome for less than 0.10$.

Lc-WGS greatly benefits from the UKB reference panel, mainly for very low-coverage data (0.1-0.5x) and at rare variants (<0.1% MAF). For 0.25x and 0.5x, our method imputes variants at 0.01% MAF with r2=0.8 and r2=0.89 respectively. As a comparison, imputation using the UKB Axiom array reaches an accuracy of r2=0.71. By looking at non-homozygous reference calls, we show that the process of imputation increases “in-silico” the coverage of lc-WGS samples by an order of 10x, resulting in a striking reduction of sequencing costs.

We evaluated the performance of the UKB reference panel against 1000 Genomes by imputing 276 genomes from the Simons Genome Diversity Project. As expected, the UKB reference panel substantially improves the imputation of European individuals, especially those with Northern European ancestry. As a proof of concept, we imputed four ancient Viking samples and show that UKB reference panel brings a significant boost of accuracy (for 0.1x, 1000G r2=0.83; UKB r2=0.92), implying that Viking ancestry is recovered remarkably well from the UK Biobank.

Furthermore, in order to quantify the impact of these results for disease association, we performed GWAS analyses using 10,000 UKB individuals across 100 quantitative traits, comparing lc-WGS and SNP array to high coverage data. We found that the UKB Axiom array is quantitatively similar to 0.25x data (beta r2=0.9, p val r2=0.87), and inferior to 1x data (beta r2=0.97, p val r2=0.95). Looking at precision and recall of the hits, 0.5x captures high coverage GWAS signals better than the UKB Axiom array, suggesting it is better suited for fine mapping.

Overall, we present the remarkable performance of lc-WGS imputation from the UKB reference panel with a competitive financial cost, that could lead to a boost of sample size and signal of future genomic studies.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4262 Machine learning based integration of large-scale priors for genetic fine-mapping, colocalization and in-silico functionalization

Authors:

K. Bhattacharyya[1,2], S. Bhattacharjee[2]; [1]Univ. of Calcutta, Kolkata, India, [2]Natl. Inst. of BioMed. Genomics, Kalyani, India

Identification of causal genetic variants together with their potential regulatory mechanism can offer the most critical insights into the proper understanding of the complex genetic etiology of complex traits. So far, genome wide association studies have been remarkably successful in identifying genetic variants associated with various complex traits. However, the interpretation of these GWAS variants in terms of causality and the perception regarding the role played by these variants in the genetic etiology of complex traits have remained significantly ambiguous due to three major reasons- A. complex LD structure among the associated variants making it almost impossible to separate false positives from true causal variants, B. majority of the associated variants are in noncoding region, C. nonavailability of large gold standard sets of regulatory variants for training and validation through Machine Learning models. Knowledge from annotations can help to overcome the low power of fine-mapping due to strong LD. This highlights the requirement of novel methods that can dynamically incorporate large-sets of priors based on available knowledge. Here, we propose a statistical method combining Machine Learning with existing Bayesian techniques to efficiently integrate large-scale functional annotations and epigenomic landscape in a tissue/cell-type specific manner. A unique feature of our method is its ability to not only infer causal variants, but also potential gene(s) through which their action is mediated along with potential regulatory mechanisms. The proposed systematic prioritization method focuses on utilization of Bayesian statistical methods for Colocalization (e.g. Coloc), fine-mapping (e.g., Sum of Single Effects Regression Model or SuSiE) and Machine Learning technique XGBoost. We apply the proposed method for integrative analysis of GWAS summary data, GTEx v8 data, LD structure, Functional Annotations for quantitative lipid traits in relevant tissues. The main utility of such prioritization is to facilitate interpretation of these quantitative lipid traits associated loci to draw inference regarding genetic causality of CVD. Systematic and efficient prioritization of causal variants in complex trait associated loci along with identification of regulated cis-Genes is a critical step in guiding wet-lab validation experiments and reducing the resource burden. Further, improved gene-prioritization will enhance our understanding of biological processes driving complex diseases and help to pinpoint potential drug targets.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4263 Machine learning-derived risk of atherosclerosis progression reveals novel genetic associations and empowers drug target discovery

Authors:

L. Leal[1], L. Chen[2], S. Hu[2], H. Climente-González[2], R. Nielsen[1], R. Gupta[1], R. R. Kitchen[1], J. M. M. Howson[2]; [1]Novo Nordisk Res. Ctr. Oxford, Oxford, United Kingdom, [2]Novo Nordisk Res. Ctr. Oxford, Human Genetics CoE, Oxford, United Kingdom

Atherosclerosis (ASCVD) progresses with age yet, typically, ASCVD diagnoses appear in electronic medical records (EHR) as discrete events such as presence/absence of coronary artery disease (CAD). Binary phenotypes such as these have lower statistical power than continuous phenotypes and do not reflect the underlying incremental biology of atherosclerosis progression. EHR provide an opportunity to model phenotypes longitudinally and identify novel genetic associations in GWAS, however there remain challenges that reduce performance of GWAS, specifically regarding underdiagnosis and missing data. We hypothesised re-encoding individuals on an atherosclerosis continuum would be beneficial to understand atherosclerotic biology and identify potential drug targets for atherosclerosis.

We introduce a reproducible pipeline for phenotype re-definition and present its implementation with ASCVD in UK-Biobank. Taking 51,056 cases and 325,949 controls, it trains a Machine Learning (ML) classification model integrating Polygenic risk scores (PRS) with clinical and life-style variables and computes a continuous progression phenotype of ASCVD. We generated interpretable ML models and found the CAD PRS was amongst the top-10 high impact features, and that a high number of individuals were predicted to have advanced ASCVD but had not yet had a CVD event. We then performed GWAS of the progression phenotypes, identifying 74 associated genetic regions, 33 of which were not in the binary GWAS. We replicated 35 (of 74) associations in the CAD GWAS by Aragam et al 2022 (181,522 CAD cases, 1,165,690 controls), despite our smaller cohort size. Taking ASCVD progression as an outcome, we implemented systematic Mendelian randomisation (MR) and genetic colocalisation using tissue-specific eQTLs and plasma pQTLs and identified 285 potential causal genes for ASCVD, which do not map to previously known CAD loci. Thirty-nine of these did not implicate major lipids. 12 genes were differentially expressed in aortas of atherosclerotic individuals with consistent effect direction to that from MR. One of these novel targets is ZNF197, a regulatory protein for which increased gene expression was found associated with reduced atherosclerosis.

By moving from a binary to a continuous phenotype using ML, we leveraged association signals in GWAS and identified potential causal genes for ASCVD. Comprehensive genetic analysis together with molecular data annotations allowed us to characterise non-lipid related genes with potential for novel mechanisms of action. Following the use case for ASCVD, this pipeline can now be implemented to other disease areas.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4264 Mapping sleep's phenotypic and genetic links to multiple organs: a systematic analysis of multi-organ images in the UK Biobank

Authors:

Z. Fan[1], Y. Li[2], J. Shu[2], X. Yang[2], B. Li[2], J. Lin[3], Q. Wang[4], P. Paschou[2], T. Li[5], H. Zhu[5], B. Zhao[6]; [1]Univ. of Pennsylvania, Philadelphia, PA, [2]Purdue Univ., West Lafayette, IN, [3]Yale Univ., New Haven, CT, [45]Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, [6]Univ. of Pennsylvania, Dept. of Statistics and Data Sci., Philadelphia, PA

Sleep is essential for human health, but a systematic analysis of the relationship between sleep and various organs, along with their genetic underpinnings, is lacking. Using imaging features of organ (brain, heart, and abdomen) structures and functions as clinical endophenotypes, we present a systematic genetic investigation of sleep and muti-organ connections from over 40,000 subjects in the UK Biobank. We identified novel phenotypic and genetic links between sleep and a wide range of imaging traits, such as brain structures, white matter integrity, brain activities, cardiac structures and functions, as well as abdominal compositions. We prioritized imaging modalities and traits for specific sleep conditions, such as the resting brain function measures in the somatomotor network with narcolepsy. Overlapping genetic influences were detected between sleep and multiple organs, some of which showed evidence of shared causal genetic variants. In conclusion, large-scale imaging genetic data illuminate the implications of sleep on whole-body health and their genetic links. An interactive web browser (www.ig4sleep.org) has been developed to facilitate exploring our results.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4265 Mendelian Randomization analysis of metabolites associated with severe obesity in the Hispanic Community Health Study/ Study of Latinos (HCHS/SOL)

Authors:

K. Young[1], V. Buchanan[2], M. Graff[1], M. Krishnan[3], H. Highland[4], B. Yu[5], C. Avery[1], S. Buyske[6], J. Cai[3], M. Daviglus[7], A. Howard[3], C. Isasi[8], R. Kaplan[8], R. Loos[9], Q. Qi[8], R. Rohde[3], J. Rotter[10], L. Van Horn[11], P. Gordon-Larsen[3], E. Boerwinkle[5], K. North[3]; [1]Univ North Carolina, Chapel Hill, NC, [2]Univ. of North Carolina - Chapel Hill, Westfield, IN, [3]Univ. of North Carolina, Chapel Hill, NC, [4]Univ North Carolina at Chapel Hill, Chapel Hill, NC, [5]Univ. of Texas Hlth.Sci. Ctr. at Houston, Houston, TX, [6]Rutgers Univ, Piscataway, NJ, [7]Univ. of Illinois, Chicago, IL, [8]Albert Einstein Coll. of Med., Bronx, NY, [9]Univ. of Copenhagen, Copenhagen, Denmark, [10]Lundquist Inst., Harbor-UCLA Med Ctr, Torrance, CA, [11]Northwestern Univ., Chicago, IL

Severe obesity (SevO: body mass index (BMI)≥40) is a major risk factor for heart disease and Type 2 Diabetes. Both disproportionately impact Hispanic/Latinos, 24.5% of whom are projected to have SevO by 2030. However, the underlying metabolic dysfunction characteristic of SevO remains unknown.

To assess if metabolomic associations with SevO are causal or consequential, we first identified metabolites associated with SevO in the Hispanic Community Health Study/Study of Latinos (HCHS/SOL). For the top 20 SevO associated metabolites, we extracted HCHS/SOL metabolite GWAS (mGWAS) summary statistics to derive metaboQTLs, and summary statistics from 1) a multi-population meta-analysis of SevO (N>70,000), and 2) UK Biobank (UKBB) to implement bidirectional Mendelian Randomization (MR).

Anthropometry and data for 640 metabolites were available for 551 females (27% SevO) and 371 males (15% SevO). In the meta-analysis, 224 metabolites were associated with SevO ($p<0.05/640$). Of the top 20 SevO metabolites, we excluded 8 with no mGWAS signals. For the remaining 12, we included independent SNPs ($p<5E-8$) as instrumental variables (IVs) in the forward MR. For the 59 independent SNPs in the SevO GWAS, we extracted summary statistics from the pan-ancestry UKBB for both measured BMI (BMI_m) and BMI calculated from bioimpedance (BMI_B) (35 SNPs) as IVs in the reverse MR.

In the forward MR, no metabolites were causally associated with SevO, though N-Acetylasparagine was significantly causally associated with both BMI phenotypes (ß (SE) =0.004 (0.001), BMI_m p =9.50E-6; BMI_B p=2.39E-4). The strongest genetic signal for this metabolite was rs10189885 (ß (SE): 0.981 (0.026), p=7.30E-308), an intronic variant in *ALMS1P1*, a pseudogene associated with Alström syndrome, a recessive condition whose phenotype includes hyperinsulinemia-associated childhood obesity.

In the reverse MR, BMI_m showed causal associations with four metabolites, including phenylalanine (Robust IVW ß (SE): 0.357 (0.115), p=1.9E-4) and 2-methylbutyrylcarnitine (Robust IVW ß (SE): 0.432 (0.0.118), p=2.62E-4). Causal estimates for both metabolites were robust to horizontal pleiotropy. Phenylalanine is an essential amino acid found in most protein food sources and has been positively associated with obesity as well as 2-year risk of hypertriglyceridemia in Hispanic/Latino children. 2-methylbutyrylcarnitine is a product of branched chain amino acid catabolism and a biomarker of meat and fish consumption that has been previously positively associated with obesity in Hispanic/Latino children. Our work is a first step towards disentangling the metabolic causes and consequences of SevO.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4266 Mendelian randomization approach suggests the potential of repurposing anti-diabetic drugs for the prevention of Parkinson's disease.

Authors:

J-Y. Lee[1], D. Park[2], J. Park[3], J. Sung[1]; [1]Seoul Natl. Univ., Seoul, Korea, Republic of, [2]Ajou Univ. Hosp., Suwon, Korea, Republic of, [3]Seoul Natl. Univ. Sch. of Publ. Hlth., Seoul, Korea, Republic of

Anti-diabetic drugs have shown some evidence suggesting a protective role in the prevention of Parkinson's disease (PD), however, the results were controversial. Mendelian randomization (MR) analysis using genetic instrumental variables associated with target genes of drugs has the potential to simulate the effect of the drug. Therefore, the objective of this study is to evaluate the potential repurposing of anti-diabetic medications for the prevention of PD. We identified the target genes of oral hypoglycemic drugs from Drugbank. Subsequently, we identified the SNPs associated with the expression of these genes using the expression quantitative trait loci (eQTL) data obtained from whole blood samples in the GTEx project. Utilizing these SNPs as instrumental variables, we performed an MR analysis to investigate the impact of drug target genes on the occurrence of PD. We obtained the GWAS data from the Diabetes Meta-Analysis of Trans-Ethnic association studies (DIAMANTE) Consortium for type 2 diabetes and from the International Parkinson's Disease Genomics Consortium for PD.
From the MR analysis, we revealed associations between several genes targeted by the anti-diabetic drugs and the incidence of PD. Notably, *GPD1*, inhibited by metformin, exhibited a significant increase in the risk of PD (p-value from MR analysis using Wald ratio method = 0.006). Additionally, *CPT1A* and *CFTR*, both inhibited by glyburide, showed tendencies towards a protective and harmful effect on PD, respectively, although these associations were not statistically significant (p-value from Wald ratio methods = 0.06 and 0.09).
Our findings suggest a potential role for anti-diabetic drugs in the prevention of PD. Inhibiting target gene expression with metformin or glyburide was associated with an increased risk of PD, which raises important considerations for using these drugs in individuals at higher risk of PD. Conversely, the potential protective effect of inhibiting *CPT1A* by glyburide, although not significant, indicates the possibility of exploring this drug as a potential medication option. In conclusion, this study highlights the potential involvement of specific target genes of anti-diabetic medications in the development of PD, providing enhanced opportunities for optimizing medication selection in diabetic patients at higher risk of developing PD.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4267 Mendelian randomization shows no causal effect of serum vitamin D levels on thyroid function.

Authors:

T. Zemunik, N. Pleic, I. Gunjača, M. Babić Leko; Univ. of Split, Sch. of Med., Split, Croatia

Background: Vitamin D plays an essential role in the normal functioning of many organs, including the thyroid gland. As a result, it is not surprising that insufficient levels of vitamin D are seen as a potential factor contributing to the emergence of several thyroid disorders, such as autoimmune thyroid disease and thyroid cancer. Nevertheless, the exact relationship between serum vitamin D levels and thyroid function remains incompletely understood. Here we use a Mendelian randomization approach (MR) to investigate the causal effect of serum vitamin D levels on the indicators of thyroid function; thyroid-stimulating hormone (TSH), free thyroxine (fT4), and thyroid peroxidase antibodies (TPOAb). **Methods:** We performed a two-sample MR study using summary statistics from the most comprehensive genome-wide association studies (GWAS) of serum vitamin D levels (25 hydroxyvitamin D (25(OH)D), n= 443,734), TSH (n=119,715), fT4 (n=49,269) and TPOAb (n=18,297). The multiplicative random-effects inverse variance weighted (IVW) method was chosen as the main analysis. Sensitivity analysis was performed using the weighted median and mode analysis, as well as MR-Egger, MR-PRESSO and Causal Analysis Using Summary Effect estimates (CAUSE). **Results:** Our analysis suggests that serum vitamin D levels are not causally associated with the determinants of thyroid function. Genetically predicted vitamin D levels were not causally associated with TSH (β=-0.02, P=0.512), nor with fT4 (β=0.05, P=0.229). The direction of the association between vitamin D and fT4 levels was consistent across different MR analyses, while this was not the case in the TSH analysis. The insignificant relationships were further confirmed by the CAUSE analysis for both TSH and fT4. Vitamin D levels showed a negative association with both TPOAb levels (β=-0.02, P=0.712), and TPOAb positivity (OR=0.57, P=0.141), however, these associations were not significant. **Conclusions:** This MR study does not support a causal association between genetically predicted serum vitamin D levels and the determinants of thyroid function. Further studies, particularly randomized controlled trials, are necessary to evaluate the therapeutic advantages of using vitamin D supplementation for the treatment of thyroid dysfunction.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4268 Meta-analysis of Alzheimer's disease GWAS identifies novel loci and implicates regulation of a hypoxic-responsive gene as a disease mechanism

Authors:

A. Diwadkar[1], J. Regan[1], X. Wang[2], H. Markus[2], L. Carrel[1], D. Liu[3]; [1]Penn State Univ., Hershey, PA, [2]Penn State Coll. of Med., Hershey, PA, [3]Penn State Univ., Hershey, PA

Alzheimer's disease (AD) is the most common form of dementia and is characterized by formation of amyloid-beta (Aβ) plaques and neurofibrillary tau tangles in the brain. AD is known for its phenotypic heterogeneity and sex differences where females tend to have an earlier age of onset, faster disease progression, and more severe disease outcomes. AD is strongly influenced by genetic risk factors, with an estimated heritability of 60 to 80%. To date, over 95 loci have been associated with AD risk. Yet, much of the disease heritability remains unexplained. There is a need to enlarge sample sizes, improve sample diversity, and better accommodate phenotype heterogeneity to improve power. To enhance the analyses of AD GWAS, we propose a novel meta-analysis approach IMAC (Integrated Meta-Analysis for Cross-trait GWAS) a mixed effect meta-regression model that uses factors of genetic correlation matrix as covariates to adjust for heterogeneity of AD phenotypes while accounting for sample overlaps. IMAC includes fixed effect, random effect analysis, and meta-regression as special cases. Unlike other cross-trait methods such as MTAG which assumes that genetic effects between phenotypes follow a multivariate normal distribution, our method is more flexible and can accommodate cross-trait genetic effect differences. For example, variants with similar effects between different AD subphenotypes would be most powerfully analyzed by a fixed effect method. Alternately, variants that differentially affect AD subphenotypes would benefit from models that incorporate factors as covariates to characterize genetic effect heterogeneity. We applied this method to perform a meta-analysis on 4 large publicly available AD and AD proxy GWAS summary statistics namely, Bellenguez et al. (2022), Wightman et al. (2021), de Rojas et al. (2021) and Jansen et al. (2019). We performed imputation on the meta-analysis results using FIZI to identify 30 known and 9 novel AD loci. Further, we applied statistical fine-mapping using susieR and utilized eQTL, transcriptomic, and epigenomic datasets in brain-related cell types to identify functional regulatory variants for experimental validation using a massively parallel reporter assay. The novel risk loci identified include a locus near the gene HIF-1α (hypoxia-inducible factor-1α), which is produced in both acute and chronic hypoxia. Hypoxia is a strong risk factor for AD. In AD models, HIF-1α can mediate hypoxic effects leading to neuronal apoptosis, increased levels of Aβ, and aberrant phosphorylation of tau proteins. Overall, our method provides novel risk loci for follow-up studies that may further elucidate the pathophysiology of AD.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4269 Meta-analysis of cardiac eQTLs improves identifying heart failure mechanisms

Authors:

S. Figarska[1], L. Zhang[1], S. B. Montgomery[2], J. R. Priest[1]; [1]Tenaya Therapeutics, South San Francisco, CA, [2]Stanford Univ., Stanford, CA

There are limited sources of eQTLs in the heart, more specifically the left ventricle (LV), impeding the study of the genetic and molecular mechanism of heart failure. One of the sources is GTEx that offers data from 386 LV samples. These samples were collected postmortem from mainly healthy individuals aged 20-71 yrs, 33% females, mostly European ancestry. A separate source is data from Myocardial Applied Genomics Network (MAGNet) which includes 313 LV freewall tissue samples harvested at the time of cardiac surgery from subjects with heart failure (HF) undergoing transplantation and from unused donor hearts. To expand insight into LV eQTLs and inform the mechanisms of heart failure risk loci, we reprocessed MAGNet data to make it more comparable with GTEx. Specifically, we imputed MAGNet genetic data using TOPMed, we calculated latent factors and re-called eQTLs with a MAF>0.01 threshold. For each set of eQTLs we subsequently tested colocalization with GWAS on HF in individuals of European descent, obtained from GBMI (Global Biobank Meta-Analysis Initiative). Colocalization was performed with coloc for each locus where there was a GWAS SNP p<1e-7 and within 250kb there was an eQTL SNP in either MAGNET or GTEx with a p<1e-4. We found colocalizations between ABO and SPATS2L eQTLs and HF GWAS, for both GTEx and MAGNet. Notably, colocalization between PTPN11, known as a mendelian cause of Noonan syndrome which includes cardiac manifestations, colocalized with HF only for GTEx data. On the other hand, colocalization between NPC1, a gene previously shown to be associated with risk of coronary heart disease, and HF, was observed only in MAGNet data sets. We have meta-analyzed both eQTL datasets to create the largest available set of eQTL data in left ventricle and report 10,706 genes with an eQTL (FDR 0.05). We expect these data to increase statistical power for finemapping and colocalization and shed more light on molecular mechanisms of genetic risk for heart failure.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4270 Meta-analysis of set-based multiple phenotype association test based on GWAS summary statistics from different cohorts

Authors:

L. Zhu; Michigan Technological Univ., Houghton, MI

Genome-wide association studies (GWAS) have emerged as popular tools for identifying SNPs that are associated with complex diseases. Standard analysis of a GWAS involves assessing the association between each individual SNP and a disease. However, this approach suffers from limited reproducibility and difficulties in detecting multi-SNP and pleiotropic effects. Although joint analysis of multiple correlated phenotypes for GWAS can identify and interpret pleiotropic loci which are essential to understand pleiotropy in diseases and complex traits, most of the multiple phenotype association tests are designed for a single SNP, resulting in much lower power when the number of SNPs increases and SNPs are correlated, especially when their effect sizes are small and only their cumulative effect is associated with multiple phenotypes. To overcome these limitations, set-based multiple phenotype association tests have been developed to enhances statistical power and facilitates the identification and interpretation of pleiotropic regions. Nevertheless, most existing methodologies fail to consider the impact of varying sample sizes and sample overlapping when combining GWAS summary statistics from different studies. In this research, we propose a new method, named Meta-TOW-S, which conducts joint association tests between multiple phenotypes and a set of SNPs (such as SNPs in a gene) utilizing GWAS summary statistics from different cohorts. Our approach applies set-based tests using the Optimal Weighted Combination of SNPs and accounts for sample size differences across studies by employing the Cauchy combination method. Meta-TOW-S combines the advantages of set-based tests and multi-phenotype modeling, exhibiting computational efficiency and enabling analysis across multiple correlated phenotypes, while accommodating overlapping samples from different GWAS cohorts. To assess the performance of Meta-TOW-S, we developed a phenotype simulator package that encompasses a comprehensive simulation scheme capable of modeling multiple phenotypes with multiple SNPs, including noise structures and diverse correlation patterns among phenotypes. Simulation studies validate that Meta-TOW-S maintains a desirable type-1 error rate and enhances power across various simulation scenarios. Furthermore, we demonstrate that Meta-TOW-S outperforms single phenotype-based methods by detecting a larger number of significant genes using real GWAS summary statistics obtained from multiple consortia.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4271 Metabolite prediction models in UK Biobank: A metabolome-wide association study (MWAS)

Authors:

L. Huang[1], A. L. Tapia[2], J. Chen[3], B. Rowland[3], J. D. Rosen[3], J. Wen[4], Y. Li[3,5,6], L. Raffield[5]; [1]Curriculum in Bioinformatics and Computational Biology, Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, [2]Dept. of Psychiatry, Univ. of Pittsburgh, Pittsburgh, PA, [3]Dept. of Biostatistics, Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, [4]Dept. of Genetics, Univ. of North Carolina at Chapel Hill, Chapel Hill, Chapel Hill, NC, [5]Dept. of Genetics, Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, [6]Dept. of Computer Sci., Univ. of North Carolina at Chapel Hill, Chapel Hill, NC

As potential disease risk factors, multi-omics readouts including transcripts, proteins, and metabolites are all of the interest. However, due to cost and tissue availability, sample sizes for omics data have severely lagged behind genetic data. To fill this gap, we use genetic and metabolomic data from UK Biobank (UKBB) to perform a metabolome-wide association study (MWAS). Our analyses included 101,349 European ancestry participants (EUR) (defined based on genetic similarity to 1000G) with 161 QC+ metabolites measured using the Nightingale platform. We identified metabolite quantitative trait loci (mQTL, p < 1e-6) with a subset of 45,581 individuals, built metabolite prediction elastic net models with these mQTLs in an independent subset of 45,466 individuals, and assessed the performance of the prediction models in the remaining 9,812 individuals. Our results showed that a number of metabolites can be reasonably predicted: median testing R^2 0.095, and 74 metabolites with testing R^2 > 0.1. We also applied our pre-trained prediction models to non-EUR ancestry UKB participants. Models show worse performance than in EUR participants (median testing R^2 0.053, 0.047, and 0.038, respectively for African, East Asian, and South Asian ancestry participants), similar to the transferability issue reported in TWAS or polygenic risk score studies. We proceeded with the 74 metabolites with testing R^2 > 0.1 and tested associations with 29 blood cell traits in the 333,501 EUR UKB samples without measured metabolites. We also tested the associations between measured metabolites and blood cell traits, using the 96,494 unrelated samples with both phenotypes. At a Bonferroni threshold of 2.4e-5, we had a median of 68 (range: 34-73) significant associations per blood cell trait using measured metabolites; and a median of 44 (range: 1-61) using predicted metabolites. The findings are largely expected because most of the Nightingale metabolites are lipid-related, and lipid traits have well-established relationships with blood cell indices. Association results are reasonably consistent between measured and predicted metabolites. We also performed MWAS between metabolites and ICD-10 disease case/control status. We identified numerous significant associations; for example, higher leucine is significantly associated with a higher risk of gout (M10), essential hypertension (I10), and disorders of lipoprotein metabolism and other lipidaemias (E78). In summary, MWAS in biobank samples can be a powerful approach to reveal molecular mechanisms and potentially discover novel therapeutic targets.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4272 Metabolome-wide Mendelian randomization for youth type 2 diabetes

Authors:

K. FAGBEMI[1], R. Avocegamou[1], M. Yazdanpanah[1], N. Yazdanpanah[1], D. MANOUSAKI[2]; [1]CHU Sainte-Justine, Univ. of Montreal, Montreal, QC, Canada, [2]Dept. of Pediatrics, Faculty of Med., Univ. of Montreal,, Montreal, QC, Canada

BACKGROUNDPreviously a rare disease in children, type 2 diabetes (T2D) has an increasing prevalence in youth, and is closely related to the pediatric obesity epidemic. Given the role of metabolism in diabetes, we aimed to identify circulating metabolites having a causal link with youth T2D as early disease biomarkers, using two-sample Mendelian randomization (MR).**METHODS AND RESULTS**
For our MR study, we tested causal effects on youth T2D of metabolites with available SNP-instruments in four GWAS. Effects of these SNPs on pediatric T2D were derived from the multi-ethnic ProDiGY GWAS (3,006 cases/6,061controls) or its European subset (664 cases/1434 controls). In multi-instrument MR analyses, steiger testing evaluated the directionality of the MR associations, and reverse two-sample MR analyses were performed for metabolites with evidence of reverse causation in our forward MRs. Multivariable MR (MVMR) tested effects of the metabolites on youth T2D conditioning on childhood BMI. Genetic colocalization was applied to identify shared causal variants between the MR-prioritized metabolites and youth T2D. Among 586 unique metabolites tested in the multiethnic MR, 25[DM1] metabolites demonstrated a causal effect on youth T2D (FDR P-values < 0.1), most of which clustered in the glycerophospholipid class. Among these, PC aa C32:2 and PC ae C36:2 had a protective effect (OR = 0.83, 95% CI: 0.77-0.89; OR = 0.89, 95% CI: 0.82-0.96). Steiger testing provided evidence for reverse causality for 4 metabolites (HDL diameter, caproate (6:0), alanine, serum cholesterol). However, the findings of the reverse MR analyses for the 4 metabolites did not support a causal effect of youth T2D on these metabolites. MVMR analyses showed that, conditioning on childhood BMI, the MR associations of HDL.D, serum cholesterol and PC aa C34:3 with youth T2D disappeared, while those of PC aa C32:2 and PC ae C36:2 remained. Among the candidate metabolites, only the glycerophosphocholine PC ae C36:2, colocalized with youth T2D (H4=0,804).In the European MR analysis, 8 of the 586 tested metabolites showed a causal effect on youth T2D (FDR P-value< 0.1), 6 of which are glycerophospholipids. 1-stearoyl-2-linoleoyl-GPE (18:0/18:2) had a protective effect on youth T2D (OR=0.90; 95% CI: 0.85-0.95), whereas the other 7 metabolites increased T2D risk, with 1-Docosahexaenoyl-GPE(22:6) associating with the highest risk (OR = 1.22; 95% CI: 1.08-1.37). Steiger testing supported the expected direction of the causal association for all 8 metabolites.We have identified causal effects of several glycophospholipids on youth T2D in our MR studies, including the PC ae C36:2, which colocalized with youth T2D.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4273 Meta-learning for instant genetic classification with neural networks.

Authors:

D. Bonet Sole[1,2], D. Mas Montserrat[1], X. Giro-i-Nieto[2], A. Ioannidis[1]; [1]Stanford Univ., Stanford, CA, [2]Univ.t Politècnica de Catalunya, Barcelona, Spain

Large-scale biomedical research databases provide the opportunity to adopt machine learning and deep learning approaches for genetic studies and precision medicine. However, training deep learning models and adjusting hyperparameter configurations for new tasks and datasets is computationally demanding and time-consuming. Traditional statistical and machine learning methods remain the preferred choice for most genomic studies, while neural network alternatives require extensive hyperparameter tuning or work only under very specific limitations. We present a meta-learning framework designed to perform instant classification on genetic classification tasks. A pre-trained hypernetwork receives a labeled SNP dataset and generates the weights of a task-specific neural model tailored to the provided dataset that can be directly used for inference, removing the need for training a new model. We effectively decouple the complexity of specialized genetic prediction models that perform individual tasks from a general meta-model. In other words, we train a high-capacity meta-model to encode task-specific variations in the parameters of a smaller model. We report extensive experiments on the UK Biobank and HapMap3 datasets for phenotype prediction tasks along with ancestry inference from genotype data, showing that our method outperforms competing neural networks approaches, traditional methods, and boosting machines, while being significantly faster. We provide invariance to permutations of input samples with scalability with respect to the dataset size, as well as feature-permutation invariance. Additionally, our approach demonstrates robust adaptability across a variety of tasks with little to no fine-tuning, positioning our approach as a strong solution for numerous applications and rapid model deployment. Our approach introduces a promising paradigm for obtaining instant predictions, with the potential to substantially decrease the computational burden of deep learning.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4274 Methods for multi-phenotype colocalization analysis in a single cohort

Authors:

S. Hanks[1], X. Yin[2], M. Laakso[3], X. Wen[4]; [1]Univ. of Michigan, Ann Arbor, MI, [2]Nanjing Med. Univ., Nanjing, China, [3]Inst. of Clinical Med., Kuopio, Finland, [4]Univ. of Michigan, Ann Arbor, MI

Genetic colocalization is the process of identifying genetic variants that are causal for multiple association signals at a single locus. Most existing methods explicitly assume that the traits are measured in independent, non-overlapping samples. However, there are increasing numbers of cohorts with multiple phenotypes collected on the same samples, such as the METSIM study measuring ~1.4K metabolites in ~6K individuals. The consequences of using these existing methods with a single cohort design are currently unknown. To investigate the consequences of violating the independent cohort assumption, we used genetic data from the METSIM study to simulate two continuous traits, varying the proportion of the error variance that is attributed to individual-level confounding. We used fastENLOC to perform colocalization on the simulated data and found that Type I error rates are well-controlled (FDR less than 5%) when treating a single cohort as independent samples. We also propose a new method to perform colocalization in a single cohort by explicitly estimating the shared unobserved confounding across the two traits with probabilistic principal component analysis (PPCA). We then control this estimate as a covariate in the fine-mapping stage of the colocalization analysis. With this new method, Type I error remains well-controlled, and power increases up to 22%. In the future, we will use this method to perform colocalization analysis with metabolite data in the METSIM study.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4275 Methyl-TWAS: A powerful method for precise pseudo-transcriptome-wide association studies (TWAS) using long-range DNA methylation.

Authors:

S. Kim[1], Y. Qin[1], H. Park[1], M. Yue[1], Z. Xu[1], E. Forno[1], W. Chen[1], J. Celedon[2]; [1]Univ. of Pittsburgh, Pittsburgh, PA, [2]Children's Hosp of Pittsburgh of UPMC, Univ. of Pittsburgh, Pittsburgh, PA

While transcriptome-wide association studies (TWAS) using genotype data have gained much popularity, these show low prediction accuracy of imputed gene expression. This can be because SNPs are neither tissue specific nor disease-specific and are not affected by environmental factors while gene expressions are tissue- and disease-specific and largely affected by environmental factors. Also, majority of GWAS SNPs indirectly regulate gene expression by regulating DNA methylation.We previously developed geneEXPLORE that incorporates long-range methylation CpGs to maximize prediction accuracy, outperforming other methods in pan-cancer. However, this study did not apply the method for TWAS. Owing to the good prediction accuracy of gene expression using long-range DNA methylation, here we propose to conduct TWAS using long-range DNA methylation rather than genotype data. To this end, we developed Methyl-TWAS, the first TWAS method that incorporates DNA methylation instead of genotype data. Using nasal epithelium, we found that the predictive power of DNA methylation on gene expression is much higher than that of SNPs. The genes better predicted by Methyl-TWAS are in immune pathways whereas the genes better predicted by PrediXcan are in metabolic and biosynthetic pathways. Further, Methyl-TWAS predicts immune genes and differentially expressed genes significantly better than randomly selected genes while PrediXcan does not. Owing to the accurately predicted expressions of genes, Methyl-TWAS outperforms PrediXcan in identifying DEGs of atopic asthma. Also, in an external test data set, Methyl-TWAS identified differentially expressed genes of atopic asthma accurately as much as using observed expression data. Methyl-TWAS would give opportunities for researchers to accurately conduct TWAS using increasing numbers of publicly available DNA methylation data to study various immune diseases.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4276 MiXeR-TAG: A new method for classification of overlapping and trait specific variants.

Authors:

N. Parker[1], O. Frei[1], A. Shadrin[1], G. Hindley[1], O. Smeland[1], O. Andreassen[2], A. Dale[3]; [1]Univ. of Oslo, Oslo, Norway, [2]Univ. of Oslo, Oslo, Oslo, Norway, [3]Univ. of California San Diego, San Diego, CA

Previous studies have identified a need for assessment of genetic overlap between pairs of traits beyond genetic correlation. Bivariate MiXeR uses a Gaussian mixture model to estimate the number of shared and trait-specific genetic variants between two complex traits. However, identification of which genetic variants are shared or trait-specific was not previously available.

Here we present a novel extension to bivariate MiXeR to assign genetic variants to one of 3 categories: specific to trait 1, specific to trait 2, or shared by both traits. In addition, we report trait-specific variants located in the same independent genomic block. We also use this novel MiXeR based trait annotation of genetic variants (MiXeR-TAG) to perform gene enrichment analyses for each category. P-values from the original GWAS were used to threshold variants to be significantly associated with at least 1 trait. For each variant likelihood estimates are derived for the 4 categories and variants were assigned the category with the highest value. For gene enrichment analyses, we selected the nearest gene assigned to independent significant SNPs (independent at R^2<0.1 and distance 250kb) and used hypergeometric tests to assess significant overlap. We used GWAS summary statistics for four trait comparisons to test this method: (1) cerebral cortical thickness (TH) and height, (2) cerebral cortical surface area (SA) and height, (3) Crohn's disease (CD) and body mass index (BMI), and (4) CD and rheumatoid arthritis (RA). Enrichment analyses showed that TH genes were upregulated in the brain (p_{fdr}=2.7e-10). The same was true for SA genes (p_{fdr}=2.2e-2). Gene ontology enrichment analyses implicated neuron development and the post-synapse cellular component for both TH and SA genes. Enrichment among GWAS catalog genes showed TH genes were enriched for genes linked to cognitive ability and several psychiatric disorders (p_{fdr}<0.05). Also, height genes were enriched for the height GWAS genes (p_{fdr}=2.7e-122). Genes shared by CD and BMI were upregulated in adipose tissue (p_{fdr}=1.93e-6) while RA genes were upregulated in blood vessels (p_{fdr}=3.7e-3). Genes shared by both CD and BMI as well as CD and RA were enriched for cytokine and immune system processes based on gene ontology groups and KEGG pathways (p_{fdr}<0.05). Also, BMI specific genes were enriched for BMI GWAS genes (p_{fdr}=1.8e-120).

Overall, genes mapped to variants assigned to trait specific or shared categories by MiXeR-TAG exhibited gene enrichment for biologically viable and relevant gene sets. The MiXeR-TAG approach provides a novel method for understanding the shared and unique genetic architecture of complex traits.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4277 Modelling genotype-phenotype dose-response dynamics to prioritise novel drug targets: a statistical workflow for common and rare variants

Authors:

L. Stefanucci[1], Z. Ke[1], C. Jones[1], D. Ochoa[2], S. Jeremy[1], E. Holzinger[3], J. Maranville[3], E. McDonagh[4], K. Alasoo[5], I. Dunham[6], M. Ghoussaini[7], G. Trynka[8]; [1]Wellcome Trust Sanger Inst., Cambridge, United Kingdom, [2]European Molecular Biology Lab., European Bioinformatics Inst. (EMBL-EBI), Hinxton, United Kingdom, [3]BMS, Cambridge, MA, [4]Stanford Univ., PharmGKB, Palo Alto, CA, [5]Univ. of Tartu, Tartu, Estonia, [6]Wellcome Sanger Inst., Wellcome Genome Campus, Cambridge, United Kingdom, Saffron Walden, United Kingdom, [7]Wellcome Trust Sanger Inst, Cambridge, United Kingdom, [8]Wellcome Sanger Inst., Hinxton, United Kingdom

Unexpected toxicity and erroneous drug target selection are the primary culprits behind the high attrition rates encountered during drug development. Human genetics holds great promise to mitigate these challenges: drugs supported by genetic evidence exhibit a higher success rate than those lacking such support. When multiple independent variants are identified for the same gene-trait association (i.e. allelic series), they can be used to construct dose-response curves and titrate the relationship between protein function and phenotypes. The construction of these relationships can provide crucial information to the drug development process. For instance, a positive relationship indicates that mimicking the loss of protein function (e.g. by targeting with an inhibitor) becomes the key mechanism for disease prevention. Mendelian randomisation (MR) leverages GWAS and QTL to construct these dose-response relations. However, the number of instrument variants available limits the power to estimate these relations. Yet, including additional rare variants, beyond QTL and GWAS, could further validate MR observations. Furthermore, rare variants provide insights at the dose-response curve extremes, which are more relevant for drug development.We developed a framework that can use Open Targets data (2,959,742 genetic variants and 22,274 common traits) to correlate genetic variants with disease risk, focusing in particular on rare variants. To estimate the coding-variant effect on protein function, when experimental data were not available, we trained a gradient-boosted decision tree model (Specificity=0.98; Sensitivity=0.81). To infer the latent dose-response relation, we used a Bayesian hierarchical framework, independently modelling the different classes of variants. Extensive simulations show an accurate predictive capacity (MAPE=2.01%). Summary statistics modelling confirmed the biological mechanism and the expected dose-response correlation for a well-characterised genotype-phenotype association, such as PCSK9 and hypercholesterolemia. By factoring in rare variants, our approach unveiled a stronger correlation (β=0.24; CI=[0.12-0.29]) compared to MR (β=0.074, sd=0.004) between PCSK9 concentration and high cholesterol. This promising method can be expanded to virtually all available genes and phenotypes.To conclude, our workflow models dose-response relationships between genes and diseases, combining the effect of common and rare variants. This approach can systematically reveal gene roles in disease aetiology and, notably, provides insight into the directionality of the effect and, thus, drug modulation.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4278 MR-link 2: Improved *cis* Mendelian randomization validated through three independent validation datasets.

Authors:

A. van der Graaf[1], R. Warmerdam[2,3], U. Võsa[4], L. Franke[5,3], Z. Kutalik[1]; [1]Univ. of Lausanne, Lausanne, Switzerland, [2]Univ. Med. Ctr. Groningen, Groningen, Netherlands, [3]Oncode Inst., Utrecht, Netherlands, [4]Univ. of Tartu, Tartu, Estonia, [5]Univ. Med. Ctr. Groningen, Groningen, Netherlands

Background: Causality shapes human biology, but it is difficult to prove as it usually requires randomized trials. Mendelian randomization (MR) can identify causal relationships from observational data but has increased false positive rates (FPRs) when genetic instruments are scarce. This typically occurs when molecular traits are considered: often only a single *cis* associated region may be available for analysis.
Methods: We present a new, summary statistics *cis* MR method: MR-link 2. MR-link 2 models all SNPs in a region to estimate the causal effect while accounting for pleiotropy. We compare MR-link 2 to three *cis* MR methods: MR-IVW, MR-IVW-R and MR-PCA and two colocalization approaches: coloc and SuSiE-coloc. We test each method in three independent validation datasets of causal relationships: i) simulations, ii) gold standard metabolite networks and iii) canonical causal relationships between complex traits.
Results: In simulations that include pleiotropy, MR-link 2 has well-calibrated FPR and high power, resulting in area under the receiver operator characteristic curves (AUC) superior to other methods. In some cases, more than 0.2 units better than the best competing method, MR-PCA. Our simulations show that when MR-link 2 assumptions of infinitesimal genetic architecture are violated, FPR increases slightly, still it outperforms competing methods in 75% of simulated scenarios.
Using 921 metabolite pairs as true causal relationships derived from 3 metabolic pathway databases (KEGG, MetaCyc and Wikipathways) and metabolites from 4 metabolite QTL (mQTL) studies, MR-link 2 achieves an AUC of 0.696, outperforming all tested competing methods (max AUC: 0.664). MR-link 2 also has lower bias in causal effect estimation (0.036) than the best competing approach, MR-IVW (0.040) when testing the causal effect of metabolites on themselves using mQTLs coming from different studies.
Applied to 9 canonical causal relationships in complex traits (e.g., smoking causes cancer), and 5 true negatives (e.g., adult blood pressure causes child onset asthma), all MR methods correctly distinguish causal from non-causal pairs. When comparing individual locus-based causal effect estimates, MR-link 2 has the lowest heterogeneity across loci (median Cochran's Q: 686, lowest competing: MR-IVW, 1438) and best calibrated FPR (0.09) for the true negative pairs at ($P<0.05$) compared to other methods (lowest FPR: MR-IVW, 0.15).
Conclusion: MR-link 2 is a MR method for *cis* causal inference. In three validation datasets, MR-link 2 outperforms competing approaches. MR-link 2 has lower FPR while still retaining high power, due to being more robust to MR assumption violations.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4279 MTClass: Identification and annotation of multi-phenotype cis-eQTLs using machine learning.

Authors:

R. Li[1], S. Qin[2]; [1]Emory Univ., Atlanta, GA, [2]Emory Univ, Atlanta, GA

Dysregulation of gene expression contributes to the pathophysiology of many human diseases. Mapping genetic variants that affect the expression of one or more genes, termed expression quantitative trait loci (eQTLs), is a growing area of interest. Multi-phenotype eQTL mapping identifies variants that affect the expression of the same gene across multiple phenotypes (e.g., tissues, exons), thereby potentially recognizing variants with stronger and broader functional impacts that are more likely to lead to phenotypic consequences. However, measuring the effect of a variant on gene expression across many phenotypes is a challenge. To solve this problem, we propose a model-free strategy named MTClass. Instead of calculating a p-value under a statistical testing framework, we attempt to classify vectors of multi-phenotype expression values in terms of donor genotype for the variant of interest. We then rank variant-gene pairs based on classification performance. Using data from the GTEx Consortium, we utilized our strategy in an example with 13 brain tissues, 317 donors, and 47% missing data. To accommodate missing expression measures in donor-tissue combinations, we used multiple imputation with predictive mean matching. MTClass was able to identify variants and genes with stronger functional impacts than competing methods such as multi-phenotype association tests. Our method consistently outperformed two other methods in terms of neighboring GWAS SNP hits calculated for the top 5,000 variants. Using gene-set enrichment analysis, we observed more significant enrichment from our method's top 100 genes than from those of other methods. We extended the same framework to classify variants that affect the expression of multiple exons in brain tissues. In relevant brain regions, we found greater overlap with known disease-related genes among our top 100 genes compared to single-tissue approaches and other multivariate methods. Finally, we constructed a "2D case" in which we used an artificial neural network to classify genotypes across multiple exons and 9 tissues with no missing data. We found that other methods were not able to do this and that MTClass can uncover variants that are both eQTLs and sQTLs across multiple tissues. Together, these results emphasize that multi-phenotype eQTLs from our method likely have functional consequences, and they suggest the value of further pursuing such variants.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4280 mtPGS: Leverage Multiple Correlated Traits for Accurate Polygenic Score Construction

Authors:

C. Xu, S. Ganesh, X. Zhou; Univ. of Michigan, Ann Arbor, MI

Background: Accurate polygenic scores (PGS) facilitate the genetic prediction of complex traits and aid in the development of personalized medicine. Multivariate PGS methods leverage traits that are correlated or relevant to the target complex trait of interest to facilitate its prediction. However, existing multivariate PGS methods often make relatively simple modeling assumptions on the shared genetic architecture between the target trait and the relevant traits, and these methods have largely ignored the environmental correlation between traits due to sample overlap. Further efforts are needed for the development of multivariate PGS method that relies on flexible modeling assumption and incorporates environmental correlations for PGS construction. **Methods:** We develop a statistical method called multi-trait assisted PGS (mtPGS), which can construct accurate PGS for a target trait of interest through leveraging multiple traits relevant to the target trait. Specifically, mtPGS borrows SNP effect size similarity information between the target trait and its relevant traits to improve the effect size estimation on the target trait, thus achieving accurate PGS. In the process, mtPGS flexibly models the shared genetic architecture between the target and the relevant traits to achieve robust performance, while explicitly accounting for the environmental covariance among them to accommodate different study designs with various sample overlap patterns. In addition, mtPGS uses only summary statistics as input and relies on a deterministic algorithm with several algebraic techniques for scalable computation. **Results:** The predictive performances of mtPGS were evaluated through comprehensive simulations and applications to 25 traits in the UK Biobank (UKB). In the simulation settings with distinct genetic architectures and various combinations of SNP heritability and genetic/environmental correlation, mtPGS achieved an average of 7.05%-504% accuracy gain compared to the seven other state-of-the-art PGS methods. mtPGS also achieved robust predictive performances in the simulations with different scenarios of sample overlap and different number of relevant traits. For the prediction of 25 traits in the UKB, mtPGS achieved an average of 2.17%-51.12% accuracy gain over competing PGS methods. mtPGS compares favorably to the other PGS methods with relatively low computing time and memory cost. **Conclusion:** Our proposed PGS method, multi-trait assisted PGS (mtPGS), represents an accurate and scalable method for PGS construction in the large-scale biobank datasets.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4281 Muesli intake protects against coronary artery disease: A Mendelian randomization study on 13 dietary traits.

Authors:

J. Park[1], B. Petrazzini[1], S. Bafna[1], A. Duffy[1], I. Forrest[1], H. M. Vy[1], C. Marquez-Luna[1], M. Verbanck[2], J. Narula[1], R. S. Rosenson[1], D. M. Jordan[1], G. Rocheleau[1], R. Do[1]; [1]Icahn Sch. of Med. at Mount Sinai, New York, NY, [2]Université Paris Cité, Paris, France

Background: Diet is a key modifiable risk factor of coronary artery disease (CAD). However, the causal effects of specific dietary behaviors on CAD risk remain unclear, as observational studies are potentially biased by confounding and reverse causation. Given the genetic heritability of lifestyle traits and the advancement of dietary data in population biobanks, genetic epidemiology could help infer causality between specific dietary traits and CAD risk.
Methods: We performed two-sample MR using inverse variance weighted (IVW) regression to estimate the causal effects of 13 common dietary traits on CAD risk, using cross-sectional genetic and dietary data on up to 420,531 UK Biobank participants and 184,305 CARDIoGRAMplusC4D participants of European ancestry. To examine the underlying mechanisms mediating any detected associations, we then conducted follow-up MR analyses involving 171 plasma metabolite levels from UK Biobank. Finally, we performed sensitivity analyses using weighted median estimation, MR-Egger regression, and MR-PRESSO.
Results: Genetically-proxied muesli intake was negatively associated with CAD risk (OR 0.74, 95% CI 0.65-0.84, $P = 5.39 \times 10^{-4}$). Sensitivity analyses using weighted median estimation demonstrated a nominally-significant association in the same direction. Additionally, we identified acetate (OR 0.037, 95% CI 0.01-0.12, $P = 1.15 \times 10^{-4}$) as the plasma metabolite potentially mediating this association.
Conclusions: Consuming muesli, a mixture of oats, seeds, nuts, and dried fruit, may causally reduce CAD risk. Circulating levels of acetate, a gut microbiota-derived short-chain fatty acid, may causally mediate this cardioprotective effect. These findings could prioritize clinical trials for dietary interventions, provide mechanistic insights into a gut-heart axis, and inform dietary policy for population health.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4282 Multi ancestry multi trait meta analysis of autoimmune diseases.

Authors:

K. Anderson Fears, D. Liu, B. Jiang, X. Wang; Pennsylvania State Univ., State College, PA

Genetic association signals are often shared between traits or ancestries, and jointly analyzing datasets via meta-analysis across different traits and ancestries will likely improve power. To take advantage of the wealth of GWAS data available for including multiple traits or multiple ancestries, we propose a new method for a combined multi-trait and cross ancestry meta-analysis. The new method uses a variable number of principal components from a genetic correlation matrix as covariates in a meta-regression model to account for phenotype heterogeneity and adjusts for sample overlaps using residual correlation from LD score regression. To test the effectiveness of our new method, we compiled the largest multi-ancestry SLE dataset consisting of GWAS summary statistics for 740,000 individuals in six different ancestries for 14 autoimmune diseases. For many of these diseases both genetic and environmental risk factors are shared with many precursors to the diseases only becoming highly specific immediately prior to development of final stages. These similarities in triggers have made specification in testing arduous and gives the perfect testing grounds for our method. To test our method with those available we compare our method with the multi-trait analysis of genome-wide association (MTAG) method, using Popcorn for genetic correlation, as well as a standard fixed effect meta-analysis. Our results show that IMAC outperforms MTAG in simulation, with controlled type 1 error compared to MTAGs conservative type 1 error, and higher statistical power across all scenarios of heritability (0-1) and sample sizes (20,000, 50,000 and 100,000). On real world data, IMAC, using a single principal component with a p-value cutoff of 0.2, identifies 57, 81, 105, 127 and 134 novel loci compared to MTAGs 61, 61, 30, 28, and 29 across five correlation sets. Our findings highlight the importance of including multi-ancestry and multi-trait information together, which along with increasing sample sizes, increases power for rare variant detection and has the potential to reduce health disparities.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4283 Multi-ancestry transcriptome predictions with functionally informed variants improve transcriptome-wide association studies in TOPMed MESA

Authors:

X. Hu[1], D. Araujo[2], S. Rich[1], J. Rotter[3], M. Cho[4], L. Raffield[5], Y. Li[5], D. Liu[6], H. Wheeler[2], H. Im[7], A. Manichaikul[1], TOPMed Multi-Omics Working Group; [1]Univ. of Virginia, Charlottesville, VA, [2]Loyola Univ. Chicago, Chicago, IL, [3]Lundquist Inst., Harbor-UCLA Med Ctr, Torrance, CA, [4]Brigham and Women's Hosp., Boston, MA, [5]UNC - Chapel Hill, Chapel Hill, NC, [6]Pennsylvania State Univ., Hershey, PA, [7]Univ. of Chicago, Chicago, IL

Reliable prediction of genetically regulated gene expression is key to accurate transcriptome-wide association studies (TWAS). Reference transcriptome prediction models for TWAS have been constructed primarily based on individuals of European ancestry. With the emergence of multi-ancestry GWAS, there is a need for reliable multi-ancestry transcriptome prediction models for downstream TWAS efforts. Furthermore, the genomic variants overlapping with key annotations (e.g., fine-mapping, 3D genomics informed regions, and epigenetic processes) are more likely to be functionally relevant to influence gene expression. We developed multi-ancestry transcriptome prediction models with functionally informed variants (FIVs) by leveraging PBMC RNA-seq from 1,287 TOPMed MESA multi-ancestry samples and corresponding whole-genome sequencing data. Then we examined the performance of models on both prediction accuracy and TWAS. We built four prediction models including one benchmark model (Elastic Net, EN), and three models with FIVs, i.e., EN with fine-mapped variants; Prediction Using Models Informed by Chromatin conformations and Epigenomics, PUMICE; and PUMICE with fine-mapped variants. The prediction accuracy of four models was then assessed in Geuvadis cohort using 449 multi-ancestry samples with LCL RNA-seq. To examine model's performance on TWAS, we leveraged summary statistics from two recent multi-ancestry GWAS, the Global Lipids Genetics Consortium (GLGC) GWAS (N~1.65 million), and lung function GWAS (N=580,869). We then examined TWAS precision by overlapping Bonferroni-significant TWAS genes with previously identified GWAS trait-related putative causal genes (i.e., Mendelian and Mouse knockout genes, genes with coding variants, genes with rare exonic association, nearest genes). While the gene expression prediction accuracy was similar across the four models in both discovery and validation analyses, the TWAS from models with FIVs outperformed. The TWAS from models with FIVs identified more putative causal genes than the TWAS from EN model for three out five lipid traits and for three out of four lung function traits respectively. For example, the TWAS from PUMICE and EN identified 87 and 80 putative causal genes respectively for total cholesterol. Similarly, the TWAS from PUMICE identified 13 putative causal genes for peak expiratory flow rate, while the TWAS from EN only identified 6 genes. Our study demonstrates the value of including FIVs in multi-ancestry transcriptome prediction models for improving TWAS precision.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4284 Multi-class Modeling Identifies Shared Genetic Risk for Late-onset Epilepsy and Alzheimer's Disease

Authors:

M. Fu, T. Tran, UCLA Precision Health Data Discovery Repository Working Group, UCLA Precision Health ATLAS Working Group, K. Vossel, T. Chang; Univ. of California, Los Angeles, Los Angeles, CA

Introduction: Previous studies show a strong link between late-onset epilepsy (LOE) and Alzheimer's disease (AD). Individuals with AD have a higher risk of developing LOE, while LOE can be a subtype and risk factor for AD. However, their shared genetic risk beyond the *APOE* gene remains unclear. Our study aims to identify the shared genetic factors of AD and LOE, interpret the biological pathways involved, and determine the effect of shared genetic risk on AD through LOE. **Methods:** We used electronic health records (EHRs) from UCLA Health for discovery analyses and validated our findings in the All of Us database. We defined phenotypes using phecodes with patients' records aged 60-90. A two-step Least Absolute Shrinkage and Selection Operator (LASSO) workflow was used to identify shared and disease-specific genetic variants between AD and LOE based on prior functional genomic data. Shared risk variants were mapped to genes for biological interpretations. We calculated an AD-LOE shared risk score and used it as a proxy in a causal mediation analysis. **Results:** In the UCLA European genetic inferred ancestry (GIA) sample (N=17,031), the two-step LASSO method identified 34 shared genetic loci between AD and LOE, located at *19q13* (the *APOE* region), *8p21*, *6p12*, and *11q12*. These loci were further mapped to 65 genes, which showed enrichment in molecular functions and pathways such as tau protein binding and plasma lipoprotein metabolism. Disease-specific risk genes, such as *CR1* (AD-specific) and *PTK2B* (LOE-specific), were also identified. Individuals with high predicted shared risk scores have a higher risk of developing AD, LOE, or both in their later life compared to those with low-risk scores. LOE partially mediates (20%) the effect of AD-LOE shared genetic risk on AD. Validation results from All of Us were consistent with findings from the UCLA sample. **Conclusion:** Our study used machine learning to identify shared genetic risks of AD and LOE in the European GIA population. In addition to providing substantial evidence for the significant contribution of the *APOE-TOMM40-APOC1* gene cluster to shared risk, we uncovered novel genes that may contribute to shared or disease-specific risk. We developed models using the UCLA sample and validated them on the All of Us database to ensure our findings are reliable and generalizable. Our study is one of the first to use All of Us genetic data to investigate AD. It provides insights into the mechanisms underlying AD and LOE and could lead to disease prevention, targeted treatment strategies, and innovative drugs for the co-occurrence of these two diseases.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4285 Multidimensional analysis of pedigree, epidemiologic, and molecular data identify risk and perceived causative factors for ME/CFS.

Authors:

R. Moslehi[1], A. Kumar[1], A. Bhanushali[1], M. Tracy[1], A. Dzutsev[2]; [1]Univ. at Albany, State Univ. of New York, Albany, NY, [2]NCI, NIH, Bethesda, MD

Background: Myalgic encephalomyelitis (ME)/chronic fatigue syndrome (CFS) is a complex disabling disorder with no known etiology or approved treatment. Estimates of the prevalence suggest that up to 3.4 million Americans may be afflicted. It has been suggested that ME/CFS may be triggered by an infectious illness including COVID-19, hence the prediction that 10 million new cases of ME/CFS may be diagnosed globally due to the pandemic. We conducted a molecular epidemiologic study to identify risk factors and biologic mechanisms for ME/CFS. Methods: Our clinic-based case-control study involved 59 carefully selected ME/CFS patients and 54 appropriately matched healthy controls. We compared cases and controls with respect to the following: 1. prevalence of autoimmune disease (AID) and cancer among their first-degree relatives, 2. prevalence of epidemiologic factors, and 3. serum levels of 48 cytokines. Statistical methods used were logistic regression and cumulative incidence analysis to calculate odds ratios (OR), relative risks (RR), 95% confidence intervals (CI) and p-values. We used machine learning approaches to identify a cytokine profile of ME/CFS. Results: Our analysis revealed that ME/CFS cases were five times more likely than controls to have a family history of AID (OR=5.30, p=0.002). The life-time risk of AID among first-degree relatives of cases was significantly higher compared to the relatives of controls (RR=3.72, p=0.0006). First-degree relatives of cases also had a significantly higher risk of early-onset (diagnosed <60 years of age) cancer compared to the relatives of controls (RR=2.81, p=0.03). Comparison of epidemiologic factors identified history of allergies requiring medication (OR=6.00, p<0.0001), exposure to contaminants (OR=4.35, p=0.0002), history of illness requiring hospitalization (OR=4.33, p=0.0004), ≥4 episodes of significant illness requiring hospitalization (OR=24.36, p<0.0001) and ≥2 episodes of significant stress (OR=3.07, p=0.03) as risk factors for ME/CFS. The most common self-identified perceived *causes* of ME/CFS reported by cases in response to an open-ended question were Infectious Illness (27.3%), Infectious Agents (15.9%), and Stress (15.9%). Our analysis identified a cytokine profile of ME/CFS, which classified patients with 84% accuracy (kappa=0.68, p=0.025, sensitivity=0.75, specificity=1.00) in a separate test set in random forest models. Conclusions: Findings from our multidimensional analysis of pedigree, epidemiologic and molecular data suggest certain risk factors for ME/CFS and links with AID and cancer, providing etiologic clues and druggable targets for treatment of ME/CFS.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4286 Multi-omics Integration Identifies Genes Influencing Traits Associated with Cardiovascular Risks

Authors:

S. Acharya[1], L. Liao[1], W. Jung[1], E. Kang[1], V. A. Moghaddam[1], M. Feitosa[2], M. Wojczynski[3], S. Jin[1], J. Anema[4], K. Schwander[5], J. O. Connel[6], M. Province[1], M. Brent[1]; [1]Washington Univ. in St Louis, St Louis, MO, [2]Washington Univ Sch Med., Saint Louis, MO, [3]Washington Univ Schl Med, St Louis, MO, [4]Washington Univ. Sch. of Med., St. Louis, MO, [5]Washington Univ. Sch. of Med., St Louis, MO, [6]Univ. of Maryland, Bethesa, MD

The Long Life Family Study (LLFS) comprises 4,953 participants aged 24-110 years in 539 pedigrees displaying exceptional longevity. We aim to identify genetic mechanisms that protect LLFS participants against age-related cardiovascular risks. We hypothesized that (1) among genes that lie near non-coding genomic variants associated with a trait, genes whose RNA level is also associated with the trait are more likely to be causal, and (2) genes with low P-values that interact with each other in a molecular system and participate in the same biological process are more likely to be causal. Based on these hypotheses, we developed a freely available multi-omics integration pipeline that identified candidate causal genes for 11 cardiovascular risk traits. Using our pipeline, we aggregated gene-level statistics from rare-variant analysis, GWAS, and gene expression-trait association by Correlated Meta-Analysis (CMA). It then identifies Protein-Protein Interaction (PPI) network modules enriched for both genes with low CMA P-values and genes annotated to biological process terms overrepresented in these modules. Each input to CMA is generated by a linear mixed model. All 11 traits were adjusted for family relatedness, age, sex, field center, and 10 genetic PCs. Pulmonary traits were further adjusted for height and smoking, low-density lipoprotein and total cholesterol for statin use, and waist and pulse pressure for BMI. Across all traits, CMA identified 51 significant genes after Bonferroni correction ($P \leq 2.8 \times 10^{-7}$). *CETP*, *NLRC5*, *SLC45A3*, and *TOMM40* lie within 50 Kb of a known trait-associated variant (*previously associated genes*). Module analysis identified 65 genes (2 CMA-significant) that (1) have CMA P-value $\leq 5 \times 10^{-3}$, (2) lie in a PPI module enriched for genes with low P-values, and (3) are annotated with a biological process that is enriched among module genes, 10 of which were previously associated with the same traits. Permutation analysis showed that these criteria yield a false positive rate of 1 in 14876. For high-density lipoprotein levels, *ENPP2* (P < 1×10-4), *ABCG1* ($P < 6 \times 10^{-4}$), *APOC3* ($P < 2 \times 10^{-5}$), and *CETP* ($P < 2 \times 10^{-15}$) interact in a PPI module whose genes have P-values that are significantly lower than expected and are significantly enriched for lipid-related biological processes. *ABCG1*, *APOC3*, and *CETP* are previously associated genes participating in cholesterol transport. *ENPP2* has been implicated in adipose tissue expansion in obesity and participates with *APOC3* in the glycerolipid catabolic process. Overall, module analysis identified highly plausible candidate causal genes whose P-values after CMA alone were merely suggestive.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4287 Multiple approaches to generating polygenic risk scores for lung adenocarcinoma in East Asian never-smokers

Authors:

B. Blechter[1], K. Wang[1], T-Y. Chen[2], K. Shiraishi[3], K. Matsuo[4], K. Chen[5], J. Shi[1], I-S. Chang[6], J. Dai[7], K. Young Tae[8], P. Choudhury[9], J. Y. Wong[10], W. Hu[1], J. Choi[1], R. Hung[11], X-O. Shu[12], D. Lin[13], W. Zheng[14], Z. Yin[15], W. Seow[16], Q. Cai[17], N. Chatterjee[18], O. Gorlova[19], C. Amos[20], H. Shen[7], C. Hsiung[2], S. Chanock[1], N. Rothman[1], T. Kohno[21], Q. Lan[1], H. Zhang[1]; [1]Div. of Cancer Epidemiology and Genetics, Natl. Cancer Inst., Rockville, MD, [2]Inst. of Population Hlth.Sci., Natl. Hlth.Res. Inst.s, Zhunan, Taiwan, [3]Natl. Cancer Res. Inst., Tokyo, Japan, [4]Div. of Molecular Med., Aichi Cancer Ctr. Res. Inst., Nagoya, Japan, [5]Dept. of Epidemiology and Biostatistics, Natl. Clinical Res. Ctr. for Cancer, Key Lab. of Molecular Cancer Epidemiology of Tianjin, Tianjin Med. Univ. Cancer Inst. and Hosp., Tianjin Med. Univ.,, Tianjin, China, [6]Natl. Inst. of Cancer Res., Natl. Hlth.Res. Inst.s, Zhunan, Taiwan, [7]Dept. of Epidemiology, Sch. of Publ. Hlth., Nanjing Med. Univ., Nanjing, China, [8]Cancer Res. Inst., Seoul Natl. Univ. Coll. of Med., Seoul Natl. Univ. Hosp., Seul, Korea, Republic of, [9]American Cancer Society, Atlanta, GA, [10]Lab. of Genomic Instability and Cardiopulmonary Outcomes, Natl. Heart, Lung, and Blood Inst., Bethesad, MD, [11]Sinai Hlth., Univ. of Toronto, Toronto, ON, Canada, [12]Vanderbilt Univ. Med. Ctr., Nashville, TN, [13]Dept. of Etiology & Carcinogenesis and State Key Lab. of Molecular Oncology, Cancer Inst. and Hosp., Chinese Academy of Med. Sci. and Peking Union Med. Coll., Beijing, China, [14]Vanderbilt Univ Med Ctr., Nashville, TN, [15]Dept. of Epidemiology, Sch. of Publ. Hlth., China Med. Univ., Shenyang, China, [16]Saw Swee Hock Sch. of Publ. Hlth., Natl. Univ. of Singapore, Singapore, Singapore, [17]Vanderbilt Univ., Nashville, TN, [18]Dept. of Biostatistics, Johns Hopkins Univ., Baltimore, MD, [19]Dept. of Med., Section of Epidemiology and Population Sci., Inst. for Clinical and Translational Res., Houston, TX, [20]Baylor Coll. of Med., Houston, TX, [21]Div. of Genome Biology, Natl. Cancer Res. Inst., Tokyo, Japan

Background Lung cancer among never-smokers is a significant global health burden, with the highest incidence rates found in East Asian (EAS) women. Polygenic risk scores (PRSs), representing cumulative genetic susceptibility, can stratify women by lung cancer risk. However, PRS development has focused on European (EUR) individuals who are active smokers. We developed and validated single-and multi-ancestry PRSs for lung adenocarcinoma (LUAD) in EAS never-smokers, using the largest available genome-wide association study (GWAS) dataset. *Methods* Several PRS methods were evaluated to assess LUAD risk in EAS women using summary statistics from 3,564 never-smoking cases and 16,238 controls of EAS ancestry, and from 2,058 never-smoking cases and 5,575 controls of EUR ancestry. Single-ancestry PRS methods that used GWAS data from only EAS women included a PRS incorporating 25 variants that have previously reached genome-wide significance, PRS-25, as well as a Bayesian-based PRS, LDpred2. PRS tuning and validation was conducted in the Female Lung Cancer Consortium in Asia, an independent dataset of 4,438 never-smoking EAS cases and 4,544 controls, by estimating the area under the receiver operating characteristics curve (AUC). The AUC of LDpred2 PRS was then projected at different sample sizes using estimates of effect-size distribution and heritability derived from EAS populations using GENESIS. Next, we developed a multi-ancestry PRS using CT-SLEB, an approach incorporating genetic data from EAS and EUR individuals that models the genetic correlations across ancestries. We also estimated the lifetime absolute risk of LUAD based on age-specific lung cancer incidence and overall mortality rates. *Results* The single ancestry methods PRS-25 and LDpred2 PRS had AUCs of 0.62 and 0.63, respectively. Using LDpred2 PRS, we projected that a future study with a 1:1, 1:4 or 1:10 case-control ratio would need 105,000, 65,000 or 55,000 EAS never-smoking LUAD cases, respectively, to attain an AUC of 0.70. We found that the multi-ancestry PRS method CT-SLEB had an AUC of 0.64. Compared to women in the middle quintile, those in the lowest and highest 5% of the CT-SLEB PRS had 0.42 and 4.40-fold risk of developing LUAD, respectively. Further, the lifetime risk (age 30-80) of LUAD in women in the lowest and highest 5% of the PRS were 0.78% and 6.69%, respectively. *Conclusions* Improved methods of using GWAS data to generate a PRS could result in increasing AUC estimates and guide sample size calculations for new studies aiming to further characterize the genetic architecture of LUAD in never-smokers. This approach could inform primary and secondary prevention efforts in the future.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4288 Multi-Trait Analyses Reveal Novel Genetic Etiology for Congenital Heart Disease and Congenital Diaphragmatic Hernia

Authors:

Y. Xie[1], R. Wu[1], K. Ng[2], W. Dong[2], H. Li[1], M. Lek[2], M. Brueckner[2], H. Zhao[1]; [1]Yale Univ. Sch. of Publ. Hlth., New Haven, CT, [2]Yale Univ. Sch. of Med., New Haven, CT

Compared with common diseases, genetic factors are likely to play a more critical role in the etiologies of birth defects in contrast to non-genetic factors (i.e., environment). With its low occurrence in the population, rare de novo variants (DNVs) have been shown to be a major contributor to birth defects. Meanwhile, the development of transcriptomic technologies enables researchers to understand genomic function at the single-cell resolution and may help reveal the pathogenesis of birth defects. Currently, the integration of DNVs and single-cell expression data has not been thoroughly explored. To bridge this gap, we developed an analytical pipeline to discover the shared genetic mechanisms between birth defects by incorporating DNV data from multiple congenital traits and single-cell data from developing tissues. For DNVs, we curated publicly available data from 2,645 congenital heart disease (CHD) trios, 827 congenital diaphragmatic hernia (CDH) trios, and 1,789 autism-unaffected trios as healthy controls. For single-cell expression, we utilized published mouse embryonic data between 9.5 and 13.5 days of gestation and human fetal data from 72 to 129 days. We applied the EncoreDNM method to test genetic correlation across traits both globally and by partitioning genes using highly expressed and uniquely expressed gene information from single cells. Specifically, we found an overall significant correlation enrichment between CHD and CDH (ρ=0.5, p=8.1e-5). In the partitioned analysis, we found significant correlations between CHD and CDH in intermediate mesoderm over 50 times of replication (mean ρ=0.81, mean p=3.4e-4) and foregut (mean ρ=0.80, mean p=4.3e-4), and correlation enrichment in the stomach (mean ρ=0.81, mean p=3.3e-4) and Liver (mean ρ=0.70, mean p=6.0e-6). To identify specific genetic factors, we applied M-DATA to conduct multi-trait analyses among the curated cohorts. By jointly analyzing CHD and CDH, we identified three additional genes for CHD (*MYRF*, *BRAF*, *PPL*), and five additional genes for CDH (*PTPN11*, *POGZ*, *RAF1*, *GATA6*, *KDM5B*). Among these genes, *BRAF*, *KDM5B*, *RAF1*, and *POGZ* were highly expressed across all tissues, and *PTPN11*, *GATA6*, and *MYRF* were highly expressed in intermediate mesoderm and foregut. *PPL* was uniquely expressed exclusively in the intermediate mesoderm, and *MYRF* displayed both high and unique expression in the stomach. These results may shed light on the common etiology of birth defects and motivate the development of statistical methods that can integrate DNVs and single-cell expression data from developing tissues.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4289 Multi-trait GWAS for diverse ancestries : Mapping the knowledge gap

Authors:

H. Julienne[1], L. Troubat[2], H. Aschard[1]; [1]Inst. Pasteur, Paris, France, [2]AP-HP, Assistance Publique - Hôpitaux de Paris, Paris, France

Approximately 95% of samples analyzed in univariate genome-wide association study (GWAS) are of European ancestry. As genetic hindsights gained on European ancestry generalize only partially to other ancestries, this bias could result in increased health disparities. While the genetic community increasingly recommends including more individuals from non-European ancestries in GWAS studies, bias in method development has received less attention. The majority of methods in statistical genetics are developed on European ancestry only (65% of recent publications we surveyed). This European by default mode creates a snowballing effect where secondary analysis increases further the genetic knowledge gap between European and non-European ancestries. Yet, existing data in non-European populations, often of modest sample size, could benefit from innovative approaches maximizing statistical power.

In this context, we tested if the properties of our pipeline for conducting multi-trait GWAS (JASS, Joint Analysis of Summary Statistics) - that is the ability to detect biologically relevant associations missed by univariate GWAS without increasing inflation(1) - would generalize to non-European ancestries. To this end, we conducted the joint GWAS of 19 hematological traits and glycemic traits accross five ancestries (European (EUR), admixed American (AMR), African (AFR), East Asian (EAS), South-East Asian (SAS)).Results:

We detected 367 new genome-wide significant associations in non-European populations (15 in AMR, 72 in AFR and 280 in EAS), which represents respectively 7%, 25% and 21% of all associations in the AFR, AMR and EAS populations.

Overall, multi-trait testing increases the replication of European associated loci in non-European ancestry by 15%. Pleiotropic effects were highly similar at significant loci across ancestries (e.g. the median correlation between multi-trait genetic effect of EUR and EAS was 0.88). For hematological traits, strong discrepancies in pleiotropic effects are tied to known evolutionary divergences: the ARKC1 loci which is adaptive to overcome the p.vivax induced malaria.

Altogether these analyses suggest that multi-trait GWAS methods can be a valuable tool to narrow the genetic knowledge gap between European and non-European populations. (Funding: ANR-20-CE36-0009)

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4290 † Multivariable Mendelian Randomization adjusting for heritable confounding suggests that C-reactive protein is a biomarker but not causal for multiple complex diseases

Authors:

R. Shi, J. Morrison; Univ. of Michigan, Ann Arbor, MI

C-reactive protein (CRP) is a marker of inflammation found to be associated with immune response, cardiovascular disease, and neuropsychiatric disorders in observational studies. However, it is unknown whether CRP causally affects these traits, or is correlated due to reverse causation or unobserved confounding. Mendelian randomization (MR) is a method for estimating causal effects using genetic variants as instrumental variables, thereby avoiding the biases of observational studies. However, MR may produce biased results if factors that confound the exposure and outcome are heritable, leading to horizontal pleiotropy. This is a major concern for studies of CRP, because CRP increases in response to inflammation due to many heritable conditions.Multivariable mendelian randomization (MVMR) can be used to eliminate bias due to heritable confounding when GWAS summary data are available for confounders. Prior univariable MR studies of CRP have suggested causal effects of CRP on several traits such as schizophrenia, bipolar disorders and colorectal cancer. In this study, we used MVMR to estimate the causal effect of CRP on 14 outcomes, including cardiovascular and neuropsychiatric diseases, previously linked with CRP in literature. One major challenge of using MVMR to adjust for heritable confounding, is that heritable confounders may be unknown in advance. We combined literature derived confounders with data driven confounder selection, initially considering thousands of traits present in the MRC-IEU GWAS database and finally selecting a small number of candidate confounders. In initial univariable MR analyses, we found evidence for a causal protective effect of increased plasma CRP level on the risk of schizophrenia, Alzheimer's disease, coronary artery disease, and colorectal cancer, consistent with previous MR results and opposite of observational associations. However, after adjusting for heritable confounders, we found no evidence of causal effects of CRP on any of the 14 traits examined. We found that BMI may be a source of heritable confounding for several relationships, with other confounders including triglycerides, smoking, and diabetes. These results suggested that previous univariable MR analyses of CRP may be biased by heritable confounding and that CRP may have fewer causal effects on human diseases than previously suggested. We have also demonstrated a novel approach of confounder selection using public GWAS databases which may increase the number of traits that can be effectively analyzed using MVMR.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4291 Multivariable Mendelian Randomization identifies LDL cholesterol-independent proteins putatively causal for coronary artery disease

Authors:

D. Shungin[1], A. Hartley[2], A. Carter[2], J. Howson[2], J. Bowden[3]; [1]Human Genetics Ctr. of Excellence, Novo Nordisk, Måløv, Denmark, [2]Human Genetics Ctr. of Excellence, Novo Nordisk, Oxford, United Kingdom, [3]Human Genetics Ctr. of Excellence, Novo Nordisk & Univ. of Exeter, Oxford/Exeter, United Kingdom

Coronary artery disease (CAD) remains a significant global health burden and identifying modifiable risk factors beyond low-density lipoprotein-cholesterol (LDL) is critical for improving prevention and treatment strategies. To identify proteins with LDL-independent effects on CAD, we performed a multivariable Mendelian Randomization analysis (MVMR) across GWAS results for 2,932 plasma proteins in up to 48,645 participants from the UK Biobank Pharma Proteomics Project.

We created genetic instruments for 2,019 proteins with cis-protein quantitative trait loci (pQTL) within a +/-500kb region ($P<5x10^{-8}$); for LDL we derived instruments from GWAS in the UK Biobank (N=400,223; European ancestry) while CAD estimates were extracted from the largest published meta-analysis (Aragam et al, 2022, Ncontrols/cases= 984,168/181,522; European ancestry). We applied a recent MVMR method (Batool et al, 2022) that derives principal component instruments from a set of highly correlated SNPs to increase statistical power of analyses in cis-region. We modelled two exposures at a time, a protein and LDL, and repeated this pair-wise analysis for all proteins to derive causal effects for these genetically predicted traits.

We identified multiple proteins with putative causal effects on CAD independent of LDL, i.e. their causal effects are not attenuated after inclusion of LDL in MVMR. Amongst the prioritized proteins we detected IL1RN that is implicated in CAD in addition to LDL, via a non-LDL (inflammatory) mechanism i.e. both IL1RN (odds ratio [OR]=1.06 per 1 standard deviation [SD] increase of IL1RN, 95%CI:1.03-1.11) and LDL are causally related with CAD in MVMR (OR=1.39 per 1SD increase of LDL, 95%CI:1.30-1.47). Furthermore, we validated the known effects of drug targeting APOBR on CAD which are driven primarily by LDL (OR=1.39 per 1 SD increase of LDL, 95% CI:1.30-1.48), rather than protein levels (OR=1.00 ,95% CI:0.99-1.01 for 1-standard deviation increase of APOBR). Follow-up sensitivity univariable MR analyses for prioritized and non-prioritized proteins further estimated the putative causal effects of each protein on CAD when not accounting for LDL, underscoring the utility of MVMR in estimating effects of related exposures with shared genetic predictors.

Overall, these findings shed light on alternative pathways involved in CAD pathogenesis and provide valuable insights for development of targeted therapies. Further validation of these causal relationships and identification of underlying mechanisms will pave the way for interventions aimed at reducing CAD risk in the population.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4292 † Multivariable Mendelian randomization reveals potential causal genes that contribute to blood pressure traits in diverse populations

Authors:

Y. Yang[1], N. Lorincz-Comi[1], X. Zhu[2]; [1]Case Western Reserve Univ., Cleveland, OH, [2]Case Western Reserve Univ., Shaker Heights, OH

High blood pressure (BP) is a major risk factor for cardiovascular disease. Identifying genes that cause BP variation is crucial for uncovering molecular mechanisms and discovering drug targets. Popular non-experimental methods to investigate causality - such as colocalization, TWAS, and univariable Mendelian randomization (UVMR) - leverage summary-level gene expression data but are not robust because of high correlations among gene expressions in a locus. It is also unclear whether causal genes discovered in European (EUR) populations are transferrable to non-EUR populations. To address these issues, we performed multivariable Mendelian randomization (MVMR) to robustly identify causal genes influencing systolic and diastolic BP (S/DBP) across diverse populations. We implemented MVMR using MR Joint Outliers-aNd-Exposures Selection (Mr.Jones), which was recently developed by our group to prioritize causal genes and identify horizontal pleiotropy with variable selection penalties. We used summary-level GWAS data for eQTLs in blood tissue from the eQTLGen consortium (n=30K) and for BP traits from Surendran et al. (n=800K), BioBank Japan (n=150K), and Liang et al. (n=30K) for EUR, East Asian (EAS), and African (AFR) populations, respectively. We targeted the top 10 most significant loci in the European SBP GWAS and searched for causal genes in these loci across populations. Each locus was defined as a 2Mb window centered around the lead SNP. In EUR, we examined 295 genes and identified 28 and 27 genes with P<5E-8 for SBP and DBP, respectively. In EAS and AFR, we analyzed 279 and 232 genes, respectively, resulting in the discovery of 6 and 3 genes for SBP, and 6 and 2 genes for DBP. Gene expressions in the top 10 loci explained 0.63 and 0.68 of local heritabilities for SBP and DBP in EUR. In EAS and AFR, the fractions were 0.30, 0.22, 0.20, and 0.17, respectively. Our analysis also resolved many inconsistencies in the literature, particularly with regard to TRAFD. Despite historically inconsistent effect directions found in the TWAS hub across six studies, our analysis observed TRAFD as a significant risk gene for SBP with Z-scores > 20. By leveraging eQTL and GWAS data, our MVMR analysis holds promise for discovering novel genes causally associated with BP traits. In subsequent analyses, we plan to scan the whole genome in search of causal BP genes and examine their functional enrichments.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4293 Native Hawaiians and Pacific Islanders: Is APOE a Biomarker of Disease?

Authors:

J. Tavana[1], R. J. Hansen[1], B. J. Vaitohi[1], C. S. Pakalani[1], A. T. Perez[1], F. Matthes[2], S. Polaapau[2], S. F. Funaki[2], T. D. Packard[1], J. S. Kauwe[1], P. G. Ridge[1]; [1]Brigham Young Univ., Provo, UT, [2]Brigham Young Univ. Hawaii, Laie, HI

Native Hawaiians and Pacific Islanders (NHPI) are among the least represented racial groups in Alzheimer's disease (AD) research (for example, there are only 36 samples in the NACC). The prevalence of AD is unknown in NHPIs. Likewise, no large studies have reported disease biomarkers, including assessing APOE genotypes for correlation with AD. In one small study in Guam natives, APOE e4 was not correlated with AD. We have enrolled 648 NHPIs in our cohort with AD8, CDR scores, or NACC cognitive tests. We have collected APOE genotypes for 234 individuals.

Study participants were recruited during brain health educational workshops that utilize visual and interactive learning. During these workshops, participants are invited to provide a blood or saliva sample after signing an IRB approved consent form. Samples are then processed in our lab and APOE genotypes are collected.

Table 1 summarizes the cognitive assessments and APOE genotypes for 234 individuals. APOE allele frequencies vary widely based on race. However, in our dataset, the e4 frequency in NHPI females and males is the same and higher than reported in other races.

We are underpowered to detect significant trends in subjects with and without dementia (Table 2). Our preliminary data reveal e4 allele frequencies in subjects with and without dementia. No trend towards e4 differences in impaired versus normal subjects.

We are augmenting and refining the existing cohort. We are genotyping APOE in 414 additional NHPIs and formally diagnosing each subject. We will have sufficient samples to determine whether or not APOE correlates with AD status. If APOE is not a biomarker of AD in NHPIs, this suggests that the genetic architecture of AD in NHPIs is unique and demonstrates the importance of prioritizing NHPIs in future AD research.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4294 Network diffusion-based risk scoring models for coronary artery diseases in UK biobank individuals: Fusing polygenic impact of intermediate clinical factors on complex disease

Authors:

Y. Nam[1], S-H. JUNG[1], K. Jaeyoung[2], J-S. Yun[3], D. Kim[1]; [1]Univ. of Pennsylvania, Philadelphia, PA, [2]Sungkyunkwan Univ., Seoul, Korea, Republic of, [3]The Catholic Univ. of Korea, Seoul, Korea, Republic of

Background: Intermediate clinical risk factors, such as biomarkers derived from urine- or blood-based laboratory results, have been studied for their genetic or clinical interactions with complex diseases. While composite polygenic risk scores (PRSs) have demonstrated the increased accuracy in risk prediction for common diseases by aggregating the genetic predisposition of multiple intermediate risk factors, there are limitation in leveraging extensive genetic relationships in scoring. Genetic correlation networks can help elucidate the complex relationships between phenotype and clinical factors. A combined model, incorporating extensive relationships from correlation networks, is likely to enhance the ability to stratify high-risk individuals compared to standard PRS or linear combinations of independent scores. **Methods:** We developed a network-based composite risk scoring model (net-cPRS) that aggregates the propagated polygenic impact of intermediate phenotypes on complex disease from networks. Initially, we built a genetic correlation network using linkage disequilibrium score (LD scores) with GWAS summary statistics across the phenotype of interest (binary trait) and intermediate phenotypes. Subsequently, label propagations were applied to estimate the polygenic impact on the phenotype by diffusing the query on the binary trait to clinical factors. The estimated impacts depict the direction and magnitude of genetic impacts at the summary level and allow for the combination of multiple risk scores for intermediate clinical factors using pre-defined coefficients in regression models. A fine-tuning process was performed to estimate individual risk scores using individual genetic profiles. **Results:** As a proof-of-concept study, we developed net-cPRS for coronary artery diseases (CAD) using data from 370K European individuals in UK Biobank. Genetic correlation networks were built using LD scores between independent GWAS summary of CAD and 12 intermediate factors, including BMI, HDL/LDL-C, TG, BP, FG, HbA1C and eGFR. Network-diffused genetic impacts for CAD were obtained for 12 intermediate factors and applied as weights in a logistic model. The scoring model was fine-tuned using individual genetic profiles with 5-fold cross-validations. In terms of prediction accuracy, the average AUC of net-cPRS was 0.747, demonstrating improved prediction ability compared to standard PRS (0.65) and composite PRS (0.721). **Conclusion:** Proposed net-cPRS, which incorporates genetic relationships, can improve the prediction ability of PRS models. However, further experiments are needed to apply net-cPRS to additional phenotypes.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4295 Network-based approach to gene prioritization at GWAS loci.

Authors:

L. Martini[1], B. Hobbs[2], A. Fazzone[3], L. Becchetti[4], M. Gentili[1], M. Cho[5]; [1]Harvard, Boston, MA, [2]Brigham and Women s Hosp., Needham Heights, MA, [3]Centai, Turin, Italy, [4]Sapienza, Roma, Italy, [5]Brigham Women's Hosp., Boston, MA

Translation of risk loci identified by genome-wide association studies to disease-causing genes and biologic pathways is challenging. Often, researchers attempt to identify a causal gene at each locus using a series of bioinformatics methods. However, the genetic risk to most diseases is likely distributed across a large number of biologic processes with a network of molecular interactions contributing to disease biology. Considering each locus in isolation does not leverage the relationship between genes associated across loci.
In this work, we propose the Relations-Maximization Method (RMM-GWAS), a new network-based approach to identify effector genes from a GWAS. Our method builds on the hypothesis that disease-relevant genes across GWAS loci share biological processes, pathways, and correlated gene expression levels. It integrates network topology and gene biological information to prioritize a set of genes, one per disease locus, such that the overall number of shared interactions is maximized.
We assessed the performance of our framework, comparing it with well-established baselines: dmGWAS, SigMod, LEAN, DEPICT and DOMINO. We compared them along three different axes: i) prediction of FDA-approved drug targets, ii) prediction of genes that have been previously identified to be involved in murine disease-related phenotypes, and iii) functional analysis of the set of statistically significant Reactome pathways associated with the prioritized gene sets. We compared them on the COPD GWAS and we found RMM-GWAS to be the only heuristic to prioritize genes that are targets of FDA approved drugs. Furthermore, genes prioritized by RMM-GWAS are enriched in pathways and biological processes previously linked to COPD including inflammation and immune response, WNT/beta-Catenin/TCF/FZD4 interactions, and C-type lectin receptor signaling.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4296 Neuroimaging PheWAS of Essential Tremor Genes: Matching Genes to Neuromorphology

Authors:

X. Bledsoe[1], E. Gamazon[2]; [1]Vanderbilt Univ., Nashville, TN, [2]VUMC Clare Hall, Univ. of Cambridge, Nashville, TN

Essential tremor (ET) is a progressive and debilitating condition characterized by motor tremor in the hands and arms that can eventually spread to impair head motion, mastication, and speech. Significant progress has been made in assessing the joint molecular and neurophysiologic causes of ET. At a neural level, current data suggest that ET results from a dysfunctional "tremor pacemaker" in the olivary nucleus operating in concert with cerebellothalamocortical circuits. At a molecular level, GWAS data and expression analyses have highlighted variation in DNA and RNA transcripts that correlate with disease status. While these data are meaningful contributions, it is difficult to discriminate between brain measures that drive the disease from those which result from it.

We address this gap by first performing a transcriptome-wide association study (TWAS) of ET by applying the JTI-PrediXcan methodology to summary statistics from the largest GWAS of the condition to date. I performed a transcriptome based, tissue-specific neuroimaging phenome wide association study (PheWAS) to identify brain measures that are putatively impacted by the gene expression profile associated with ET. For each gene that was significantly associated with ET (Bonferroni threshold of 0.05), I identified the associated cortical and subcortical neuroimaging measures, matching on tissue context (study-wide false discovery rate ≤ 0.05).

The TWAS identified 90 different gene products for which genetically regulated gene expression is associated with ET. These include top loci from the initial publication as well as novel genes. From the neuroimaging PheWAS, 14 of the top 15 neuroimaging measures reflect structural connectivity measures. The three most impacted measures are the left and right inferior cerebellar peduncles and the middle cerebellar peduncle. These findings were driven by altered expression of *BACE2* and *LINC00323.* Additionally, the bilateral corticospinal tract and posterior limbs of the internal capsule are each represented in the top 15 associations.

Our TWAS data replicate and extend the set of transcript-level associations reported by Liao et al. Findings from the neuroimaging PheWAS replicate prior work highlighting the cerebellar peduncles as regions involved in ET pathophysiology and then show that expression of BACE2 and LINC00323 are associated with these structures. Additionally we capture associations between ET gene transcripts and 4 neuroimaging measures of the descending motor tract. These measures indicate that gene expression alterations that occur secondary to ET SNPs are associated with motor neuroanatomy in patients without overt disease.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4297 New insights into the genetic ethology of 57 essential and non-essential trace elements in humans

Authors:

M. Moksnes[1], A. F. Hansen[1], B. Wolford[2], L. Thomas[2], H. Rasheed[3], A. Simic[1], L. Bhatta[1], A. Brantsæter[4], I. Surakka[5], W. Zhou[6], P. Magnus[4], P. R. Njølstad[7], O. Andreassen[8], T. Syversen[1], J. Zheng[9], L. Fritsche[10], D. Evans[11], N. Warrington[12], T. Nøst[2], B. ÅSvold[2], T. Flaten[1], C. Willer[13], K. Hveem[1], B. Brumpton[2]; [1]NTNU, Trondheim, Norway, [2]Norwegian Univ. of Sci. and Technology, Trondheim, Norway, [3]NTNU, Norway, Trondheim, Norway, [4]Norwegian Inst. of Publ. Hlth., Oslo, Norway, [5]Univ. of Michigan, Jackson, MI, [6]Broad Inst. of Harvard and MIT, Cambridge, MA, [7]Univ. of Bergen, Bergen, Norway, [8]Oslo Univ. Hosp., Oslo, Oslo, Norway, [9]Univ. of Bristol, Bristol, United Kingdom, [10]Univ. of Michigan Sch. of Publ. Hlth., Ann Arbor, MI, [11]Univ Queensland, St Lucia, Australia, [12]Univ Queensland, Brisbane, Australia, [13]Regeneron, Tarrytown, NY

Trace elements are important for human health but may exert toxic or adverse effects. Mechanisms of uptake, distribution, metabolism, and excretion are partly under genetic control but have not yet been extensively mapped. Here we report a comprehensive multi-element genome-wide association study (GWAS) of 57 essential and non-essential trace elements. We performed GWA meta-analyses of 14 trace elements in up to 6564 Scandinavian whole-blood samples, and GWASs of 43 trace elements in up to 2819 samples measured only in the Trøndelag Health Study (HUNT). We identified 11 novel genetic loci associated with blood concentrations of arsenic, cadmium, manganese, selenium, and zinc in genome-wide meta-analyses. In HUNT, several genome-wide significant loci were also indicated for other trace elements. Using two-sample Mendelian randomization, we found several indications of weak to moderate effects on health outcomes, the most precise being a weak harmful effect of increased zinc on prostate cancer. However, independent validation is needed. Our new understanding of trace element-associated genetic variants may help establish consequences of trace elements on human health.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4298 New insights into thyroid hormone regulation: a transcriptome and proteome-wide association study.

Authors:

S. Monteiro-Martins[1], R. Sterenborg[2,3], M. Medici[2,3], A. Köttgen[1,4], A. Teumer[5,6], on behalf of the ThyroidOmics Consortium; [1]Univ. of Freiburg, Freiburg, Germany, [2]Radboud Univ. Med. Ctr., Nijmegen, Netherlands, [3]Erasmus Med. Ctr., Rotterdam, Netherlands, [4]Johns Hopkins Univ., Baltimore, MD, [5]Univ. Med. Greifswald, Greifswald, Germany, [6]DZHK (German Ctr. for Cardiovascular Res.), Partner Site Greifswald, Greifswald, Germany

Thyroid hormones, including thyrotropin (TSH) and free thyroxine (FT4), are regulated via the hypothalamic-pituitary-thyroid (HPT) axis and play a crucial role in regulating metabolic processes in humans. We aimed to better understand molecular mechanisms underlying TSH and FT4 levels through tissue-specific transcriptome-wide association studies (TWAS) and plasma protein-wide association studies (PWAS). TWAS was performed based on HPT-relevant tissues from GTEx v8, and PWAS on plasma protein levels measured by the Somalogic v4.0 assay, using GWAS-summary statistics for TSH and FT4 levels from up to 271,040 European ancestry individuals in the ThyroidOmics Consortium. Statistical significance was set at $p<1.74\times10^{-06}$ (TWAS) and $p<3.78\times10^{-05}$ (PWAS). The relationship between eQTLs, pQTLs, and trait GWAS summary statistics was analyzed through conditional colocalization analysis. TWAS identified 297 and 113 transcripts associated with TSH and FT4, respectively, with 25 associated to both. Of these, 24 for TSH and 14 for FT4 were not in loci revealed by the underlying GWAS, demonstrating the increased power of our approach. Colocalization revealed a shared genetic basis between mRNA levels ($PPH_4>0.8$) with TSH for 158 genes and with FT4 for 45 genes. Interestingly, there were tissue-specific colocalizations between CAPZB expression and TSH levels related to independent genetic variants in different tissues, including a putative regulatory variant in the thyroid.PWAS identified 18 and 7 proteins associated with TSH and FT4, respectively, with 2 shared proteins, HEXIM1 and QSOX2. Of these, 5 TSH and 7 FT4-associated proteins were not encoded within significant GWAS loci. Colocalization was observed for 5 plasma protein levels with TSH and 5 with FT4. Comparing results from PWAS and TWAS, 10 TSH genes and 1 FT4 gene were significant in both. Of these, ANXA5 expression was inversely associated with TSH in all studied tissues and plasma protein levels (TWAS-$p=7.61\times10^{-12}$ [whole blood], $p=6.40\times10^{-13}$ [hypothalamus], $p=1.57\times10^{-15}$ [pituitary], $p=4.27\times10^{-15}$ [thyroid], PWAS-$p=1.18\times10^{-13}$), supported by colocalization. Our analyses revealed new thyroid function-associated genes and prioritized candidates in known GWAS loci, contributing to a better understanding of transcriptional regulation and protein levels relevant to thyroid function.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4299 Next-generation map of constrained coding regions from hundreds of thousands of humans

Authors:

S. Lulla, P. McHale, B. Pedersen, A. Quinlan; Univ. of Utah, Salt Lake City, UT

A longstanding interest in human genetics has been to identify the subset of the human genome that is critical to normal human development. In pursuit of this question, we previously created a map of constrained coding regions (CCRs) [1], or regions where the absence of genetic variation implies negative selection due to critical function or disease pathology. We defined a region as the coding distance between protein-changing variants, and we weighted region lengths by sequencing coverage. We then quantified the degree of constraint of each region, where regions with the greatest distance between protein-changing variants have the highest predicted constraint. This map provided a more fine-scale metric of intraspecies constraint than other previously established gene-wide measures and nominated new protein-coding regions that are essential to human development.

We will present progress on a more rigorous approach to modeling constraint that calculates a ratio of observed to expected constraint in small sliding windows across protein-coding exons. Such an approach allows us to relax the original requirement of the complete absence of protein-changing variants in a CCR to capture cases where genetic variation in a region is present but extremely sparse, which will occur more often as cohort sizes increase. The gnomAD v2 and forthcoming population-scale datasets provide unprecedented deep samples of human coding variation, allowing us to construct a higher-resolution map of constraint and assess which CCRs from the previous dataset survive given the expanded dataset. We anticipate that our efforts will yield a much higher resolution map of constrained coding regions in the human genome, providing a valuable resource for rare disease interpretation and identifying coding sequence that is essential to human development.

References: 1. Havrilla JM, Pedersen BS, Layer RM, Quinlan AR. A map of constrained coding regions in the human genome. Nat Genet. 2019;51: 88-95.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4300 *NODAL* variants are associated with a continuum of laterality defects from simple D-transposition of the great arteries to heterotaxy.

Authors:

Z. Dardas[1], J. Fatih[1], A. Jolly[1], M. Dawood[1], H. Du[1], C. Grochowski[1], E. G. Jones[2], S. Jhangiani[3], X. Wehrens[1], P. Liu[1], W. Bi[1], E. Boerwinkle[4], J. Posey[1], D. Muzny[3], R. Gibbs[3], J. Lupski[1], Z. Coban Akdemir[5], S. Morris[2]; [1]Baylor Coll. of Med., Houston, TX, [2]Texas Children's Hosp., Houston, TX, [3]Baylor Coll. Med., Houston, TX, [4]Univ. of Texas Hlth.Sci. Ctr. at Houston, Houston, TX, [5]UTHlth.Sch. of Publ. Hlth., Houston, TX

Background: NODAL signaling plays a critical role in embryonic patterning and heart development in vertebrates. Genetic variants resulting in perturbations of the TGF-β/NODAL signaling pathway have reproducibly been shown to cause laterality defects in humans. To further explore this association and improve genetic diagnosis, the study aims to identify and characterize a broader range of *NODAL* variants in a large number of individuals with laterality defects. **Methods and Results:** We re-analyzed a cohort of 307 proband-only exomes of individuals with clinically diagnosed laterality congenital heart disease (CHD) using family-based, rare variant genomic analyses. For those with candidate contributory variants, variant allele confirmation and segregation analysis were studied by Sanger sequencing in available family members. To this cohort, we added 12 affected subjects with known *NODAL* variants and CHD from institutional research and clinical cohorts to investigate an allelic series. Missense, nonsense, splice site, indels, and/or structural variants of *NODAL* were identified as potential causes of heterotaxy and other laterality defects in 33 CHD cases. We describe a recurrent complex indel variant for which the nucleic acid secondary structure predictions suggest secondary structure mutagenesis as a possible mechanism for formation. We identified two copy number variant (CNV) deletion alleles spanning *NODAL* in two unrelated CHD cases. Furthermore, 17 CHD individuals were found (16/17 with known Hispanic ancestry) to have the c.778G>A:p.G260R *NODAL* missense variant which we propose reclassification from a variant of uncertain significance (VUS) to likely pathogenic. Quantitative human phenotype ontology-based analyses of the observed clinical phenotype for all cases with p.G260R variation, including heterozygous, homozygous, and compound heterozygous cases, reveal clustering of individuals with biallelic variations. This finding provides evidence for a genotypic-phenotypic correlation and an allele-specific gene dosage model. **Conclusion**: Our data further support a role for rare deleterious variants in *NODAL* as a cause for sporadic human laterality defects, expand the repertoire of observed anatomical complexity of potential cardiovascular anomalies, and provide evidence that the population specificity of *NODAL* variants should be taken into consideration in genetic counseling and clinical genomic testing for this condition.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4301 Non-additive interactions of rare variants and lifestyle factors contribute to obesity

Authors:

D. Banerjee[1,2,3], A. Tyryshkina[1,3], S. Girirajan[1,2,3]; [1]Pennsylvania State Univ., State College, PA, [2]Bioinformatics and Genomics Program, Pennsylvania State Univ., University Park, PA, [3]Dept. of Biochemistry and Molecular Biology, Pennsylvania State Univ., University Park, PA

Obesity is a complex disorder where a multitude of genes and environmental factors interact to express the disease phenotype. Efforts till now have primarily concentrated on searching for GWAS implied common variants or monogenic rare variants that increase the disease risk. In this study, we used RareComb, a computational framework that uses apriori algorithm along with binomial tests, to identify oligogenic combinations of two or more genes with rare variants and lifestyle factors that contribute to obesity in a cohort consisting of ~200,000 individuals from UK Biobank. Our results demonstrate that oligogenic combinations of rare variants play a significant role as obesity risk factors, similar to monogenic or polygenic variants. In fact, mean effect sizes of the discovered combinations were ~2.3 times higher than that of previously identified obesity-associated genes with low-frequency mutations (t-test, p=4.36E-10). Additionally, we illustrate how obesogenic lifestyle factors in conjunction with rare variants further exacerbate the disease risk. Individuals carrying combinations of obesogenic lifestyle factors and risk genes had significantly higher mean BMI compared to those without such combinations (t-test, p< 3.04E-53). Our results indicate that musculoskeletal genes, such as *SCN10A* and *TTN*, primarily contribute to obesity risk regardless of lifestyle factors, whereas nervous system genes, such as *APC* and *PRPH*, interact with lifestyle factors to increase the risk. We further identify a novel class of variably expressive genes that either increase the risk of or protect from obesity and its related disorders depending on their interaction with distinct partner genes. Our findings challenge the existing landscape of genetic screening for complex disorders which primarily focuses on monogenic variants or polygenic risk scores and emphasize the importance of considering oligogenic risk when assessing predisposition to complex disorders like obesity.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4302 Non-additive polygenic scores.

Authors:

R. Ohta[1], Y. Tanigawa[2], Y. Suzuki[1], M. Kellis[2], S. Morishita[1]; [1]The Univ. of Tokyo, Kashiwa, Chiba, Japan, [2]MIT, Cambridge, MA

Polygenic scores (PGSs), which combine multiple SNVs to explain a phenotype, have recently received attention due to their potential for contributing to predicting disease risk. Most existing PGS methods exploit the additive model to construct a best-possible linear model of the effect sizes. However, Palmer *et al.* showed that non-additive heritability was estimated to exist in some phenotypes and to be especially significant in autoimmune diseases such as rheumatoid arthritis and psoriasis.

Considering this background, it is promising to build PGSs under additive and non-additive models; however, the traditional PGS methods are designed to rely heavily on additive effects, and it is non-trivial to revise them with non-additive effects. More precisely, the technical difficulty in devising non-additive models is how to assign optimal independent SNV scores to three allelic dosages (0, 1, and 2), which allows us to handle any genetic inheritance models, including the additive, recessive, dominant, and overdominant models. Here, we propose a method that can build PGSs under additive and non-additive models and is also efficient because it computes analytic exact solutions for optimal SNV scores.

Applied to twelve representative phenotypes in UK Biobank, our software program named **GenoBoost** was competitive across all traits, achieved the highest accuracy for five phenotypes, and ranked second for three phenotypes among seven PGS methods. GenoBoost improved Nagelkerke's R^2 by 4.9% on average compared to LASSO. Non-additive GenoBoost had higher accuracy than additive GenoBoost for rheumatoid arthritis and psoriasis, which is consistent with Palmer *et al.*

Among variants in non-additive GenoBoost PGS models, approximately 40-67% of variants were classified as non-additive, which was consistent with GWAS classification. When prioritizing the associated variants under general non-additive models, GenoBoost exploits one metric, SNV utility, as an association indicator to handle general genetic models, unlike GWAS using *p*-values for each model. Focusing on non-additive variants with the eight largest SNV utilities in the PGS models, four out of eight were in the experimentally validated non-additive genes but not reported in GWAS catalog.

In conclusion, we propose the first PGS algorithm that can incorporate general genetic models to the best of our knowledge and outperformed other PGS methods for five out of twelve phenotypes. Non-additive GenoBoost has the ability to classify genetic inheritance models and prioritize the non-additive variants. Our results demonstrate that the non-additive model is essential in predicting risk for polygenic diseases.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4303 Non-APOE genetic risk scores for late-onset Alzheimer disease in a diverse clinical population

Authors:

C. Rose, S. Liu, W. S. Bush, D. C. Crawford; Case Western Reserve Univ., Cleveland, OH

Alzheimer disease (AD), a leading cause of death in the US, is the most common cause of dementia. Most known risk factors such as age, sex, and genetics are not modifiable. Large genome-wide association studies (GWAS) in mostly European-descent populations have identified >75 genetic variants associated with late-onset AD (LOAD). Few GWAS and risk scores derived from them include diverse populations. In the present study, we accessed the Epidemiologic Architecture for Genes Linked to Environment (EAGLE) BioVU, a resource of de-identified electronic health records (EHRs) linked to Illumina Metabochip genotypes assayed on DNA samples extracted from clinical blood draws at a major medical center in Davidson County, Tennessee. From the approximately 11,000, 1,000, and 1,500 African American (AA), Asian American (A), and Hispanic (H) patients in EAGLE BioVU, respectively, we identified a total of 77 cases of LOAD using International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) code 331.0. Controls were defined as absence of the case-defining ICD-9-CM code for patients ≥65 years of age at last clinic visit (n=3,752). We queried the Polygenic Score (PGS) catalog and the largest GWAS for European-descent and African-descent populations in the literature for LOAD-associated ($p<5x10-8$) single nucleotide variants (SNVs), effect alleles, and genetic effects (beta coefficients). Of the 126 SNVs considered, seven were assayed directly by the Metabochip. Tests for Hardy-Weinberg Equilibrium ($p<0.00005$) removed a rare SNV (rs6656401in AA) and rs889555 (in A). Unweighted and weighted risk scores were calculated for each patient stratified by population and case-control status. The unweighted risk scores ranged from [0,8], [0,6], and [0,7] with means (±standard deviations or SD) 2.44 (±1.12), 2.07 (±1.32), and 2.11 (±1.32) for AA, A, and H, respectively. The mean (±SD) weighted risk scores were 0.165 (±0.105), 0.142 (±0.107), and 0.178 (±0.149) for AA, A, and H, respectively. When stratified by case-control status, the unweighted risk scores ranged from [0,5] and [0,8] for cases and controls, respectively. A test of association between case status and genetic risk score was performed using logistic regression across all populations. Genetic risk scores were not associated with LOAD in this diverse patient population (OR=0.468; 95% CI: 0.046-3.688; p=0.498). Work is ongoing to develop risk scores for non-LOAD dementia (vascular dementia, dementia with Lewy body, frontotemporal dementia) as well as to develop additional computable phenotypes for LOAD and related dementias leveraging ontologies in this diverse clinical population.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4304 Novel methods for estimating risk parameters associated with polygenic scores using case-parent trio designs

Authors:

Z. Wang, D. Ray, L. Grosvenor, T. Beaty, H. Volk, C. Ladd-Acosta, N. Chatterjee; Johns Hopkins Univ., Baltimore, MD

Background Hidden population stratification in GWAS of unrelated individuals can lead to overestimation of genetic effects, complicating the applications and interpretations for polygenic scores (PGS) in many settings, including risk prediction and Mendelian randomization. Family-based studies, where genetic effects are estimated through within-family comparisons, provide an opportunity to test for PGS-trait associations while correcting for potential population stratification bias.
Methods We develop a likelihood-based method for estimating genetic effects, parental nurturing effects, and gene-environment interactions using case-parent trios. We assume the disease risk follows a log-linear model and PGS follows a normal distribution, allowing for family-specific terms in both components to account for arbitrary population structures. We show that the likelihood calculations can be simplified into a parental PGS component and a transmission-based likelihood, from which genetic effect estimates can be derived in closed forms. We apply the proposed method to multi-ancestry case-parent-trio studies of non-syndromic orofacial clefts (OFCs) (GENEVA study, European ancestry (EUR) N_{trio} = 778, East Asian ancestry (EAS) N_{trio} = 1126) and autism (SPARK study, multi-ancestry N_{trio} = 1517) to characterize the risk of these childhood disorders associated with established PGS (OFCs: 24 SNPs, autism: 35087 SNPs) and their interactions with several maternally mediated risk factors.
Results Simulation studies demonstrate this proposed method can produce unbiased effect estimates and correct coverage probabilities under complex population structures. In contrast, the polygenic transmission disequilibrium test (Weiner et al., Nat Genet 2017), while producing a valid test for genetic effects, yields attenuated estimates of effect-size due to the use of an incorrect scaling factor. Data analyses using our proposed method indicate transmission-based estimates of PGS effects on the risk of autism (RR_{EUR} = 1.33 per SD change, 95% CI = [1.22,1.44]) and OFCs (RR_{EUR} = 1.73 per SD change, 95% CI = [1.58,1.89]; RR_{EAS} = 1.59 per SD change, 95% CI = [1.46,1.73]) are similar to those reported from population-based studies. Further for OFCs, we observe evidence of maternally mediated nurturing effect (p = 0.03) and gene-environment interaction with maternal primary and environmental smoking status (p = 0.0025 and p = 0.024) in EUR.
Conclusion We provide a novel and comprehensive framework for estimating a collective of risk parameters associated with PGS using case-parent trios. An R package is publicly available on Github.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4305 Novel multi-omics analysis methods for the prioritization of therapeutic targets: a versatile platform

Authors:

N. Raghavan[1], A. Musolf[1], J. Miller[1], Y. Yan[1], K. Shkura[2], B. Guo[1], J-H. Sul[1], S. Longerich[1], D. Fagegaltier[1], M. Kennedy[1], A. Tamburino[1], R. Mathew[1]; [1]Merck & Co., Inc., Cambridge, MA, [2]MSD, London, United Kingdom

For many diseases, disparate multi-omics datasets that provide evidence for potential disease-associated genes and proteins are publicly available. However, deriving insights on genes/proteins to prioritize for therapeutics is difficult due to a general lack of integration of these data. We created a framework for therapeutic target prioritization with an integrated -'omics forward approach.

Multiple lines of evidence including statistical genetics approaches followed by functional genomics have highlighted the crucial role specific cell types play in disease and pathology. We leveraged high quality, disease-relevant, public multi-omics data and created a ranking scheme of genes expressed in cell types of interest. Genetics, epigenetics, transcriptomics, and proteomics data were used to score each gene. Genetic and epigenetic data were used to infer the evidence for causality of the gene to disease and of disease relevant chromatin accessibility in these cells. Differences in protein and RNA levels may be indicative of the cause or effect of the disease process. Thus, gene expression, protein level changes in disease-affected regions and/or cell type specific evidence that may be very relevant to the disease were scored highly. Scores were generated within each datatype, scaled, and hypothesis-driven score integration methods were applied.

Final scores prioritized known genes implicated in the disorder, suggesting proper calibration of the score. We also identify novel genes that have roles in disease-implicated processes but with limited research available in published literature, demonstrating the high value of the method to discover novel targets. This framework can integrate proprietary proteomic and disease-specific assay data, along with metrics that reflect the ability to target the protein through known therapeutic modalities. This scoring schema is flexible: with appropriate modifications relevant to the disease pathophysiology and data availability this framework is applicable across a range of human diseases to prioritize genes for drug discovery.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4306 omicsIMC: A Comprehensive Benchmarking Platform for Robust Comparison of Imputation Methods in Mass Spectrometry-based Omics Data

Authors:

W. Lin[1], K-j. Su[1], C. Qiu[1], Q. Tian[2], Z. Luo[1], H. Shen[3], H. Deng[1]; [1]Tulane Univ., New Orleans, LA, [2]Tulane Univ., Mandeville, LA, [3]Tulane Univ, New Orleans, LA

The analysis of mass spectrometry-based omics data such as lipidomics, is pivotal for elucidating biological processes and identifying biomarkers. However, missing information presented in spectrometry-based omics data is a critical challenge, potentially undermining downstream analyses which will hamper the comprehension of biological processes and the identification of biomarkers. To solve this issue, much effort has been done and several imputation methods have been developed with three categories: conventional, model-based and machine learning-based, however comprehensively and systematically comparing existing methods is still lacking, especially for lipidomics data. In this work, we compared 21 distinct imputation methods tailored for handling missing values in simulated spectrometry-based omics data and real lipidomics data from three aspects: (i) differential analysis, (ii) clustering, and (iii) functional enrichment analysis. Based on our simulation results, the machine learning-based methods generally show better performance than others with higher prediction accuracy. Notably, we built omicsIMC (mass spectrometry-based omics Imputation Methods Comparison platform), the first platform that provides researchers with a versatile framework for simulating and evaluating a diverse range of imputation strategies, encompassing machine learning, model-based, and conventional approaches. The strength of omicsIMC lies in its ability to address missing data challenges across diverse mass spectrometry-based omics datasets. By facilitating informed decision-making in imputation method selection, omicsIMC can empower researchers to conduct reliable and accurate downstream analyses using their mass spectrometry-based omics data. The comprehensive benchmarking and versatility of omicsIMC will make it a valuable tool for the scientific community engaged in mass spectrometry-based omics research.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4307 One third of genetic loci associated with coronary artery disease, including *CDKN2B-AS1*, are independent of known heritable clinical risk factors.

Authors:

C. P. Barnhill[1], B. W. Dulken[1,2], M. M. Soliai[1], J. B. Chang[1], S-G. Ji[1]; [1]BridgeBio Pharma, Inc., Palo Alto, CA, [2]Dept. of Pathology, Stanford Sch. of Med., Stanford, CA

Despite advances in the standard of care, atherosclerotic cardiovascular disease is the leading cause of morbidity and mortality in the United States and worldwide. One population with unmet need is those with advanced atherosclerosis and myocardial infarction (MI), but no standard modifiable risk factors (SMuRF-less, 14-27% of individuals with ST-elevated MI). Previous genome-wide association studies (GWAS) have identified over 150 loci associated with coronary artery disease (CAD) risk. Additionally, Mendelian randomization studies have revealed causal associations for risk factors such as hypertension, diabetes, and blood lipids. We sought to determine which of these loci confer risk independent of known heritable clinical risk factors in order to prioritize loci for SMuRF-less CAD. However, two challenges emerged: first, limiting to a sample without any of the risk factors would drastically reduce our power to detect a signal; second, directly conditioning on heritable risk factors as covariates in a GWAS could potentially introduce collider bias. To circumvent these issues, we conducted a multitrait-based conditional and joint association analysis (mtCOJO). GWAS summary statistics from the UK Biobank contain 92 genome-wide significant loci associated with CAD. We applied mtCOJO to remove the effects of 14 known heritable risk factors and found that only a third of loci remained genome-wide significant (n = 32). Among the loci that no longer held significance were genes known or suspected to be associated with circulating lipoprotein levels, such as *LPA* and *PSRC1*. *CDKN2B-AS1* was the most significant locus for CAD risk, maintaining its significance after conditioning without a significant heterogeneity of effect ($p_{initial} = 1.01 \times 10^{-73}$, $p_{conditional} = 6.59 \times 10^{-76}$, $p_{heterogeneity} = 0.26$). We applied two statistical fine-mapping tools, FINEMAP and SuSiE, to variants in the *CDKN2B-AS1* locus, which yielded two independent credible sets of variants. The lead variant in the first set, rs1333042, is associated with an elevated risk of CAD (OR = 1.26, $p = 6.30 \times 10^{-116}$) and increased splicing of exon 1 to exon 5 ($p = 1.70 \times 10^{-6}$) in *CDKN2B-AS1* transcripts in GTEX. The lead variant in the second set, rs3731239, is associated with protection from CAD (OR = 0.87, $p = 8.90 \times 10^{-38}$), reduced expression of *CDKN2B-AS1* ($p = 1.40 \times 10^{-5}$) in GTEX, and increased expression of *CDKN2B* ($p = 3.60 \times 10^{-25}$) in eQTLgen. In summary, we propose a novel approach to prioritize loci for SMuRF-less CAD and suggest that splicing of *CDKN2B-AS1* and its impact on *CDKN2B* expression constitute a significant genetic factor contributing to residual CAD risk, independent of known heritable risk factors.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4308 Optimizing variant impact prediction in autoinflammatory disease

Authors:

B. Matthiasardottir[1,2], S. Blackstone[1], S. Mount[2], D. Kastner[1]; [1]NIH, Bethesda, MD, [2]Univ. of Maryland, College Park, MD

Background: Autoinflammatory diseases are immunological disorders caused by dysregulation of the innate immune system, characterized by seemingly unprovoked attacks of inflammation without evidence of infection, high-titer autoantibodies, or antigen-specific T cells. The identification of genes underlying these diseases has revolutionized our understanding of innate immunity and provided the basis for life-saving therapies. We have sequenced the exomes of approximately 2000 patients and their unaffected family members. Exomes have been analyzed for protein coding variants by family-based and genotype-first approaches. The diagnostic yield of our cohort remains low, and many variants of uncertain significance (VUS) occur in known inflammatory genes.
Methods/Results: We have found that validated pathogenic variants in genes associated with the immune system often score below standard thresholds for variant impact prediction. This indicates that we might be missing variants, or a diagnosis, due to a high false negative rate in the genes underlying these illnesses. Methods for variant impact prediction have relied on training models of known variants found in ClinVar. Variants associated with autoinflammatory diseases are underrepresented in ClinVar, so it is possible that existing mutation pathogenicity prediction tools are not well-suited to these genes. By statistically comparing experimentally validated variants found in ClinVar to validated variants found in the Infevers database, we've shown that the optimal accuracy thresholds for binary classification of variants associated with autoinflammation are lower than those for other genes. The optimal accuracy thresholds of predictors such as REVEL and CADD in autoinflammatory genes are 0.324 and 17.24 vs. 0.65 and 20.0, respectively. This same trend is seen for ten other predictors included in the analysis. The area under the curve is the same, indicating that we do not introduce more noise. We speculate that this difference is due to the nature of selection on these genes, with episodes of strong selection for increased inflammation due to pathogen exposure balanced by selection for reduced inflammation otherwise, which may change the relationship between conservation and functional importance. In support of this, we are exploring the relationship between measures of selection and optimal accuracy thresholds further. Preliminary studies strongly support the hypothesis that autoinflammatory genes are under strong but variable selection.
Conclusion: Our data support the establishment of new variant curation guidelines for autoinflammatory disease.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4309 Outsmarting data limitations: probabilistic data augmentation via simulation of low-coverage whole-genome sequencing

Authors:

K-Y. Chen[1], S-F. Chen[2], R. Dias[3], A. Torkamani[4]; [1]Scripps Res. California, La Jolla, CA, [2]Scripps Res., La Jolla, CA, [3]Univ. of Florida, Gainesville, FL, [4]Scripps Translational Sci. Inst., La Jolla, CA

Low-coverage whole-genome sequencing (lcWGS) has emerged as a cost-effective technique with significant applications in population genomics, enhancing the reliability of statistical analyses, refining bioinformatics workflows, and advancing model construction including genotype imputation and machine learning methodologies. Despite these advantages, the scarcity of publicly available lcWGS data, relative to high-coverage whole-genome sequencing (hcWGS), and the computational burden and time required to generate large-scale lcWGS data, have posed significant challenges. Moreover, there's a notable lack of resources specifically designed to simulate lcWGS data directly into Variant Call Format (VCF) files. This paucity is particularly evident when we seek to carry out downstream analyses and apply probabilistic data augmentation approaches for the advancement of artificial intelligence models, especially for downstream applications in genome-wide association studies (GWAS). To address these challenges, we introduce a cost-effective strategy to simulate lcWGS VCF files from existing hcWGS datasets. To simulate lcWGS data from hcWGS genotype information, we use an admixed population dataset derived from1000 Genome Project (1KGP) from different read depths to develop a multivariate Gaussian mixture model for each genotype. The multivariate Gaussian mixture models were trained on allele presence probability that derived from phred-scaled likelihood scores (PL) in lcWGS, which were aligned to their corresponding hcWGS genotype data. Our findings suggest that the probability distributions of allele presence in simulated low-coverage conditions correspond significantly with the actual probability distributions of allele presence in real low-coverage conditions across different levels of sequencing depth. Furthermore, we demonstrate that applying the lcWGS simulation tools has potential in preserving the power and predictive accuracy of polygenic risk scores. Our work offers a novel and cost-effective method for lcWGS simulation, providing a powerful tool to potentially catalyze the development of new downstream bioinformatics tools for sequencing data analysis. This approach could also benefit probabilistic data augmentation strategies for artificial intelligence.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4310 Partitioning genetic nurture: Investigating the specific parental traits driving the signal of indirect genetic effects on educational attainment

Authors:

A. Perry[1], A. Akbari[1], A. Durvasula[2], N. Patterson[3], D. Reich[1]; [1]Harvard Univ., Cambridge, MA, [2]Harvard Med. Sch., Cambridge, MA, [3]Broad Inst MIT & Harvard, Cambridge, MA

In industrial societies, education is increasingly predictive of a number of life outcomes, including an individual's health, housing prospects, and job opportunities. Thus, it is important to understand the variables, societal and biological, impacting an individual's educational opportunities. Genome-Wide Association Studies (GWAS) have identified genetic variants correlated with educational attainment, implying that an individual's predisposition to higher education is impacted in part by their genotype. Not all of the variants correlated with educational attainment, however, directly modulate the factors that lead an individual to attain more schooling. Previous research has found that some of these variants affect educational attainment by means of "indirect" genetic effects, whereby variants expressed in one individual impact the educational attainment of a different individual. One example of indirect genetic effects that is likely to be particularly relevant to educational attainment is parental effects, where a parent's value of some non-education-related phenotype affects the educational attainment of their children. Because of these parental effects, genetic variants correlated with a parent's non-education-related traits may be correlated with educational attainment in population-level GWAS. Prior research has postulated that parental effects are particularly important in explaining the genetic variants correlated with educational attainment, but a key unanswered question is which parental traits are driving this signal of indirect genetic effects. This study seeks to address this question by correlating parental polygenic risk scores (PRS; a measure of an individual's genetic predisposition toward a phenotype) for many traits to their children's educational outcomes. The study identifies parental traits which, at the phenotypic level, are correlated to children's educational attainment, and differ qualitatively from the PRS score that shapes educational attainment in the children. To control for direct genetic effects, this approach involves computing PRS on the parental genetic information not transmitted to children, and then studies the correlation between parental PRS and child educational attainment phenotypes.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4311 Pathway-Specific Polygenic Scores for Atrial Fibrillation Reveal Heterogeneous Genetic Risk Profiles

Authors:

S. Gunn[1], C. Roselli[2], K. Lunetta[1], S. Choi[1]; [1]Boston Univ. SPH, Boston, MA, [2]Broad Inst., Cambridge, MA

Background

Here we derive pathway-specific PGS (PPGS) for pathways associated with atrial fibrillation (AF) and use these PPGS to demonstrate the heterogeneity of genetic risk in individuals with AF.

Methods

Deriving Pathway-PGS

To determine pathways of interest, we used the cardiovascular knowledge portal MAGMA results for GO biological processes annotations for AF, with a Bonferroni threshold to determine a set of 76 pathways. We grouped the pathways in functional categories: conduction, contractile, cardiac development, or other. To produce each PPGS we subset the recent AFGen GWAS (in press) to SNPs within the pathway-specific set of gene bodies, and estimated weights using the PRS-CS approach. We also built a genome-wide PGS (GWPGS) with PRS-CS.

Evaluating Pathway-PGS

We computed and standardized the scores in a white British sample of the UKBB (N = 408884). First, we evaluated the prediction performance of the PPGS in the sample. Next, we considered AF cases with high overall genetic risk, defined by top 10% of GWPGS, (N = 6624) and applied k-means clustering to infer profiles of genetic risk. To summarize the clusters, we computed the overall and functional category-specific means of the standardized PPGS.

Results and Discussion

As expected, we found all PPGS were associated with AF in the UKBB (p-values < 10-20). When we consider 5 clusters we see a significant difference in overall mean PPGS and functional category-specific mean PPGS by cluster using ANOVA (p-values < 2e-16). Three clusters show relatively uniform risk across the PPGS: low, medium-high and high groups with overall mean PPGS of -0.364, 0.699, 1.15, respectively. However two clusters show varying levels of PPGS across the functional categories: one cluster has relatively low PPGS in the conduction and contractile categories (group-specific mean PPGS of -0.166 and -0.039 respectively), with higher PPGS in cardiac development (mean cardiac-development PPGS 0.551), and another cluster has the opposite pattern, relatively higher PPGS for conduction and contractile categories (0.475, 0.555) and lower PPGS in cardiac development (-0.195).

Conclusions

Using a novel method for deriving pathway-specific PGS, we show that even in individuals with high overall GWPGS for AF, there are multiple profiles of genetic risk. In future work, we plan to investigate additional sources of genetic risk for AF to identify the missing genetic risk factors for the high GWPGS but low PPGS group. We hope that these pathway-specific PGS can improve our treatment of AF by better understanding the sources of genetic risk at play in this complex trait.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4312 PENGUIN: Progressively Establish Novel Genetics Underlying Interrelation

Authors:

J. Huang[1,2], R. Sun[3], E. Hu[4], G. Deiana[5], J. M. Ranson[6], V. Napolioni[5]; [1]Southern Univ. of Sci. and Technology, Shenzhen, China, [2]Inst. for Global Hlth.and Dev., Peking Univ., Beijing, China, [3]Dept. of Publ. and Ecosystem Hlth., Cornell Univ. Coll. of Vet. Med., Ithaca, NY, [4]Southern Med. Univ. of China, Guangzhou, China, [5]Sch. of BioSci.s and Vet. Med., Univ. of Camerino, Camerino, Italy, [6]Coll. of Med. and Hlth., Univ. of Exeter, Exeter, United Kingdom

Background: The abundance of publicly available genome-wide association study (GWAS) summary statistics is providing golden opportunities for studying the interrelation of complex traits. Well-established methods include the calculation of genetic **correlation** (by LDSC etc.), the assessment of genetic **causation** (through Mendelian Randomization), and the evaluation of genetic **colocalization**. The 1st (correlation) and 2nd C (causation) mainly utilize data at the **DNA** level, while the 3rd C (colocalization) involves data at the **RNA** level. **Methods**: We now propose a heuristic 5-C approach that further includes **coevolution** and **corroboration**. The 4th C (coevolution) utilizes data at the **protein** level and enables the study of protein interaction and genetic evolution among populations of different ancestry. The 5th C (corroboration) represents a broad suite of big data mining approaches that independently support the putative interrelation. These 5-C modules are progressive, meaning that each new module is more zoomed-in and reliable that its precedent module. We nickname this 5-C approach "**PENGUIN**", for "***Progressively Establish Novel Genetics Underlying Interrelation***". **Results**: We developed a pilot version of PENGUIN and made available at github.com/jielab/penguin. The input data for PENGUIN is summary statistics of GWAS. We chose five widely studied **exposures** (adult height, body mass index [BMI], smoking behavior, low-density cholesterol [LDL], C-reactive protein [CRP]) and five **outcomes** of global public health concern (coronary artery disease [CAD], type 2 diabetes [T2D], lung cancer, dementia, COVID-19 severity) for a pilot test. We confirmed well-established interrelation (e.g. BMI and height on CAD), and also discovered novel interrelation (height on COVID-19 severity). We demonstrate that PENGUIN is a powerful and yet user-friendly platform that incorporates established software in a progressive and logic manner. **Conclusion**: Similar to the 5-C situation analysis widely adopted in business marketing (Company, Collaborators, Customers, Competitors, Climate), we designed and developed a 5-C protocol (**Correlation, Causation, Colocalization, Coevolution, Corroboration**) to scientifically generate and conveniently replicate genetic interrelation reported on daily basis.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4313 Personalized transcription factor binding from deep learning sheds light into the cis and trans regulation of gene expression.

Authors:

T. Adeluwa[1], S. Gona[1], S. Baca[2], M. Freedman[2], A. Gusev[2], B. Liu[3], R. Madduri[4], T. Amariuta[5], H. Im[6]; [1]The Univ. of Chicago, Chicago, IL, [2]Dana-Farber Cancer Inst., Boston, MA, [3]Natl. Univ. of Singapore, Singapore, Singapore, [4]Argonne Natl. Lab., Naperville, IL, [5]Univ. of California San Diego, Brookline, MA, [6]Univ. of Chicago, Chicago, IL

Transcription factors (TFs) play a pivotal role in gene regulation and phenotype expression, and their altered binding could potentially result in various human diseases (Reshef et al., 2018). Therefore, the development of reliable genetic predictors for TF activity would significantly advance our understanding of gene regulation and phenotypic implications. Previous research focused on the genetic determinants of Androgen Receptor (AR) binding by measuring AR in prostate tumor samples from 120 subjects (Baca et al., 2022). However, the cost of ChIP-Seq experiments poses significant challenges to extend this approach across a diverse range of TFs and tissue or cell types. To overcome these hurdles, we utilized cutting-edge methods such as ENFORMER, which predicts epigenetic features directly from DNA sequences, to construct a model for individualized TF binding prediction. While currently available ENFORMER models generates a limited selection of TFs and cell types, our study extends this framework by hypothesizing that TF binding can be precisely profiled as a function of genome-wide patterns of numerous epigenetic key features, as shown by the IMPACT method (Amariuta et al., 2019). We used ENFORMER to predict the 5,313 features utilized in the IMPACT framework based on DNA sequences. Subsequently, through logistic regression, we developed predictors of genome-wide TF binding per individual, based on personalized epigenetic features. A key advantage of our approach is that it relies on the same set of 5,313 ENFORMER-predicted features for the training of each new TF binding dataset. We validated our model by comparing predicted and observed AR binding in prostate cancer samples, resulting in 597 Bonferroni significant sites out of 12,812 tested. Additionally, by contrasting our predictions with those from Baca et al., who used linear predictors trained with in vivo measured TF binding across 120 individuals, we achieved correlations up to 75%, with a median of 45% (36.6% of which had BF significant p-values). Lastly, we explored the potential influence of AR binding on gene regulation by correlating the mRNA levels of a gene with the predicted AR binding at the gene's promoter. 167 out of 15,021 demonstrated significant association after Bonferroni correction. Our predictors create a promising platform for linking transcription factors in different cell types and tissues to diseases, enhancing our understanding of the genetic underpinnings of complex, polygenic diseases.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4314 Pervasive biases in GWAS using family history of Alzheimer's disease as proxy phenotypes

Authors:

Y. Wu[1], **Z. Sun**[1], J. Fletcher[2], Q. Lu[1]; [1]Univ. of Wisconsin-Madison, Madison, WI, [2]Univ. of Wisconsin, Madison, WI

Genome-wide association studies (GWAS) of Alzheimer's disease (AD) often rely on family health history as a proxy phenotype to boost sample size and statistical power. Recent AD GWASs have combined case-control associations with GWAS-by-proxy (GWAX) associations through meta-analysis. Despite its increasing popularity, there is a lack of careful investigation into the methodological issues and the quality of GWAX association results. In this study, we performed GWAX on parental AD history in the UK Biobank (47,993 proxy cases and 315,096 controls). While GWAX and GWAS exhibited a high genetic correlation (rg=0.887, p=4.1e-7), we observed substantial divergences between the two approaches, particularly in analyses involving genome-wide single nucleotide polymorphisms (SNPs) such as genetic correlation estimation, Mendelian randomization, and polygenic risk score applications. Surprisingly, the genetic correlation with education was positive in GWAX (rg=0.167, p=1.7e-11), contrary to the negative correlation observed in GWAS (rg= -0.133, p=2.4e-5). Such discrepancies were not limited to AD but were also observed in for other reported parental illnesses such as Parkinson's disease, breast cancer, severe depression, and stroke. Furthermore, we found similar issues in AD GWAX using the All of Us data (3,899 proxy cases and 23,218 controls). To identify the bias in AD GWAX, we developed a theoretical framework to decompose GWAX into AD and non-AD components. The positive genetic correlation with cognition was explained by the non-AD component (rg=0.260 with education, p=5.2e-11). Moreover, the non-AD component in GWAX exhibited correlations with lower risks of health outcomes such as coronary artery disease (rg= -0.136, p=2.8e-3) and diabetes (rg= -0.206, p=2.8e-3), indicating the presence of substantial survival bias in samples reporting parental AD history. Although adjusting for parental age in GWAX reduced the genetic correlation with education, additional unaccounted factors in the current GWAX were suggested. We also provide evidence that the genetic correlation between family health history awareness and education (rg=0.285, p=1.4e-10) contributes to these biases in GWAX associations. Our findings highlight that GWAX based on the reported parental health history introduces substantial and systematic biases in AD genetic associations due to the non-representativeness of the biobank's health history survey. Naively combining GWAX with regular case-control GWAS can lead to misleading and erroneous results in genome-wide analyses.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4315 Phenome-Wide Admixture Mapping in Hispanic/Latino(a) populations in a Large Hospital-Based Biobank

Authors:

S. Cullina[1,2], R. Shemirani[1,2], B. Rowan[1,2], N. Abul-Husn[3,2,4], S. Asgari[1,4,5], E. Kenny[1]; [1]Icahn Sch. of Med. at Mount Sinai, New York, NY, [2]Inst. for Genomic Hlth., Icahn Sch. of Med. at Mount Sinai, New York, NY, [3]23andMe, Sunnyvale, CA, [4]Dept. of Med., Icahn Sch. of Med. at Mount Sinai, New York, NY, [5]Dept. of Genetics and Genomic Sci., Icahn Sch. of Med. at Mount Sinai, New York, NY

Biobank-scale association studies in Hispanic/Latino(a) populations (HLs), who share genetic ancestry from Europe, Africa and the Americas, are lacking due to underrepresentation of HLs in genomic research and the insufficiency of genomic discovery pipelines designed for large-scale analysis of admixed populations. We developed a pipeline for admixture mapping, which tests for associations based on differences in disease prevalence and allele frequency between ancestral populations of admixed individuals, within the diverse Bio*Me* biobank in New York City.**Methods:** Global admixture proportions were inferred in ~53,000 Bio*Me* participants and 15,149 HLs were identified based on both self-reported race/ethnicity and genetically inferred admixture patterns. TOPMed imputed genotype data was phased and GNOMIX was used to infer three-way (African, European, and Native American (NA) ancestry) Local Ancestry (LA) haplotypes using a global reference panel constructed for maximal NA representation. A generalized linear mixed model (SAIGE) was used for admixture mapping, testing the association between LA haplotypes and 1,018 phecodes derived from electronic health records, as well as a curated set of over 200 biomarkers and quantitative traits. STEAM, a method that accounts for population structure, was used to derive a genome-wide significance (gws) threshold.Results: We achieve high quality local ancestry calls with correlation between global and local ancestry proportions reaching R^2 0.99. We identify 54 phecode-haplotype associations reaching gws (P&lt4.87x10^{-6}). We replicate known admixture association, such as African haplotypes overlapping the HBB gene associated with “Sickle cell anemia” (P&lt2.74x10^{-9} , Odds Ratio (OR) (95% Confidence Interval (CI)) = 4.72 (2.8-7.95)) as well as identify putatively novel associations, for example, a NA haplotype on chr7q35 associated with “Alzheimer's disease” (P&lt6.57x10^{-7}, OR= 2.33 (1.65-3.28)) and a European haplotype on chr10q26.2 associated with “Cerebrovascular disease” (P&lt9.6x10^{-7}, OR (95% CI) = 1.31 (1.17-1.45)). Current work includes fine-mapping and functional characterization of significant associations and selection scans of these regions using iSAFE.**Conclusion:** Efforts to diversify genomic research will recruit more admixed individuals to biobanks, driving the need for calibrated pipelines for genomic discovery in admixed populations. This approach enhances association testing power compared to GWAS, specifically for population-specific risk loci, and provides valuable insights into population differences in the clinical manifestation of disease risk loci.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4316 Phenome-wide association study of asthma genetic loci reveals potential subtypes of disease.

Authors:

M. J. Saward[1], R. J. Packer[1,2], L. G. Heaney[3], I. Sayers[4], K. A. Fawcett[1]; [1]Dept. of Population Hlth.Sci., Univ. of Leicester, Leicester, United Kingdom, [2]Leicester NIHR BioMed. Res. Ctr., Glenfield Hosp., Leicester, United Kingdom, [3]Wellcome-Wolfson Ctr. for Experimental Med., Sch. of Med., Dentistry and BioMed. Sci., Queen's Univ. Belfast, Belfast, United Kingdom, [4]Ctr. for Respiratory Res., NIHR Nottingham BioMed. Res. Ctr., Sch. of Med., Biodiscovery Inst., Univ. of Nottingham, Nottingham, United Kingdom

Asthma is a chronic, heterogeneous lung condition, affecting over 300 million people worldwide and susceptibility is thought to be due to a combination of genetic and environmental factors. The clinical heterogeneity is thought to be the result of different pathobiological mechanisms, and therefore phenome-wide genetic studies of identified asthma associated variants may help to reveal the genes and pathways underlying different clinical features and subtypes of asthma.

A literature search for GWAS of asthma found 43 studies identifying 478 associated variants. We tested these variants in European UK Biobank participants using Regenie, and confirmed association for 440 of the variants ($P<0.05$), which represented 210 independent signals. We applied the DeepPheWAS package in R to the 210 signals, to test for associations with 1,900 traits using data from 213,402 European participants in UK Biobank. Nine traits showed significant association with at least 10% of the signals, which were: eosinophil count (95 signals), eosinophilia (65), lymphocyte count (31), FEV1/FVC ratio (27), white blood cell count (27), neutrophil count (25), standing height (23), monocyte count (22) and platelet count (21). Eosinophilia and white blood cell count were removed for non-independence with eosinophil and neutrophil counts, respectively. Four signals were associated with six out of seven traits, two with five traits, nine with four traits, 16 with three traits and 29 with two traits. Thirteen signals were associated with both eosinophil and neutrophil counts. Some further, not often reported traits of interest were found to be associated with at least 5% of the independent signals, including IGF-1 (17), gamma-glutamyl transferase (12), nasal polyps (12), coeliac disease (12), type 1 diabetes (12) and hypothyroidism (11).

Our findings show that asthma genetic signals cluster into distinct and overlapping traits providing a potential framework to understand the contribution of these signals to disease mechanisms and identify endotypes. The numbers of signals for each trait also highlights the polygenicity of the potential subtypes.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4317 Phenome-wide association study of immune genes with known eQTL and pQTL effects using the All of Us research project

Authors:

S. Boris, B. Bates, K. Bates, M. Bailey; Brigham Young Univ., Provo, UT

Investigating genetic variants associated with disease susceptibility is crucial for understanding complex diseases. Complex diseases develop through a combination of genetic and environmental factors, and it is difficult to identify the specific genetic variants responsible for them. However, by conducting large-scale studies that look at the genomic data of many individuals, it is possible to identify genetic variants that are associated with an increased risk of developing certain diseases. Here we conducted a Phenome-Wide Association Study (PheWAS) to identify genetic variants within three genes - *ERAP2*, *HLA-DQB1*, and *PPIL3* - associated with disease susceptibility. These three genes were recently reported to have strong eQTL and pQTL SNPs found in the National Cancer Institute's Clinical Proteomic Tumor Analysis Consortium (CPTAC) in both tumor and normal adjacent samples. We used the large-scale genomic dataset from the All of Us research program, which includes the whole genome sequencing data of over 245,000 individuals. We also used the electronic health records (EHR) of these individuals to determine which of them had been diagnosed with particular diseases. Using their genomic data as well as the phenotypic information present in their EHRs, we used the PheWAS R-library written by Robert Carroll et al. 2014 to identify which disease phenotypes were linked to the identified variants within *ERAP2*, *HLA-DQB1*, and *PPIL3*. Preliminary results suggest that the previously identified eQTL and pQTL hits from the CPTAC cancer-specific dataset also display potential genetic predispositions in other diseases and are dependent on genetic ancestry. By conducting this PheWAS, we demonstrate the utility of the All of Us research program in exploring the genetic basis of diseases. Furthermore, our findings highlight the importance of investigating multiple genes simultaneously to obtain an immune-specific understanding of disease risk. In conclusion, these findings contribute to the growing knowledge base regarding the genetic underpinnings of complex diseases, paving the way for further research and personalized medicine approaches, and warrant further investigation to experimentally validate our conclusions.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4318 Phenome-wide, metabolomic and proteomic association scan of *SHROOM3* haplotypes based on imputed exonic variants.

Authors:

D. Ghasemi[1], N. Dordevic[1], L. Foco[1], E. König[1], D. Emmert[1], M. Lang[1], E. Hantikainen[1], M. Gögele[1], J. Rainer[1], F. S. Domingues[1], M. Ralser[2], C. Fuchsberger[1], D. J. M. Peters[3], P. P. Pramstaller[1], C. Pattaro[1]; [1]Eurac Res., Bolzano, Italy, [2]Inst. of Biochemistry, Charité - Univ.smedizin Berlin, Berlin, Germany, [3]Leiden Univ. Med. Ctr., Leiden, Netherlands

Introduction: SHROOM3 encodes an actin-binding protein involved in cell shaping, neural tube formation, and epithelial morphogenesis. Genome-wide association studies (GWAS) identified common variants at SHROOM3 associated with chronic kidney disease, creatinine-based estimated glomerular filtration rate (eGFRcrea), and serum magnesium (Mg). While a comprehensive haplotype analysis of this locus is missing, integration of metabolomics and proteomics data may elucidate underlying physiological mechanisms. We conducted a phenome-wide, metabolomic and proteomic analysis of SHROOM3 haplotype diversity in the Cooperative Health Research in South Tyrol (CHRIS) general population study.Methods: We performed genotype imputation of the whole cohort of 13,389 participants based on whole-exome sequencing data from 3,840 of the sample. We reconstructed haplotypes tagged by 146 polymorphic functional variants in FAM47E, STBD1, CCDC158 and SHROOM3, in the ~500kb locus bounded by the recombination hotspots identified by large GWAS. The analysis encompassed 394 traits, including 74 serum, urine and anthropometric traits, 172 targeted serum metabolites, and 148 targeted plasma protein concentrations on 3,423 individuals. We fitted linear models on the inverse normal transformation of each trait, adjusted for age, sex and the first 10 genetic principal components, embedded in an expectation-maximization algorithm for haplotype estimation.Results: We identified 11 haplotypes (H1 to H11; frequency from 24.36% to 2.03%). H8 (frequency 2.67%) was most strongly associated with eGFRcrea (P=2.7e-4), the urinary albumin-to-creatinine ratio (P=3.3e-3) and multiple phosphatidylcholines. H6 (11.61%) was chiefly associated with serum creatinine, several carnitines and coagulation traits. H4 (2.81%) was associated with Mg (P=6.7e-4), eGFRcrea and basophils. H10 (2.32%) was associated with Mg, glutamine, putrescine and afamin protein. H1 (7.65%) and H7 (2.40%) were associated with multiple phosphatidylcholines. H3 (2.32%) and H9 (3.99%) were associated with thyroid-related traits, carnitines and immunoglobulins. Cluster analysis showed that different haplotypes were associated with distinct groups of traits, more specifically eGFRcrea and Mg. Conclusion: Our multiomic, haplotype association analysis highlighted strong pleiotropy at SHROOM3, associated with kidney function but also with other sets of traits. Mechanisms regulating eGFRcrea and Mg are likely distinct. Identification of haplotypes jointly associated with complex traits, metabolites and proteins highlights molecular pathways warranting further investigations.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4319 Phenotype projections enable ultra-fast biobank-scale GWAS

Authors:

M. Zietz[1], N. Tatonetti[2]; [1]Columbia Univ., New York, NY, [2]Cedars-Sinai Med. Ctr., Los Angeles, CA

Complex diseases are responsible for significant morbidity and mortality worldwide. Genome-wide association studies (GWAS) are a useful summarization of complex disease genetics. Pan-biobank GWAS have generated shareable summary statistics on thousands of phenotypes and emerged as a powerful tool in genetics. Biobank-scale GWAS, however, have two notable limitations: they are highly resource intensive, and they do not allow for hand-crafted phenotype definitions, which are often more relevant to study. Here we present Indirect GWAS (indGWAS) as a solution to these two challenges. indGWAS is a summary-statistic-based method that computes GWAS for a phenotype defined as linear combination of other phenotypes. Our method is applicable to GWAS using both linear and linear mixed models (e.g. SAIGE). To compute pan-biobank GWAS more quickly, we decompose a large set of phenotypes down to a lower-dimensional latent space, perform GWAS on the latent space, then reconstruct GWAS summary statistics in the original phenotype space. This procedure substantially reduces computation time, and this speedup gets better with more phenotypes. We found a roughly 5x reduction in time to compute GWAS across all ICD codes in the UK Biobank, but, as a large fraction of the computations can be pre-computed, GWAS on all pairs of ICD codes would afford a 2000x speedup. For custom phenotype definitions, our method approximates a map from the phenotypic latent space to the custom phenotype, estimates GWAS for the custom phenotype, and provides estimates of the errors in summary statistics due to the approximate map. GWAS on any arbitrary linear combination phenotype can be computed almost instantaneously. Overall, indGWAS enables two major improvements. First, indGWAS can greatly increase the speed of pan-biobank GWAS studies. This makes it possible, for the first time, to systematically study complex phenotypes definitions. Second, indGWAS can allow researchers to obtain approximate GWAS summary statistics for any phenotype definition of interest, without needing access to individual-level data. This will make GWAS more widely accessible and enables more rapid iteration on phenotype definitions. In summary, indGWAS enables faster pan-biobank GWAS and the ability to estimate GWAS summary statistics for custom phenotype definitions that weren't included in pan-biobank GWAS data releases. Overall, this method will push forward our understanding of complex disease by facilitating more cost-effective, accessible, and representative genetic studies using large observational data.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4320 Pitfalls in estimating SNP-based heritability for singleton genetic variants.

Authors:

H. Wang[1], P. Wainschtein[2,1], J. Sidorenko[1], M. Fikere[1,3], Y. Zhang[1], K. Kemper[1], Z. Zheng[1,4], V. Hivert[1], J. Zeng[1], M. Goddard[5], P. Visscher[1], L. Yengo[1]; [1]Inst. for Molecular BioSci., Univ. of Queensland, Brisbane, Australia, [2]Artificial Intelligence Lab., Illumina, Inc., San Diego, CA, [3]Purdue Univ., West Lafayette, IN, [4]Broad Inst. of MIT and Harvard, Boston, MA, [5]Faculty of Vet. and Agricultural Sci., Univ. of Melbourne, Parkville, Australia

The contribution of rare genetic variants to the heritability of human complex traits remains elusive. One major obstacle to addressing this question is a lack of understanding of the sources of bias affecting heritability estimation from rare variants. In this study, we investigate different factors biasing estimates of the SNP-based heritability attributable to singleton genetic variants (i.e., observed only once in the sample). Through simulations, we illustrate that singleton-based heritability estimates can be biased by factors affecting the mean of a trait only if their variance is also correlated with singleton count (SC). Moreover, we found that scale transformations inducing a correlation between the mean and the variance of a trait can yield positive or negative biases, some of which can be corrected by fitting confounders (if measured) both as fixed and random effects in a linear mixed model. Finally, we show that phenotype distribution has a stronger effect on standard errors (S.E.) of SC-based heritability estimates than previously quantified for common SNP-based heritability. We showcased these biases using 22 real phenotypes of 305,813 unrelated European-ancestry individuals in the UK Biobank with whole-exome sequence data. When adjusting for fixed and random confounding effects, we found significant SC-based heritability estimates for 3 traits, including the number of children, which estimate (5.3%, standard error = 0.5%) amounts to the cumulative proportion of variance explained by all common SNPs. Our proposed approaches to detect and correct biases in singleton-based heritability estimates are generalizable for other classes of rare variants.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4321 Pleiotropy-robust methods for high-dimensional multivariable Mendelian randomization

Authors:

N. LaPierre, M. Stephens, X. He; Univ. of Chicago, Chicago, IL

Mendelian Randomization (MR) is a popular class of methods that use genetic variants (e.g. SNPs) to estimate the causal effect of one phenotype (the "exposure") on another phenotype (the "outcome"). The assumptions of MR are violated when SNPs that affect the exposure also affect the outcome directly, or affect a third trait (a "confounder") that is a common cause of the exposure and outcome. These "pleiotropic" effects are common and represent a central challenge in MR method development. More recently, "Multivariable" MR (MVMR) methods have extended this concept to simultaneous estimation of the causal effects of multiple exposures on a single outcome. This allows users to include potential confounders as covariates in the analysis, eliminating their confounding effect. However, in many cases, potentially confounding phenotypes are unknown or unmeasured, so pleiotropy remains an issue for MVMR methods.

We consider MVMR in the high-dimensional setting where many exposures are measured. While traditional MVMR typically involves a few exposures, our work is motivated by the emergence of large datasets that associate genetic variations with dozens to thousands of molecular phenotypes, such as metabolites.

We demonstrate that we can adjust for pleiotropy, including hidden confounding factors, by leveraging some unique properties of the data that arise in such settings. A key idea is that confounding factors will induce genetic correlations among exposures, and therefore can be inferred via factor analysis (FA) methods. Additionally, this setting naturally involves *sparsity*: not all selected SNPs will affect all exposures, and not all exposures will affect the outcome.

We develop a two-stage procedure where we estimate the hidden factors in the first stage via FA and use sparse regression to estimate the effects of exposures in the second stage. This framework enables the identification of a small portion of truly causal exposures while accounting for pleiotropic effects. Furthermore, this framework is (a) *flexible*, accommodating a number of existing methods for FA and regression; (b) *modular*, in that any approach for the first stage can be combined with any approach for the second stage; and (c) work with either summary statistics or individual-level data. We highlight the value of *variational empirical Bayes* methods, which can automatically infer the degree of sparsity from the data itself. We demonstrate that several variations of our framework greatly reduce the number of false discoveries compared to existing approaches in a simulation study, while maintaining similar or even greater power, and apply these methods to blood cell traits in the UK Biobank.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4322 Polygenic etiology underlying atopic dermatitis and asthma in individuals with inborn errors of immunity and immune dysregulation

Authors:

J. Yan[1], M. Similuk[2], M. Kamen[3], C. Jodarski[4], R. Gore[1], N. Reynolds-Lallement[1], K. Lewis[1], B. Beers[1], E. Damskey[5], H. Orellana[1], S. Byers[1], R. Ghosh[2], B. A. Seifert[6], J. Lack[7], W. Cao[1], J. Bergerson[1], A. Freeman[8], S. M. Holland[7], M. A. Walkiewicz[3]; [1]NIH, Bethesda, MD, [2]NIH, Bethesda, MD, [3]NIAID, Bethesda, MD, [4]NIH, Washington, DC, [5]NIH/NIAID, Bethesda, MD, [6]NIH, Fairfax, VA, [7]Natl. Inst. of Allergy and Infectious Diseases, Bethesda, MD, [8]Natl. Inst. of Allergy and Infectious Disease, Bethesda, MD

Background Atopic dermatitis (AD) and asthma represent inflammatory disorders characterized by multifactorial etiology. They are common in the general population; however, individuals with inborn errors of immunity (IEI) such as *STAT3* hyper IgE syndrome, DOCK8 deficiency, and CARD11 deficiency can be particularly susceptible. Prior genome wide association studies suggest shared genetic etiology between AD and asthma in population-based samples. The role of common genetic variation in the relationship between the two phenotypes in the setting of IEI and immune dysregulation remains to be explored. Here, we investigate whether common polygenic risk is associated with AD and asthma among individuals with IEI and immune dysregulation disorders.
Methods Genome sequencing was completed via a centralized sequencing program in a cohort of 1,290 unrelated probands with a spectrum of immune disorders, including IEI and immune dysregulation. Common variants with minor allele frequency>0.01 were extracted, pruned using an r^2 threshold>0.2, and used to calculate polygenic scores (PGS) for AD and asthma using variant effect sizes estimated in large previously published outside studies in FinnGen and the Global Biobank Meta-analysis Initiative respectively. Scores comprised of 1,021,400 variants for AD and 834,809 variants for asthma. Association tests were completed using logistic regression controlling for sex, age, and the first 10 genomic ancestry principal components.
Results The cohort consisted of 220 (17.1%) cases with asthma, 137 (10.6%) with AD, and 56 (4.3%) with both. Participants were 52.8% female (n=681), predominantly of European ancestry (77.9%, n=1,005), and had a mean age of 39 years (range 0-85). Of the individuals with asthma and/or AD, 134 (44.5%) had a molecular diagnosis of IEI, the most common of which included *STAT3, CYBB, PIK3CD, STAT1, TNFRSF13B, DOCK8, LRBA,* and *NOD2*. The PGS for AD was associated with asthma (OR=1.96; 95% CI[1.09, 3.52]; p=0.024) but not with AD (p=0.18). The PGS for asthma was associated with AD (OR=1.24; 95% CI[1.02, 1.51]; p=0.030) and with asthma (OR=1.77; 95% CI[1.51, 2.0]; p=4.9e-12).
Discussion This study supports correlation in polygenic risk between AD and asthma among individuals with immune disorders. Notably, AD PGS was not associated with AD itself but instead with asthma, indicating possible limited power in this cohort. This cohort was enriched for IEI diagnoses; thus, results may be specific to atopic disease related to eosinophilia and elevated IgE. Additional studies are warranted to parse causal relationships underlying these results to better understand the genetic architecture of AD and asthma.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4323 Polygenic risk and early diagnosis of diabetes.

Authors:

C. German, W. Wang, 23andMe Research Team, J. Granka, S. Aslibekyan, J. Tung, S. Shringarpure, M. Holmes; 23andMe, Sunnyvale, CA

It is estimated that of the 37.3M adults in the USA with diabetes, 23% are undiagnosed. Undiagnosed disease can lead to avoidable disease-related sequelae and morbidity that affect various parts of the body, including the heart, nerves, kidneys, eyes, and immune system. Here we explore the utility of a polygenic risk score (PRS) for identifying individuals with undiagnosed diabetes and prediabetes in the UK Biobank (UKB). We include individuals self-reporting an absence of doctor-diagnosed diabetes at study baseline in the UKB with values on HbA1c to identify individuals with undiagnosed diabetes. We used a T2D PRS comprising ~1.3 million SNPs derived from a GWAS of over 250,000 cases in the 23andMe research database. We conduct a comparison of the PRS to body mass index (BMI). In this exploratory analysis, participants were restricted to white individuals to enable use of an ancestry-appropriate PRS. Undiagnosed diabetes was defined as HbA1c ≥6.5% and prediabetes as HbA1c ≥5.7% and <6.5%. Of 414,959 individuals self-reporting an absence of doctor-diagnosed diabetes and who had an HbA1c measurement at baseline, 3061 (0.7%) had undiagnosed diabetes, representing 12.2% of all (diagnosed and undiagnosed) diabetes. Over half (1,717, 56%) of undiagnosed diabetes cases were in the top 25% of the T2D PRS. Individuals with a BMI <25kg/m^2 who were in the top 12.5% of the T2D PRS had a similar risk of undiagnosed diabetes (0.3-0.6% probability) as individuals with a BMI ≥30kg/m^2 in the lowest 12.5% of the T2D PRS (0.1%-0.6% probability). Among individuals with the highest BMI (≥35kg/m^2) in the cohort, those in the lowest 12.5% of the T2D PRS had the risk of undiagnosed diabetes equivalent to the population average with a BMI of 27.2 kg/m^2. When the number of individuals evaluated by the T2D PRS was made equivalent to those with overweight and obesity (66% of the study population), the T2D PRS was similar to BMI in identifying undiagnosed diabetes: 91.4% (2782 of 3045 with data on HbA1c and BMI) for T2D PRS vs 93.2% (2839 of 3045) for overweight/obese. Prediabetes was common (14% prevalence), with measured BMI and T2D PRS providing additive risk. 1 in 10 individuals that were in the top 12.5% T2D PRS and with BMI ≥35kg/m^2 developed incident diabetes over 4 years of follow-up, as compared to the population average of 1.6%. We find that a T2D PRS is informative in identifying undiagnosed diabetes. PRS may have utility in detecting individuals at risk of asymptomatic disease that may not be caught by routine checkups.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4324 Polygenic risk score prediction accuracy convergence

Authors:

H. Aschard[1], J. Kim[2], Z. Yang[3], S. Rubinacci[4], G. Pires[1], C. Albiñana[5], C. Boetto[1], H. Julienne[1], A. Frouin[1], A. Auvergne[1], Y. Suzuki[1], S. Djebali[6], O. Delaneau[7], A. Ganna[8], B. Vilhjálmsson[5], F. Privé[5], L. Henches[1]; [1]Inst. Pasteur, Paris, France, [2]Harvard T. H. Chan Sch. of Publ. Hlth., Boston, MA, [3]Univ. of Helsinki, Helsinki, Finland, [4]Brigham and Women's Hosp., Boston, MA, [5]Aarhus Univ., Aarhus, Denmark, [6]INSERM, Toulouse, France, [7]Univ. of Lausanne, Lausanne-Dorigny, Switzerland, [8]Inst. for Molecular Med. Finland, Helsinki, Finland

Polygenic risk scores (PRS) trained from genome-wide association study (GWAS) results are set to play a pivotal role in biomedical research addressing multifactorial human diseases. The prospect of using these risk scores in clinical care and public health is generating both enthusiasm and controversy, with varying opinions about strengths and limitations across experts. The performances of existing polygenic scores are still limited, and although it is expected to improve with increasing sample size of GWAS and the development of new powerful methods, it remains unclear how much prediction can be ultimately achieved. Here, we conducted a retrospective analysis to assess the progress in PRS prediction accuracy since the publication of the first large-scale GWASs using six common human diseases with sufficient GWAS data (type 2 diabetes, coronary artery disease, breast cancer, Alzheimer disease, asthma, and obesity). We show that while PRS accuracy has grown rapidly for years, the improvement pace from recent GWAS has decreased substantially, suggesting that further increasing GWAS sample size may translate into very modest risk discrimination improvement. We next investigated the factors influencing the maximum achievable prediction using recently released whole genome-sequencing data from 125K UK Biobank participants, and state-of-the-art modelling of polygenic outcomes. Our analyses point toward increasing the variant coverage of PRS, using either more imputed variants or sequencing data, as a key component for future improvement in prediction accuracy.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4325 Polygenic scores enable discovery of widespread genetic interactions associated with quantitative traits in the UK Biobank.

Authors:

L. Ferreira[1], S. Hu[2], S. R. Myers[1]; [1]Univ. of Oxford, Oxford, United Kingdom, [2]Novo Nordisk Ltd., West Sussex, United Kingdom

Thousands of mutations have been associated with human phenotypes, enabling greater understanding of the genetics of complex traits and disease risk prediction through polygenic scores (PGSs). Because genes, and so genetic factors, operate within biological networks, interactions between mutations are expected to exist. Their modeling could aid biological understanding and improve predictive performance by PGS. However, detecting such interactions to link phenotypes to biological networks has so far been complicated by the vast search space of variant combinations, and multiplicatively small expected effect sizes of interactions between individual variant pairs.

To increase power, we created a test for interactions between a SNP and *groups* of other variants, e.g. within a PGS. Our method is enabled by a new computational algorithm for PGS construction which iteratively adjusts for detectable additive signal for a trait. In realistic simulations, this approach avoids false positives and is well-powered to find interacting networks.

We identified 144 interactions involving 52/97 quantitative traits examined in the UK Biobank. A subset of previous GWAS hits show interaction signals, implying they also modify effects at distant loci. These include important variants modifying human disease risk, e.g. at the *FTO*, *HFE*, *HLA-C*, *LDLR* and *TCF7L2* genes. Conversely, some interaction signals are *not* identified in the original GWAS: these impact the trait by altering the predictive power of its PGS, but show no significant direct effect.

We developed a test to identify, for each interaction signal, SNPs driving that interaction. This correctly identified a known interaction between *ABO* and *FUT2* affecting alkaline phosphatase, and revealed these operate within a larger network including *FUT6*, *PIGC*, *ASGR2* and *ZNF678*. A second interaction affecting eosinophil count links a noncoding *IL33* variant and an *ALOX15* missense mutation. Interestingly, a recent functional study in mice identified *ALOX15* as reducing *IL33*-induced airway eosinophilic inflammation, implicated in asthma. Additional interactions involve distinct Alzheimer's-associated *APOE* coding mutations, for distinct traits (lipoprotein(a) and the liver-function marker alanine aminotransferase), supportive of connections among Alzheimer's disease, lipids and liver function.

Our results demonstrate the power of widely-available GWAS data to uncover key "core" genes able to modulate the impacts from some, or even all, other regions of the genome. They offer the potential for functional deconstruction of observed signals to identify subgroups of variants whose impacts reinforce one another.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4326 Population Scale Analysis of Copy Number Variation in 23andMe Research Cohort

Authors:

S. Saini[1], J. Shi[1], 23andMe Research Team, A. Auton[2], S. Shringarpure[3]; [1]23andMe, Inc., Sunnyvale, CA, [2]23andMe, South San Francisco, CA, [3]23andMe, Sunnyvale, CA

The human genome contains hundreds of thousands of regions harboring structural variants, including copy number variants (CNVs), that are known to affect around 20 million bases. CNVs contribute to numerous genomic disorders including neuropsychiatric diseases and account for a high fraction of all rare loss-of-function events. Phenotypic effects of these CNVs are largely understudied because many of these are very rare in general populations, and only larger CNVs (tens of kilobases or longer) are ascertainable from SNP-array data generated by large biobanks. Traditional methods used for CNV detection from SNP-array data tend to suffer from poor resolution and inability to accurately detect breakpoints.

We used a recently developed CNV detection method, Hi-CNV, to detect CNVs in 1 million European ancestry research participants from the 23andMe, Inc cohort. Hi-CNV leverages haplotype sharing to substantially increase CNV detection power and accuracy, and is therefore well-suited for large datasets like 23andMe.

We performed extensive QC and validation of the CNV calls using multiple orthogonal approaches. The median CNV length was 8.0 kilobases (median 5.3 kilobases for deletions, median 15.4 kilobases for duplications). Compared to an older CNV detection method PennCNV, Hi-CNV is more sensitive to detecting rarer and smaller CN events, and more accurate at harmonizing CNV breakpoints. After applying standard QC, we observed 23.6 CNVs on average per sample (>4x increase than PennCNV caller). We observed the Mendelian error rate of <1%, estimated using CNV calls from 5,000 trio samples. We also validated the calls using whole genome sequencing (WGS) data from 10,000 samples. We compared the CNV type (deletion/duplication) with the WGS read-depth direction, and observed that most of the CNVs had read-depth direction below the mean for deletions (60% of the mean read-depth on average) and above the mean for duplications (139% of the mean depth on average).

We analyzed the total genomic deletion burden and duplication burden, and observed associations of deletion and duplication burden with decreased height, with deletions appearing to be significantly more deleterious than duplications. Finally, we ran genome-wide association study analysis of several traits and successfully replicated known CNV associations with Height, BMI and Triglyceride levels. Expanding our analysis to the entire 23andMe cohort will increase our power to detect CNV associations, as well as allow us to perform trans-ethnic comparisons of CNVs distribution.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4327 Power of Inclusion: enhancing polygenic prediction with admixed individuals

Authors:

Y. Tanigawa[1,2], M. Kellis[1,2]; [1]Massachusetts Inst. of Technology, Cambridge, MA, [2]Broad Inst. of MIT and Harvard, Cambridge, MA

Admixed individuals offer unique opportunities to address limited transferability in polygenic scores (PGS), given the substantial trans-ancestry genetic correlation in many complex traits. However, they are rarely considered in PGS training: most modern PGS methods operate on association summary statistics and ancestry-matched linkage-disequilibrium (LD) reference panels and face challenges in representing LD for admixed individuals.

Here, we overcome the technical limitations and present inclusive PGS (**iPGS**). Our iPGS model captures ancestry-shared genetic effects by finding the exact solution for L_1 and L_2- penalized regression on individual-level data, thus naturally applicable to admixed individuals. To jointly model the ancestry-shared and -dependent genetic effects, we develop **iPGS+refit**, which further considers population-specific effects of genetic variants, genotype PCs, and their interactions.

We first test our approach using 336,000 synthetic individuals across 33 simulation configurations with varying heritability, polygenicity, and ancestry composition in the training set. We find that the direct inclusion of the minority population, even at 5%, in iPGS training improves transferability.

We then apply iPGS to 60 quantitative traits in UK Biobank and show the benefits of inclusive training. When trained on the same number of N=237,055 individuals, we find the greatest improvements in Africans by 31.0% on average and up to 50-fold for some traits (neutrophil count, R^2=0.058) over the baseline model trained only on European individuals. When allowing iPGS to use all individuals (N=284,661) across the continuum of genetic ancestry in training, we observe an average improvement of 39.1% for African, 9.9% for South Asian, 5.8% for non-British white, 4.1% for white British, and 12.9% for the other individuals.

Moreover, we show that joint modeling of ancestry-shared and -dependent effects with iPGS+refit further increases predictive performance when heterogeneous genetic associations are present. For neutrophil count, for example, iPGS+refit shows the highest predictive performance in the African (R^2=0.115), which exceeds the best predictive performance for white British (R^2=0.090 in the iPGS model), even though only 1.49% of individuals used in the iPGS training are of African ancestry.

Lastly, we found that our iPGS and iPGS+refit outperformed PRS-CSx, a commonly-used multi-ancestry-aware PGS method, in all 60 tested traits in UK Biobank, even though our models were trained on fewer individuals.

Overall, our results highlight the benefits and importance of inclusive PGS training for developing more equitable PGS models.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4328 Powerful detection of *trans*-eQTL partially explains missing heritability of gene expression

Authors:

K. Brunton, T. Amariuta; Univ. of California, San Diego, La Jolla, CA

While most disease-associated variants from GWAS are uncharacterized yet thought to impact disease via gene expression regulation, comprehensive maps of eQTLs (genetic variants that are associated with changes in gene expression) are critical to their mechanistic interpretation. Most previous studies exclusively detect *cis*-eQTLs, e.g. variants within +/- 1Mb of the gene. However, it has been estimated that more than 70% of genetic regulation of gene expression occurs beyond the *cis* region (Liu et al. 2019 *Cell*). Identification of *trans*-eQTLs remains challenging due to large multiple testing burdens from associating gene expression with genotypes across billions of variant-gene pairs. Despite advances in reducing this burden (transPCO: Wang et al. 2022 *bioRxiv*) and increasing sample size (eQTLGen: Vosa et al. 2021 *Nat Genet*), there is negligible gene expression heritability explained by associated *trans*-eQTLs. Here, we propose two complementary approaches to powerfully detect *trans*-eQTLs and partially recover the missing heritability of gene expression. While *trans*-eQTLs are widely believed to act through genome-wide transcription factor (TF) programs, our first approach identifies upstream regulators of gene expression marked by TF binding sites. For a target gene, we identify TF sequence motifs in the gene promoter and consider *cis* regulatory variants of the TF gene as candidate *trans*-eQTLs. Our second approach ranks variants by their best linear unbiased prediction (BLUP) effect size and finds the smallest set of candidate *trans*-eQTLs that maximizes the accuracy (cross-validation R^2) of the L1-regularized linear regression. We applied these approaches to data for 574 European individuals from GTEx across 19,587 genes expressed in whole blood. Our first approach increased the average per-gene heritability estimate by 47% compared to models of only *cis*-variants (genome-wide $h^2_g = 0.12$ vs $h^2_{cis} = 0.085$), which is significantly greater than adding a random, equally sized set of variants. Our second approach increased the average per-gene heritability (in a held-out sample) by 88% compared to models of only *cis*-variants (genome-wide $h^2_g = 0.16$). *Trans*-eQTLs of competing methods explained less heritability (transPCO: $h^2_g = 0.089$, eQTLGen: $h^2_g = 0.099$). We then fit L1-regularized gene models with our candidate *trans*-eQTLs and identified 66 *trans*-eQTLs per-gene (e.g. those with non-zero coefficients), more than transPCO (3.3) and eQTLGen (3.6). We expect that these improved models of gene expression will enhance our interpretation of GWAS variants, for example via transcriptome-wide association studies.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4329 Precise genetic instruments in Mendelian randomization improve causal estimates of complex traits: Exemplified by fresh fruit intake and diastolic blood pressure reduction

Authors:

K. James, J. B. Cole; Univ. of Colorado Anschutz Med. Campus, Denver, CO

Background: MR has strict underlying statistical assumptions that are often violated by pleiotropic genetic instruments (GI), leading to inaccurate causal associations and over-simplified interpretations. This is especially common with variants from genome-wide association study (GWAS) of complex traits with large environmental influences and substantial correlation with other health/lifestyle components, such as those relating to dietary intake. To better utilize environmental MR, we need more objective methods for selecting GI that directly proxy the exposure of interest, and not their correlated traits.
Method: We evaluated three methods for selecting GI variants for two-sample MR by testing their ability to validate a known randomized controlled trial (RCT) association between fruit intake and blood pressure. The standard method, which often violates pleiotropic assumptions, included all genome-wide significant variants associated with the exposure (GWS-GI) while our unique filtering methods included a phenome-wide association study (PheWAS) strategy to identify variants with direct effects on fruit intake (direct-GI), and a biological plausibility strategy to identify variants with mechanistic ties to fruit intake (bio-GI). GI variants for fruit were obtained from a GWAS in the UK Biobank (UKB) and from a GWAS meta-analysis of the UKB and the International Consortium of Blood Pressure for diastolic blood pressure (DBP).
Results: The GWS-GI method included 43 variants and demonstrated causality between fresh fruit and decreased DBP only when relying on 50% of the GI weight to come from valid variants (weighted median: beta=-0.69, P=0.04) and not the traditional inverse variance method (IVW: beta=-0.11, P=0.81) suggesting many variants violated the assumptions. Thus, the 43 variants were subjected to a PheWAS against >2,000 health/lifestyle traits and retained if they explained more phenotypic variance in fruit intake than any other trait (direct-GI method). This method strengthened the association for both the weighted median (beta=-0.96, P=0.01) and IVW (beta=-1.04, P=0.035) methods. Two variants mapped to olfactory receptor genes providing biological evidence for influencing fruit intake (bio-GI method). The bio-GI had limited heterogeneity (Cochran's Q=0.019, P=0.9) and similar effect estimates to the other methods (IVW: beta=-0.67, P=0.1) despite being non-significant due to less GI power.
Conclusion: Objective methods that filter GI variants using complementary information can improve causal estimates as we demonstrated by strengthening the validation of a known RCT association between fruit intake and reduced DBP.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4330 Predictive imputation of gene expression from genotype data in the Louisiana osteoporosis study

Authors:

M. Song[1], K-J. Su[2], X. Zhang[2], A. Liu[2], J. Greenbaum[2], J. Luttrell[1], L. Huang[1], W. Zhou[3], C. Wu[4], Z. Luo[2], C. Qiu[2], L. Zhao[2], Q. Tian[2], H. Shen[2], P. Gong[5], C. Zhang[1], H-W. Deng[2]; [1]Univ. of Southern Mississippi, Hattiesburg, MS, [2]Tulane Univ., New Orleans, LA, [3]Michigan Technological Univ., Houghton, MI, [4]UT MD Anderson Cancer Ctr., Houston, TX, [5]U.S. Army Engineer Res. and Dev. Ctr., Vicksburg, MS

Background: Multi-omics datasets, including genomic, transcriptomic, and methylomic data, have the potential to provide refined insights into a better understanding of the biological system and the functional mechanisms of human diseases such as Alzheimer's disease, osteoporosis, and diabetes. However, missing values within omics or across different omics for large cohort studies are a common critical issue for multi-omics data analysis, with impacts that include reducing sample size and statistical power, obstructing downstream analysis, and leading to inappropriate health policy decisions. To address this issue, predictive molecular traits, such as using genomics to predict gene expression, have been proposed. In this study, our objective is to leverage the current PrediXcan method to perform gene expression data imputation from genotype data in the Louisiana Osteoporosis Study (LOS) cohort. **Methods**: We built a customer-defined gene expression prediction model following the PredictDB pipeline, using 5-fold cross-validation on 504 subjects in LOS data. This dataset includes 315 Caucasian (CA) subjects and 189 African American (AA) subjects, all of whom have both whole genome sequencing (~22X) and transcriptome data. To mitigate the effects of distinct genetic structures among different racial groups, we performed data preprocessing for these two race groups including CA and AA populations separately. Then we trained the PrediXcan model and compared the imputation performance between LOS CA and AA populations with metrics such as Pearson correlation coefficient (PCC), p-value, and $R2$. **Results:** After training the PrediXcan model with our preprocessed LOS data, we evaluated the performance with PCC and $R2$ for LOS CA and AA populations separately. For the metrics of PCC and p-values, the number of well-predicted genes, defined as PCC>0.1 and p-value<0.05, is 2,307 (2,307/13,930 =16.56%) and 1,289 (1,289/14,065=9.16%) for LOS CA and AA populations, respectively. As for the metric of $R2$, the number of well-predicted genes, defined as $R2$ >=0.01, is 1,651 (1,651/13,930 =11.85%) and 961 (961/14,065=6.83%) for LOS CA and AA populations, respectively. For the number of overlapped genes between these two race groups, there are 557 and 389 genes with metrics of PCC and $R2$, respectively. Our results showed that the imputation performance of LOS CA subjects is better than AA subjects and the reason for it is that AA subjects has a more admixed population structure than CA subjects. **Conclusions**: We developed the data preprocessing pipeline for LOS cohort and built a customer-defined prediction model following the PredictDB pipeline from the PrediXcan method.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4331 Pre-processing and quality control of whole genome sequencing data: a case study using 9000 samples from the GENESIS-HD study.

Authors:

R. Betschart[1], D. Garcia-Aguilera[2], H. Babel[1], S. Blankenberg[3], L. Guo[3], H. Moch[2], D. Seidl[2], F. Thalén[1], A. Thiéry[1], R. Twerenbold[3], T. Zeller[3], M. Zoche[2], A. Ziegler[1]; [1]Cardio-CARE, Medizincampus Davos, Davos Wolfgang, Switzerland, [2]Inst. of Pathology and Molecular Pathology, Univ. Hosp. Zurich, Zurich, Switzerland, [3]Univ. Ctr. of Cardiovascular Sci. and Dept. of Cardiology, Univ. Heart and Vascular Ctr., Univ. Med. Ctr. Eppendorf, Hamburg, Germany

Rapid advances in high-throughput DNA sequencing technologies have enabled the conduct of large-scale whole genome sequencing (WGS) studies. Before association analysis between phenotypes and genotyped can be conducted, extensive pre-processing and quality control (QC) of the raw sequence data need to be performed. This case study describes the pre-processing pipeline and QC framework we have selected for the GENEtic SequencIng Study Hamburg-Davos (GENESIS-HD), a study involving more than 9000 human whole genomes. All samples were sequenced on a single Illumina NovaSeq 6000 with an average coverage of 35x, using a PCR-free protocol and unique dual indices (UDI). For quality control (QC), one genome in a bottle (GIAB) trio was sequenced in tetraplicate, and one GIAB sample was successfully sequenced 70 times in different runs. First, we describe the sequencing technology as well as important QC metrics on the data at various stages of the processing (raw, mapping and alignment, variant calling, multi-sample calling). Finally, we provide empirical data for the compression of raw data using the novel original read archive (ORA). Our results show that the most important quality metrics for sample filtering were ancestry, sample cross-contamination, deviations from the expected Het/Hom ratio, relatedness, and too low coverage. The compression ratio of the raw files using ORA was 5:1, and the compression time was linear with respect to genome coverage. In summary, the pre-processing, joint calling, and QC of large WGS studies is feasible in reasonable time, and efficient QC procedures are readily available.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4332 Probabilistic identification of loci of interest from GWAS summary statistics with a Bayesian mixture model

Authors:

R. Kumar, M. Ritchie; Univ. of Pennsylvania, Philadelphia, PA

Genome-wide association studies (GWAS) are a popular method for analyzing the association of genetic mutations or alterations with disease or other phenotypes. Many genetic mutations are inherited with each other, in an observed mechanism called linkage disequilibrium that results from recombination (leading to a shuffling of variants between parental chromosomes) not occurring uniformly across the genome.
This means that a variant that is linked to a causal variant may be mistakenly identified as causing disease, when in reality it is merely likely to be inherited alongside the true causal variant. Generally, variants that are close to each other are likely to be inherited together (though there are exceptions to this rule), which leads to an abundance of signals that may or may not truly be actionable or useful. These often occur in regions ("loci") called "recombination hotspots", within which multiple variants may be shown to have a significant association if any one does.
As such, people have developed downstream analysis methods that make use of GWAS results to try to identify specific causal variants or identify the functional underpinnings of their effect on disease within these loci. These include methods such as genomic colocalization analyses, Mendelian randomization, and meta-analyses, among many others. However, many of these analyses require users to define the bounds of specific regions of the genome that they would like to assess, and these bounds can have quite significant impacts on the results of these analyses. There is therefore a need for a method that identifies these boundaries in a rigorous and reproducible way.
We present a method that uses Bayesian mixture models to perform statistical inference on GWAS summary statistics to identify whether individual genomic positions represent “breakpoints” between regions containing insignificant variants and regions containing significant variants, allowing one to identify probabilistically which regions represent possible regions where variants are in linkage that may be valuable for downstream studies. We will show the results of this analysis on a synthetic set of GWAS summary statistics to show its ability to provide a probabilistic estimate of loci boundaries and then reproduce its performance on various existing GWAS summary statistics that are publicly available over a variety of hyperparameters to identify putative loci of interest.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4333 Proteogenomics often fails to identify clinically valuable targets: a Mendelian randomization study.

Authors:

J. Willett; McGill Univ., Montreal, QC, Canada

Genetically supported findings are more likely to yield a successful clinical trial. Promising targets can be identified using proteogenomics Mendelian randomization (MR), which can estimate the effect of a circulating protein's level on disease and provide a biomarker to assist in drug development. Few have investigated inconsistencies between proteogenomic MR and clinical trial results. Using MR with sensitivity tests, including pairwise colocalization (PWCoCo), we determined if the MR results of *cis*-protein quantitative trait loci (cis-pQTLs), determined using Somalogic assays in the Fenland cohort, matched pre-clinically or clinically supported targets for circulating proteins. This is because the level of circulating hormones, a subset of proteins, often determines function. Of the proteins in our initial cis-pQTL MR analysis, circulating PTHrP conferred no protection against fracture, while renin and ACE did not affect hypertension. In clinical trial data, however, abaloparatide, a PTHrP analog, decreases osteoporotic fracture risk, and renin and ACE inhibitors effectively treat hypertension. These drugs are either truncated protein isoforms (such as abaloparatide) or target specific protein domains. Further, the frequency of administration of medicines can change their effect, as has been seen for PTH, which is not reflected well in cis-pQTL studies. Somalogic assays are specific for more complete and physiological isoforms. These results demonstrate the care required when interpreting proteogenomic results, given that most assays evaluate full isoforms versus specific domains and assume constant levels of the circulating proteins—which may not reflect physiological conditions. Proteomics studies could benefit from greater resolution by targeting specific domains versus whole proteins and measuring throughout the diurnal cycle.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4334 PRS methods for equitable application across ancestrally diverse biobanks

Authors:

C. Hoggart[1], Y. Ruan[2], P. O'Reilly[1]; [1]Icahn Sch. of Med. at Mount Sinai, New York, NY, [2]Broad Inst., Cambridge, MA

For polygenic risk scores (PRS) to be universally applicable and equitable they need to have the same clinical interpretation across all individuals, regardless of ancestry and admixture. Recent methods, such as PRS-CSx and BridgePRS, have improved PRS for individuals in populations of non-European ancestry by combining information across ancestrally diverse genome-wide association studies. However, these methods focus on computing PRS in individuals from a single "homogenous" ancestral population, yet many individuals are admixed with ancestry from two or more populations from different continents. Moreover, PRS trained in one ancestral population can have a markedly different mean when applied to another ancestral population, independent of phenotypic differences between populations. For PRS to be applied clinically they need to make direct prediction of the phenotype, incorporate other predictors, and be unbiased across all strata of the population to which they are applied. Here we develop, apply and compare three methods to achieve this goal: weighting ancestry specific PRS by (1) individuals' distance in PC space to each of the discovery GWAS (2) individuals' admixture proportion, and (3) normalization of a single PRS model via PCs such that the mean and variance is similar across the diverse ancestries in the cohort. The methods are applied by estimating prediction models with predictors including both PRS and other relevant covariates in both the UK Biobank cohort and the diverse BioMe cohort recruited at Mount Sinai in New York City. Validation is performed in both unseen individuals in the same cohort and out-of-cohort. Results are reported across all ancestry strata of the target cohort, for different scenarios of available training data, considering metrics such as mean square error, bias, false positive and negative rates and variance explained.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4335 PRS training and performance evaluation in two large Han Chinese cohorts

Authors:

D. Wells[1], F-J. Hsieh[2], E-C. Yeh[2], C-H. Chen[2], S. Landon[1], R. M. Sivley[1], V. Plagnol[1], P-Y. Kwok[2,3]; [1]Genomics plc, Oxford, United Kingdom, [2]Inst. of BioMed. Sci., Academia Sinica, Taipei, Taiwan, [3]Cardiovascular Res. Inst., Inst. for Human Genetics, and Dept. of Dermatology, Univ. of California, San Francisco, CA

Polygenic risk scores (PRS) have gained interest as tools for population stratification and to improve disease prevention and screening. Yet, it is well established that PRS perform best in the population they were trained in, and their portability is a major obstacle on the path toward broad adoption. In addition to evaluation of PRS performance, sufficiently large cohorts also provide an opportunity to optimise PRS for the populations of interest. We created PRS using a range of cohorts, including several Biobanks with East-Asian ancestry participants (Biobank Japan, Tohoku Medical Biobank Project, UK Biobank, Korean Genome and Epidemiology Study). We combined these data with two additional large Han-Chinese cohorts with electronic health records or health questionnaire data (Taiwan Biobank [TWB, n=85K] and Taiwan Precision Medicine Initiative [TPMI, n=120K]) to evaluate and optimise PRS trained outside of TPMI and TWB for 12 diseases: prostate/breast/colorectal cancer, type 2 diabetes (T2D), myocardial infarction, cardiovascular disease, ischemic stroke (ISS), primary open angle glaucoma (POAG), asthma, hypertension, atrial fibrillation, and osteoporosis. We then used 140K samples from TPMI and TWB to retrain these PRS and evaluated them in a held out set, using a cross-ancestry methodology. Analyses based on additional TPMI genotyping (n=250K) are ongoing and will be reported at the meeting. Measured using odds ratio per standard deviation of the PRS, controlling for age and sex, performance varied between 1.23 (ISS) and 2.37 (T2D). A small number of these estimates differed between TPMI and TWB, most likely reflecting differences in demographics, phenotype definition and data collection methodologies. On average, the retraining of the PRS using the Han-Chinese training set generated a modest but consistent improvement, with an average benefit of 3% in the OR per SD. Compared to European ancestry performances in UK Biobank, the retrained Han-Chinese performance was better for prostate cancer (OR per SD:2.32), T2D (OR per SD:2.37) and atrial fibrillation (OR per SD:1.97). In contrast, POAG (OR per SD:1.46) and ISS (OR per SD:1.23) performed substantially worse. For the remaining seven traits, we observed a moderate attenuation of performance in Han-Chinese compared to European data (average reduction 8.3%).
Overall, and after retraining using Han-Chinese data, PRS performance for this set of common traits shows limited performance reduction compared to European ancestry data. Traits with substantially lowered performances (ISS and POAG) suggest meaningful differences in disease aetiology that may warrant dedicated case control collection.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4336 PURE: Protein-trait association Using *cis*- and *tRans*-regulation Estimation.

Authors:

Z. Zhang[1], H. Melton[1], S. Liu[2], L. Wu[2], C. Wu[3]; [1]Florida State Univ., Tallahassee, FL, [2]Univ. of Hawaii Cancer Ctr., Honolulu, HI, [3]UT MD Anderson Cancer Ctr., Houston, TX

An enhanced comprehension of genetic regulation of the proteome can accelerate the elucidation of causal mechanisms for complex traits. Proteome-wide association studies (PWAS), which leverage protein quantitative trait loci (pQTL) datasets integrated with genome-wide association studies (GWAS), are frequently used to identify probable causal proteins. Analogous to transcriptome-wide association studies (TWAS), current PWAS methods predominantly focus on *cis*-acting elements, constructing protein levels prediction models using individual-level pQTL datasets, often limited by sample size. However, *trans*-acting elements can account for a significant proportion of variation in many protein markers and often play essential regulatory roles. To maximize the potential of *trans*-acting elements and summary-level pQTL data for improving the robustness and power of PWAS, we introduce a novel PWAS method, referred to as Protein-trait association Using *cis*- and *tRans*-regulation Estimation (PURE). Our method involves two steps. First, we develop prediction models using each *cis*- and *trans*-acting element leveraging summary-level data with large sample size, thereby addressing the relatively weak effects of *trans*-acting elements. Next, we estimate the associations between phenotype and genetically predicted protein levels for each *cis*- and *trans*-acting element. These associations, estimated from each locus, are then combined using an iterative algorithm to account for certain outliers and randomness in outlier detection. Applying our novel method to deCODE summary-level data, encompassing plasma protein levels of 4,907 proteins derived from 35,559 Icelanders, we constructed 2,127 protein prediction models demonstrating satisfactory performance (R^2>0.01), achieving a 58% improvement over the existing model constructed using the ARIC dataset. Further external validation of our models using the INTERVAL data yielded a high validation rate. Finally, in a case study for Alzheimer's disease using recent GWAS data of 111,326 clinically diagnosed/proxy cases and 677,663 controls, our model identified 207 likely causal proteins under Bonferroni correction. In contrast, competing methods identified 19 likely causal proteins. We will release a companion software on GitHub to enable wider use of our method.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4337 pyGEMMA: A User-Friendly Python Implementation of Linear Mixed Models for Genome-Wide Association Studies

Authors:

R. Langefeld, B. Gao, X. Zhou; Univ. of Michigan, Ann Arbor, MI

Linear mixed models (LMMs) are widely used association analysis tools for genome-wide association studies (GWAS) to account for population stratification and relatedness among individuals. While multiple efficient LMM software packages like GEMMA, Bolt-LMM, and REGENIE exist, they often lack an accessible user interface and require specific data formatting. Here, we introduce pyGEMMA, a Python/Cython-based implementation of GEMMA that provides a user-friendly framework for conducting LMM-based analyses in genetic studies. With pyGEMMA, researchers can perform end-to-end experiments with LMMs using high-level functions and scripts. pyGEMMA provides functions for both Python and R (through reticulate) for model-fitting and visualization without requiring the specific data formats required by other standard software. It includes functions for standard GWAS visualization (e.g., Manhattan plot, Q-Q plot). We also include computational efficiency improvements by leveraging tensor-based storage in NumPy to take advantage of additional BLAS operations and a variation of Brent's root finding algorithm using hyperbolic extrapolation. The package includes functionality to run numerous SNPs in parallel in order to scale to the hardware availible. We demonstrate the effectiveness of pyGEMMA through GWAS and *cis*-eQTL analyses on two datasets. Our results show that pyGEMMA produces identical results to GEMMA with at least comparable runtime, while being more convenient and user-friendly. Additional computational improvements are currently underway. By integrating GEMMA into Python, researchers can leverage its extensive functionality for LMM analysis while enjoying the ease and flexibility of Python programming. pyGEMMA is freely available on GitHub from the authors.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4338 Quantitative and qualitative contextualization of polygenic scores in genetically diverse and admixed populations.

Authors:

S. Laskey[1], P. Wilton[1], J. Zhan[1], J. Ashenhurst[2], B. Koelsch[1]; [1]23andMe, Sunnyvale, CA, [2]23andMe Inc., Sunnyvale, CA

Motivation: One of the most prominent challenges limiting the adoption of polygenic scores (PGSes) is a failure to contextualize these raw scores into meaningful interpretations, both quantitative (i.e., the absolute chance of developing a disease) and qualitative (i.e., whether risk is "increased" with respect to some baseline). PGSes often vary systematically with genetic structure and other factors, so that an individual's raw PGS — even in units of lifetime risk — may be a poor estimate of their true disease risk, particularly for admixed individuals. Many otherwise state-of-the-art PGS algorithms separate individuals into discrete, genetically homogeneous categories, which means these algorithms are poorly suited to supporting PGS contextualization for admixed people.
Methods: To contextualize PGSes quantitatively, we recalibrated estimated probabilities against condition prevalence in a cohort at the age of peak prevalence, adjusting for sex and genetic principal components. To contextualize PGSes qualitatively, we developed a baseline model trained on admixed and unadmixed individuals to estimate a person's expected PGS given their sex and genetic principal components. We compared an individual's actual PGS to this expected value and assigned them a qualitative "increased risk" interpretation if their actual PGS exceeded their expected PGS by a clinically meaningful margin.
Results: Recalibrated PGSes were far better calibrated to true condition prevalence than raw PGSes (in deciles of 11 self-reported race/ethnicity cohorts across 12 phenotypes, error between average PRS and prevalence of 24.2% and 52.7%, respectively), particularly for PGSes that vary greatly with population structure (e.g., errors of 35.6% and 89.1% for a type 2 diabetes model). A recalibration strategy representing genetic structure as continuous was also more effective than a strategy relying on discrete categories. Without adjustment for genetic structure, qualitative interpretation of PGSes was highly confounded (e.g., 100% of people with East Asian-like ancestry in this cohort were classified as having increased risk for nearsightedness), and genetic similarity categories were inadequate to support interpretations for admixed individuals. In contrast, the baseline model produced mean PGS estimates with high accuracy for admixed and unadmixed cohorts, enabling meaningful qualitative interpretations for all individuals regardless of their admixture profile.
Conclusion: For PGSes to be practically useful in the clinic and elsewhere, they must be contextualized in a way that accounts for the impact of continuous genetic population structure.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4339 Rare variant and gene-based analysis in Early-Onset Alzheimer Disease.

Authors:

M. Jean-Francois[1], J. Bradley[2], N. Gomez[1], D. Western[3], C. Wang[4], E. Lucio Da Fonseca[5], J. Thulaseedhara Kurup[6], V. Fernandez[3], C. Cruchaga[7], C. Reitz[6], G. Beecham[1], R. Sims[8], J. Williams[8], M. Pericak-Vance[9], M. Jean-Francois[1], Alzheimer's Disease Genetics Consortium (ADGC); [1]Univ. of Miami, Miami, FL, [2]Washington Univ. in St. Louis, St. Louis, MO, [3]Washington Univ. St. Louis, St. Louis, MO, [4]Washington Univ. in St Louis, St Louis, MO, [5]Univ. of Miami, Miller Sch. of Med., Miami, FL, [6]Columbia Univ., New York, NY, [7]Washington Univ., Saint Louis, MO, [8]Cardiff Univ., Cardiff, United Kingdom, [9]Univ. of Miami Miller Sch. of Med., Miami, FL

Background: Early-onset Alzheimer (EOAD) disease is a progressive dementia characterized by memory impairment and cognitive decline by 65 years of age or younger. It is estimated that 5-6% of individuals with Alzheimer disease (AD) have EOAD. In this study, we aimed to identify genetic variants and genes associated with EOAD, a less studied form of AD. To achieve this, we conducted a comprehensive analysis using a two-pronged approach that combined single variant analysis and gene-based analysis.
Methods: Whole-genome sequence data of 3,281 unrelated individuals were analyzed to identify rare variants and genes associated with EOAD. The cohort consisted of 2,545 EOAD (AA0 ≤ 65 years) affected individuals and 736 controls (Age >65). Samples were provided by the University of Miami, Columbia University, Washington University, and the Alzheimer Disease Research Centers. Single variant association analysis was performed using logistic regression on two models: model1 (SNP and PCs) and model2 (SNP, PCS, Sex, and APOE-4 dosage). Preliminary gene based optimal sequence kernel association test was performed to investigate the cumulative effects of the rare variants within the genes. Samples were also analyzed to identify carriers of any of the known EOAD genes.
Results: The preliminary gene-based analysis of model 1 showed nominal association at KLRF2 (P=2.39×10^(-5)) and DCAF1 (P=8.13×10^(-5)). A recent study has demonstrated that depletion of natural killer cells improves cognitive functions in AD mouse models, highlighting the role that KLRF2, a gene associated with natural killer cell regulation, may have with the disease. DCAF1 has been linked to T-cell regulation senescence which can lead to uncontrolled inflammation. The single variant analysis identified five novel variants that met initial association to EOAD. Several genes within 1Mb of the significant variants have been identified across all five variants that may play a role in the progression of symptoms in EOAD. Mutations in known EOAD genes (PSEN1, PSEN2, APP) were relatively rare in our cohort, with less than 2% of cases being carriers. These findings emphasize that there could be additional factors contributing to the development of the disease.
Conclusion: Several genes have been identified that show potential association with EOAD. While many of these genes have not been specifically linked to Alzheimer's Disease (AD), they are associated with various symptoms commonly experienced by individuals affected by AD. Further research is necessary to gain a deeper understanding of these genes and their implications.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4340 Rare Variant Genetic Evidence and Drug Approval Success Rates: Insights from the UK Biobank

Authors:

G. Mcinnes, M. Luth, Y. Han, D. Jakubosky, B. Cajes, S. Bruse, O. Gottesman; Empirico Inc., San Diego, CA

Despite significant technological and methodological advances to the drug discovery and development process over the past two decades, approximately 90% of investigational therapies that enter clinical development fail, most commonly for lack of efficacy. Previous studies have shown that drug targets with some level of human genetics support are more likely to be approved. However, these studies have largely considered common variants in the target gene locus, potentially overlooking the more valuable inference that can be derived from rare nonsynonymous variants, particularly loss-of-function (LoF) variants, which tend to have larger effect sizes and may better approximate the effects of a drug intervention. Additionally, whereas it is often difficult to confidently assign a causal gene for common variant phenotypic associations, both because these variants are more likely to be non-coding and because of linkage disequilibrium with nearby variants; rare nonsynonymous variants offer a more tractable path towards establishing both a causal gene and the mechanism through which the variant impacts gene or protein function. With the emergence of large-scale biobanks comprising genotyping and/or next generation sequencing data linked to rich phenotypic information, such as the UK Biobank, rare variant analyses, either individually or in aggregate via burden testing, are a potentially powerful tool for drug discovery and development. In this study, we sought to characterize rare variant genetic evidence for the targets of approved drugs, as compared to common variants. Using exome sequencing and imputed genotype data from up to 452,401 individuals in the UK Biobank, we generated associations between rare nonsynonymous variants, either individually or in aggregate, with thousands of traits, focusing on drug-target indication pairs derived from inhibitors in the OpenTargets database. Our findings revealed a notable increase in the likelihood of drug approval for target-indication pairs backed by supportive genetic evidence from rare nonsynonymous variants, particularly LoF variants. This study underscores the value of rare variant analyses for drug discovery and development and the potential for this approach to augment success rates in clinical development. This research has been conducted using the UK Biobank Resource under Application Number 34229.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4341 Rare-variant region-based tests in whole-genome sequencing of ~490,000 UK Biobank participants

Authors:

J. Liu[1], Y. Lo[1], J. Sandhuria[1], J. Davitte[1], Y. Aulchenko[2], T. Johnson[2], R. Scott[2], A. Cortes[2]; [1]GSK, Collegeville, PA, [2]GSK, Stevenage, United Kingdom

The availability of sequencing data in biobank-scale cohorts has enabled powerful rare-variant gene-based methods for identifying gene-phenotype associations. There is seldom a most powerful mask or region-based method for identifying associations, and any such analyses will likely include a range of masks, allele frequency cutoffs and methods. Due to greater and more uniform coverage, whole-genome sequencing (WGS) can identify more than 10% of coding variants when compared to whole-exome sequencing (WES), potentially enabling identification of associations that have previously been overlooked. Here, we use WGS of ~490K UK Biobank participants to perform a comprehensive set of rare-variant gene-based association analyses across ~1500 manually derived binary and quantitative phenotypes.

We defined a series of overlapping and non-overlapping loss of function, deleterious missense, missense and synonymous annotations across a range of allele frequency cutoffs to use as masks for burden, SKAT, ACAT and omnibus association methods as implemented in Regenie. We will use permutation approaches to derive gene-based genome-wide significance cutoffs and compare our results to those derived from WES. We will perform conditional analysis on region-based signals to assess the extent to which results may be driven by LD with strong nearby common-variant associations. Finally, we will assess the consistency of gene-phenotype association signals across masks, allele frequencies and association methods, and how differences can be interpreted. Our results will inform the design and interpretation of rare-variant region-based tests using WGS data, as well as provide opportunity for discovery of novel gene-phenotype associations.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4342 Real-time polygenic prediction for streaming data.

Authors:

J. Tubbs[1,2,3], H. Huang[3,4,5], T. Ge[1,2,3]; [1]Psychiatric and Neurodevelopmental Genetics Unit, Ctr. for Genomic Med., Massachusetts Gen. Hosp., Boston, MA, [2]Dept. of Psychiatry, Massachusetts Gen. Hosp., Harvard Med. Sch., Boston, MA, [3]Stanley Ctr. for Psychiatric Res., Broad Inst. of MIT and Harvard, Cambridge, MA, [4]Analytic and Translational Genetics Unit, Massachusetts Gen. Hosp., Boston, MA, [5]Dept. of Med., Harvard Med. Sch., Boston, MA

Human genetic and phenotypic data is rapidly and continuously being generated through the efforts of large consortia, biobank repositories, and consumer genetic companies. However, GWAS are conducted only at periodic intervals, meaning that **subsequently collected samples do not contribute to polygenic prediction** until the next GWAS is conducted, often after accumulating tens of thousands of additional samples. As efforts to integrate polygenic risk scores (PRS) into routine clinical care continue to expand, a method for **dynamically integrating streaming data** into PRS construction and optimization is needed to **maximize prediction accuracy** for incoming patient samples.

Built on the PRS-CS framework, we introduce a new method, **real-time PRS-CS** (rtPRS-CS), to perform online update of SNP weights for PRS construction as each new sample is collected, maximizing available power for polygenic prediction without the need to perform intermediate GWAS. Given an initial set of SNP weights estimated from a baseline training GWAS, rtPRS-CS uses **stochastic gradient descent** to iteratively update the weights by incorporating information from each subsequent sample. We performed **extensive simulation studies** and showed that **rtPRS-CS continuously improves prediction accuracy** as the number of contributing samples increases and is robust to varying baseline GWAS sample sizes, trait heritability (h2 = 0.2, 0.5, 0.8) and polygenicity (0.1%, 1% and 10% causal variants). In addition, rtPRS-CS is highly **computationally efficient**, taking &lt 1 sec to update SNP weights from a single sample using a single processing thread.

We next applied rtPRS-CS to a range of **quantitative traits measured in the UK Biobank** (UKBB; N~300,000) and observed **steadily improving prediction accuracy** for all traits as SNP weights were updated with additional samples. For example, when applying rtPRS-CS to body mass index (BMI) with initial SNP weights trained on a GWAS from ~322,000 individuals, the variance explained in BMI increased from a baseline of 8% by an average of 0.11% for each additional 10,000 samples, and the 95% confidence interval for the variance explained in the final decile of the test sample (10.5%, 11.8%) overlapped with that of the theoretical maximum (11.6%, 12.7%), as estimated in an independent sample by meta-analyzing the baseline GWAS with GWAS from the full UKBB sample. Lastly, we validated rtPRS-CS in a dataset of **schizophrenia patients and controls** collected by the Psychiatric Genomics Consortium and Stanley Global Asia Initiatives, and demonstrated that rtPRS-CS can dynamically incorporate samples to **refine PRS calculation and maximize the power of disease risk prediction**.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4343 Recurrent CNVs and psychiatric disorders: iPSYCH2015 case-cohort study provides evidence of significant associated risk increases for ASD, ADHD, and SSD but not MDD, while highlighting locus contents as the significant predictor of CNV pathogenicity, not locus type

Authors:

s. vaez[1,2,3], M. Simone[1,2,3], X. Calle Sanchez[1,2], i. Collaboraters[2], A. Raznahan[4], W. K. Thompson[2,5], T. werge[1,2,6], D. Helenius[1,2], A. Ingason[1,2,6]; [1]Inst. of biological psychiatry, Roskilde, Denmark, [2]The Lundbeck Fndn. Initiative for Integrative Psychiatric Res. (iPSYCH), Copenhagen, Denmark, [3]The Univ. of Copenhagen, Faculty of Hlth.and Med. Sci., Copenhagen, Denmark, [4]Section on Dev.al Neurogenomics, Human Genetics Branch, Natl. Inst. of Mental Hlth.Intramural Res. Program,, Bethesda, MD, [5]Herbert Wertheim Sch. of Publ. Hlth.and Human Longevity, Univ. of California, San diego, CA, [6]Lundbeck Fndn. Ctr. for GeoGenetics, GLOBE Inst., Univ. of copenhagen, Copenhagen, Denmark

Background:

Recurrent copy number variants (CNVs) have been found to confer an increased risk of neuropsychiatric disorders in clinical case-control studies. Yet, little is known about their pathogenic impact at the level of an individual or an entire population. We sought to estimate the true population prevalence of recurrent CNVs and their associated risk of psychiatric disorders.

Methods:

We used the iPSYCH2015 case-cohort study, including all individuals born in Denmark in 1981-2008 diagnosed with a major psychiatric disorder (autism, ADHD, major depressive disorder, schizophrenia spectrum disorder, or bipolar disorder)(n= 82,626) by 2015, and a random comparison sample from the same birth cohort (n=41,346). Samples were genotyped and recurrent CNVs at 18 loci were called by PennCNV and population-representative hazard ratios (HR) were derived using a Cox proportional hazard model with inverse probability of sampling (IPS) weights. Furthermore, cross-diagnosis comparisons across the CNVs, and analyses of the effect of locus features on CNV penetrance, were performed utilizing generalized estimating equation (GEE) models.

Results:

The overall population-based prevalence of duplications was higher than that of deletions. The population prevalence of overall deletion was higher compared to the UK Biobank (0.98% vs 0.70%; $P=7.7\times 10^{-12}$), whereas overall duplication prevalence did not differ between the two studies. HR estimates varied widely across the loci, with significantly increased HR most often seen to be associated with ASD and ADHD (13 CNVs each) and less often with SSD (9 CNVs), while no CNV was associated with increased HR for MDD. Pairwise comparisons of HRs (between ASD, ADHD, and SSD) revealed high correlations between diagnoses across CNVs, with the most notable exception being higher PWAS duplication and 16p11.2 deletion associated risks of ASD than of ADHD or SSD, and higher 15q11.2 deletion and 22q11.2 duplication associated risks of ASD and ADHD, than of SSD. Increased locus size and LOEUF score associated with increased HRs across all three disorders (ADHD, ASD, and SSD) and increased iPSYCH/UKBB CNV prevalence ratio.

Discussion:

We provide population-based prevalence and risk estimates, which reveal an overall higher CNV prevalence, and no evidence of increased risk association for MDD compared to case-control studies. Therefore, our findings lay the foundation for implementing genetic predictions into clinical practice.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4344 Redefining control selection: Resolving population label issues in genetic studies reliant on genetic data with a low breadth of coverage.

Authors:

J. Russ-Silsby, T. W. Laver, M. N. Wakeling, M. B. Johnson, S. E. Flanagan, E. De Franco; Univ. of Exeter, Exeter, United Kingdom

The National Academies have issued new guidelines that stress the need for a radical shift in the use of ethnic population descriptors, urging a departure from reliance on arbitrary labels. The report highlights the issue of using self-declared ethnicity as a surrogate for ancestry, a practice deemed problematic, particularly in the context of selecting controls for genetic studies. The capacity to pinpoint suitable genetic controls is instrumental in conducting valid and reproducible genetic studies, reinforcing the importance of these new guidelines. However, conventional methods for establishing genetically similar control populations are often unsuitable for sequencing techniques with lower breadth of coverage, such as targeted next-generation sequencing (tNGS). This study's objective was to establish a robust empirical strategy for determining the optimal gnomAD control population applicable to individuals sequenced using tNGS. We performed population similarity analysis on 389 unrelated, globally diverse individuals with monogenic insulin secretion disorders who had been sequenced using both tNGS and whole genome sequencing (WGS). We employed the LASER ancestry tool on the tNGS data and Principal Component Analysis (PCA) projection on the WGS data to place individuals into a reference PCA space built using the combined 1000 Genomes and Human Genome Diversity Project (HGDP) datasets. The most genetically similar gnomAD population was then determined from the PCA scores using a bespoke neural network classifier trained on the subset of the 1000 Genomes and HGDP datasets that has been population classified by gnomAD. There was high correlation between the PCA scores generated from tNGS data by this method and those from WGS data (median r^2=0.965). When using an “argmax” classification which outputs the most probable population determined by the model, 92% of individuals were identically classified using data derived from either tNGS or WGS. When a probability threshold was imposed, whereby any individual without a determined population probability greater than 0.8 is categorized as “other”, the concordance increased to 98%. Employing this approach ensures that genetics will stay at the forefront of this ethical dilemma, substituting arbitrary labels for scientifically derived, reproducible measurements of genetic ancestry. Furthermore, while we currently rely on the population labels used by gnomAD, the ability to place individuals into a multidimensional PCA space establishes a foundation to examine genetic ancestry as a spectrum and dynamically select the most appropriate controls without fixed labels.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4346 Refining the genetic architecture of protein quantitative trait loci

Authors:

C. Benner, R. Pendergrass, M. McCarthy, A. Mahajan; Genentech, South San Francisco, CA

Understanding the genetic architecture of protein quantitative trait loci (pQTLs) is instrumental to identify novel biomarkers and drug targets for complex diseases. Given the moderate to large effect sizes at causal variants, a large part of the regional heritability of proteins is expected to be captured by fine-mapping. In cases where the fine-mapping explains a large proportion of the total regional heritability, we attain a much more informative picture of the regional architecture compared to that obtained by the existing heritability estimation methods that do not identify individual causal variants.

Here, we compared regional heritability estimation across the statistical fine-mapping model FINEMAP and both the variance component model BOLT and fixed-effect model HESS in 14,189 pQTLs from 2,940 plasma proteins in 32,867 randomized European individuals from the UK Biobank study. We observed that on an average FINEMAP identified 2.5 causal variants at an association signal and captured 36% more regional heritability than the variant with the lowest P-value. At the cis-pQTL of Peptidoglycan recognition protein 2, which has been implicated in systemic lupus erythematosus, we identified seven causal variants that together explained 20% more regional heritability than the variant with the lowest P-value, 44% more than BOLT and 14% more than HESS. In pQTLs with considerable contribution to the total heritability, FINEMAP captured on average 13% and 40% more regional heritability than BOLT and HESS. We leveraged this information about the genetic architecture of pQTLs to predict the genetic component of protein levels in the rest of the UK Biobank individuals without measured protein levels. Using genetically predicted *cis*-protein levels and 1,570 disease endpoints, we identified 189 protein-disease associations after accounting for multiple-testing correction, replicating previously reported findings of Kallikrein Related Peptidase 3 prostate cancer biomarker and recapitulating the idea of repurposing Janus kinase inhibitors for treating asthma.

Comparisons between the regional heritability estimates originating from different modeling assumptions, including a fine-mapping model, illustrated how violations of model assumptions on polygenicity or unspecified genetic architecture induce inaccuracy to the existing heritability estimates. Our analysis provides insights to understand how FINEMAP, BOLT and HESS relate to each other in cases where inference of a variant-level picture of the regional genetic architecture is indeed possible.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4347 Regional burden analysis of functional variants reveals genomic hotspots enriched for regulatory elements and overlapping with recurrent pathogenic CNVs.

Authors:

N. Safarian[1], B. Trost[1], W. Engchuan[1], M. Mendes de Aquino[1], B. Thiruvahindrapuram[1], M-C. Lai[1], R. Yuen[1], L. Gallagher[1], P. Szatmari[2], S. Scherer[1], J. Vorstman[1], E. Breetvelt[1]; [1]The Hosp. for Sick Children, Toronto, ON, Canada, [2]Ctr. for Addiction & Mental Hlth., Toronto, ON, Canada

Background:We recently developed the regional burden (RB) approach, a novel analytic method that examines the regional burden of low-frequency functional genetic variants. Using this approach, we showed that a regional excess of these genetic variants is associated with increased schizophrenia (SCZ) risk as well as decreased educational attainment. In this study, we implemented the RB analysis in autism spectrum disorder (ASD). We subsequently compared the identified RB hotspots in ASD to those found in SCZ and examined overlap with recurrent pathogenic CNVs. Lastly, we explored the biological function of genomic regions most vulnerable to excess regional burden.
Methods:We combined the two largest ASD datasets (SSC and SPARK-WES-1) comparing probands (8,842) with unaffected siblings (3,968) with available sequence data. We used 22q11.21 as template and applied our sliding-window approach across the genome, using logistic regression with ASD as outcome, regional SNV excess as predictor, while accounting for sex and 20 population stratification components. We determined RB hotspots indicated by consecutive associated windows. We subsequently compared the top 20% of ASD-related RB hotspots with those for SCZ, and with regions where recurrent pathogenic CNVs occur, using permutation to test for significance. We used enrichment analyses to explore the function of the overlapping hotspots for ASD and SCZ (10% of the genome).
Results:We found similar results for ASD as for SCZ reported in our initial study for 22q11.21 (OR=2, p=0.005). The RB-hotspots for ASD overlapped with hotspots for SCZ ($p<1e-19$) and pathogenic CNVs ($p<1e-09$). The overlapping hotspots for ASD and SCZ were highly enriched for evolutionary constrained transcriptional factors ($p<1e-99$), suggesting implication of these regions in genome regulation.
Conclusion:Our findings show that genomic regions most vulnerable for the cumulative regional impact of sequence level variants strongly overlap with regions of recurrent pathogenic CNVs, as well as RB-hotspots for SCZ. The genomic regions where the RB has the strongest impact on ASD and SCZ risk are strongly enriched for transcription factors suggestion a role in genome regulation. The question raised by these results is how evolutionary strongly constrained non-coding DNA sequences are implicated in the risk for two uniquely human conditions, ASD and SCZ. Our findings suggest that RB analysis is a novel tool that may contribute to the elucidation of genomic regions influencing vulnerability for ASD and SCZ.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4348 Regional European genetic ancestry predicts type I interferon level and risk of severe viral infection.

Authors:

I. Nln[1], J. Shum[2], Y. Ghodke-Puranik[1], R. Tipon[3], D. Vsetecka[4], S. Amin[4], A. Makol[4], T. Osborn[4], K. Moder[4], V. Chowdhary[5], U. Thanarajasingam[4], E. Svenungsson[6], G. N. Goulielmos[7], T. Niewold[1]; [1]Hosp. for Special Surgery, New York, NY, [2]MarinHlth.Med. Ctr., Kentfield, CA, [3]The New York Stem Cell Fndn., New York, NY, [4]Mayo Clinic, Rochester, MN, [5]Yale Sch. of Med., New Haven, CT, [6]Karolinska Inst., Stockholm, Sweden, [7]Univ. of Crete, Heraklion, Greece

Background. Viral infections cause significant morbidity and mortality around the world, despite progress in preventative and therapeutic development. The severity of a viral infection varies widely between individuals, ranging from mild symptoms to severe organ failure and death, and it is clear that host genetic factors play a role in this variability. Type I interferon (IFN) is a critical anti-viral cytokine, and we have previously noted differences in type I IFN levels between world populations. In this study, we investigate the interrelationship between regional European genetic ancestry, type I IFN levels, and risk of severe viral infection outcomes. **Methods & Results.** We measured type I IFN in serum samples using a WISH cell reporter assay. Regional ancestry within Europe was determined using a validated set of European-American ancestry informative markers (AIMs). Principal component analysis was performed on the AIMs data to infer each subject's proportional regional European ancestry. In cohorts of European ancestry lupus patients from Europe, we noted a Northwest vs. Southeast gradient in type I IFN levels. We then studied an independent cohort of lupus patients from the Midwestern United States with varying proportional regional European genetic admixture. Here we again observed a significant gradient in IFN level, with a greater percentage of Northwestern European ancestry corresponding to higher IFN levels. Using these data, we developed a model to predict type I IFN level based on regional European ancestry (AUC=0.73, p=6.1e-6). We next examined large databases containing serious viral infection outcomes data for European countries. Lower predicted IFN level in the corresponding European country was significantly correlated with increased fatality rate due to viral infections, including COVID-19, viral hepatitis, and HIV, with correlation coefficient ranging from -0.79 (p=4e-2), -0.94 (p=6e-3) to -0.96 (p=8e-2) respectively. This association between predicted type I IFN level and outcome severity is consistent with the expectation that greater intrinsic type I IFN response is beneficial in host defense against viral infection. **Conclusion.** We observed significant gradient in circulating type I IFN level in Northwestern vs. Southeastern European genetic ancestry. Predicted type I IFN level based upon regional European ancestry was associated with population-level fatality rates in viral infection, including COVID-19, viral hepatitis, and HIV. These findings suggest that genetic testing could provide insight into an individual's fatality risk due to viruses prior to infection, across a wide range of viral pathogens.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4349 Reimagining gene-environment interaction analysis for human complex traits

Authors:

J. Miao[1], G. Song[1], Y. Wu[1], J. Hu[1], Y. Wu[1], S. Basu[1], J. Andrews[2], K. Schaumberg[1], J. Fletcher[1], L. Schmitz[1], Q. Lu[1]; [1]Univ. of Wisconsin–Madison, Madison, WI, [2]Univ. of Washington, Seattle, WA

Most human complex traits result from multiple genetic and environmental factors and their interactions. Understanding the mechanisms by which genetic and environmental factors interact (GxE) offers valuable insights into the genetic architecture of complex traits and holds great potential for advancing precision medicine. While this concept seems intuitive, GxE has not been consistently defined in the literature, especially for complex traits due to their polygenic nature. It is often unclear how to compare different GxE approaches, or even whether they can be compared, since they may be estimating entirely different parameters. Existing approaches are also plagued by technical challenges including statistical biases, computational burden, and constraints in the data. Here, we present a unified theory and method, called PIGEON, for modeling polygenic GxE effects for complex traits. We proposed two major objectives for GxE inference, one is to quantify the overall contribution of GxE to the trait and the other is to provide mechanistic insights into the interaction mechanisms, respectively. We showed that all existing methods can be linked to these two objectives but yield flawed results. Motivated by our theory, we have developed a novel approach to estimate GxE which only requires summary statistics as input. It is unbiased, computationally efficient, robust to sample overlap, heteroscedasticity, and gene-environment correlation. We illustrated the versatility of PIGEON through applications that leverage genome-wide data and exogenous environmental exposures. We deployed PIGEON to create an atlas of polygenic gene-by-sex interactions for 530 complex traits in UK Biobank. We found that the sex difference in BMI genetics is partly explained by the genetics of anorexia - anorexia polygenic score (PGS) is significantly more associated with lower BMI in females than in males (P = 5e-14). We further applied PIGEON to a randomized clinical trial, the Lung Health Study to quantify the heterogeneous treatment effect on lung function among smokers, due to individual genetic differences. We found that bronchodilator significantly reduced FEV1 decline in individuals in the high smoking initiation PGS group (BETA = 32, P = 7e-03), whereas no such effect was observed in all samples (BETA = 14, P = 0.1) or the low PGS group (BETA = -2.8, P = 0.8). In conclusion, PIGEON is a clear demonstration that modern statistical genetics that embrace the polygenicity of human traits and rely on summary data alone can also apply to GxE research. Our analyses across different exposures and outcomes showcase the broad questions that can be addressed using this analytical framework.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4350 REMETA: Efficient meta-analysis of gene-based tests in large-scale genetic studies.

Authors:

T. Joseph, J. Mbatchou, A. Ghosh, K. Watanabe, A. Marcketta, Regeneron Genetics Center, DiscovEHR Collaboration, Mayo Clinic Project Generation, M. Ferreira, G. Abecasis, **J. Marchini**; Regeneron Genetics Ctr., Tarrytown, NY

Identifying genes and variants associated with complex traits is essential for understanding the genetic architecture of disease. Meta-analysis, which combines information across studies, has proved a powerful approach for boosting association power in genetic studies without requiring access to individual level data. Meta-analysis can be performed using summary statistics from a genome-wide scan of individual variants, or from gene-based tests that aggregate variants within a gene. The former approach is effective at identifying common variants with modest effects, while the latter boosts power for detecting rare variant associations.

Here we introduce REMETA: a software tool for meta-analysis of exome-wide association studies. In contrast with many current tools, REMETA performs meta-analysis of both single variants and gene-based tests, including burden tests, the sequence kernel association test (SKAT), and the aggregated Cauchy association test (ACAT), providing a comprehensive tool for meta-analysis. REMETA is fully integrated with REGENIE and requires no additional post-processing of REGENIE's output files.

REMETA provides several novel contributions to meta-analysis of gene-based tests. First, we show that reference LD files derived from all individuals in a cohort can be accurately substituted for the exact LD files for a given study, even if it includes just a subset of individuals. Such reference LD matrices can be pre-calculated once for a cohort and used for subsequent analyses which can substantially reduce the compute and storage requirements of gene-based tests. Second, we develop a compact per-chromosome file format for efficiently storing and sharing per study LD matrices that are required for gene-based tests. Third, we develop a method for meta-analysis of gene-based tests with high case-control imbalance that is well calibrated. Finally, we implement an omnibus strategy to aggregate results across tests and correct for multiple testing.

Using data from 740,000 individuals in UK Biobank, the Geisinger Health System DiscovEHR Collaboration, and Mayo Clinic Project Generation we demonstrate the computational efficiency and ease of use of REMETA compared to other approaches, the accuracy of using reference LD files, and the calibration when applied to binary traits with high case-control imbalance. REMETA will made available at https://rgcgithub.github.io/remeta.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4351 Reproducible causal genetic variants discovery of Alzheimer's disease via knockoffs.

Authors:

J. Gu, Z. He; Stanford Univ., Palo Alto, CA

The significance of genetic factors in the development of Alzheimer's Disease (AD) is widely acknowledged, yet the causal mechanisms remain poorly explained by existing large-scale sequencing association studies of whole-exome or whole-genome. To shed light on biological insights of different AD-associated traits, we develop a novel statistical approach based on knockoffs to discover causal genetic variants with reproducibility. By constructing knockoffs that mimic the dependency structure of variant groups, feature importance scores of both original and knockoff variants are obtained through statistical inference or machine learning. Based on a new filter that accounts for group structure, we discover causal variants by sorting importance scores of those reported positive with control on false discovery rate (FDR). Experiments of simulated data and real data demonstrate its advantage over the existing methods in both power and FDR control.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4352 Revealing associations between gut microbiome and host bone mineral density via SHAP values

Authors:

L. Jiang[1], Z. Wang[2], L. Zhang[2], B. Tian[3], X. Lin[4], X. Zhang[1], A. Liu[5], Y. Gong[5], C. Qiu[5], K-J. Su[5], Z. Luo[5], Q. Tian[5], L. Wu[5], H. Shen[5], H-W. Deng[5]; [1]Tulane Univ. Sch. of Med., New Orleans, LA, [2]Beijing Jiaotong Univ. Sch. of Sci., Beijing, China, [3]Central South Univ. Sch. of Basic Med. Sci., Changsha, China, [4]Shunde Hosp. of Southern Med. Univ., Foshan, China, [5]Tulane Univ., New Orleans, LA

Previous studies have associated alterations in gut microbiota with conditions such as inflammatory bowel disease (IBD) and obesity, and these conditions may cause bone loss or increased fracture risk. Therefore, changes in gut microbiota can potentially affect bone status. In addition, some studies have examined the associations between gut microbiome and host bone mass/microstructure and have suggested that gut microbiota can possibly regulate bone tissue mechanical properties. In this project, we aim to further elucidate this association by inspecting the relationship between gut microbiome relative abundance variations and changes in bone mineral density (BMD). We collected BMD records, demographic information, and gut microbiome relative abundance at different levels (e.g., species, genus) for over 2,000 subjects in Louisiana. Instead of using statistical modeling/testing, a random regression forest was fitted, and we calculated the Shapley Additive Explanations (SHAP) value matrix that characterizes the contribution of each microbe in predicting the BMD score of each subject. Machine learning approaches like random forest normally have less prerequisites than the conventional statistical models, and recent studies have demonstrated their strong predictive power and high capability in capturing complex data structures. SHAP values were calculated based on the Shapley values from game theory, which indicates how to fairly distribute payout among players working in coalition. The SHAP value matrix allows the revealing of local associations, which could be masked by global associations (i.e., normally the focus of conventional statistical modeling/inference), between microbiome relative abundance and BMD for subgroups of individuals (with similar demographics and/or lifestyle). In our results, we found that SHAP values can help clearly visualize the global associations regarding specific microbes. Furthermore, by clustering the subjects according to the SHAP value-derived Principal Components, we obtained subgroups with distinct BMD distributions, and observed that the microbe with highest contribution in predicting BMD differs across subgroups. Overall, our results support the idea of applying SHAP in association analysis, and the microbes highlighted in different subgroups can help us develop a better understanding of the local gut microbiome-host BMD associations in different sub-populations.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4353 Revealing clinically-relevant subtypes of complex disease via contrastive learning

Authors:

A. Gorla[1], J. Mefford[2], V. Ravi[3], Y. Di[4], J. Wang[3], T. Zhu[4], A. Alwan[3], J. P. García[5], S. Huntsman[6], D. Hu[6], C. Eng[6], J. Elhawary[6], E. Ziv[6], E. G. Burchard[6], S. Sankararaman[7], N. Zaitlen[8], J. Flint[9], E. Rahmani[10]; [1]Bioinformatics InterDept.al Program, Univ. of California, Los Angeles, Los Angeles, CA, [2]Dept. of Neurology, Univ. of California Los Angeles, Los Angeles, CA, [3]Dept. of Electrical and Computer Engineering, Univ. of California, Los Angeles, Los Angeles, CA, [4]Dept. of Psychology, Univ. of Chinese Academy of Sci., Beijing, China, [5]Univ. of La Laguna,, Santa Cruz de Tenerife, Spain, [6]Dept. of Med., Univ. of California, San Francisco, San Francisco, CA, [7]Dept. of Computer Sci., Univ. of California Los Angeles, Los Angeles, CA, [8]Ctr. for Neurobehavioral Genetics, Semel Inst. for NeuroSci. and Human Behavior, Univ. of California, Los Angeles, Los Angeles, CA, [9]Dept. of Psychiatry and Biobehavioral Sci., Brain Res. Inst., Univ. of California Los Angeles, Los Angeles, CA, [10]Dept. of Computational Med., David Geffen Sch. of Med., Univ. of California Los Angeles, Los Angeles, CA

BACKGROUND:
A critical step towards realizing the promise of precision medicine is the identification of clinically-relevant disease subtypes. Existing phenotype subtyping methods primarily rely on data clustering or dimensionality reduction. However, they are tuned to capture dominant sources of variation in the data, which often originate from variation that is not descriptive of the mechanistic heterogeneity of the phenotype of interest.
OBJECTIVE:
We aimed to develop a method to identify de novo clinically-relevant subtypes from patient data that can robustly isolate subphenotypic variation even when the signal strength tends to be weak.
METHODS:
We introduce Phenotype Aware Component Analysis (PACA), a novel model-based contrastive learning algorithm. Given case/control status of any data modality that may harbor meaningful sub-phenotypic signals, PACA learns gradients of variation unique to cases (hence subphenotypic signal), while accounting for variation and imbalances of biological/technical confounders between cases and controls. We applied PACA to three clinical case/control datasets to learn new subtypes.
RESULTS:
When compared to several dimensionality reduction and contrastive learning algorithms in simulations, we show that PACA is better powered and the only well calibrated method, reporting no heterogeneity under the null. Using the PsychENCODE clinical data (N=1,321), we grouped transcriptomes from participants with either bipolar disorder (BD) or schizophrenia (SZ) diagnosis. PACA, blind to the BD/SZ labels, isolated a gradient that differentiates BD from SZ (p=2.7e-14; Pearson correlation) and reflects heterogeneity in the genetic architecture (p=1.4e-6; Subtest). Next, applying PACA to voice recordings of female Han Chinese major depressive disorder (MDD) patients and controls from the CONVERGE study (N=1,206) allowed us to leverage the voice data to define two MDD subtypes driven by previous treatment type and educational attainment (p=4.5e-3, p=3.0e-4; linear regression). These subtypes replicated (p=8.7e-3; N=6,142) in an independent sample of voice recordings collected 10 years earlier. Finally, applying PACA to DNA methylation data from Mexican and Puerto Rican pediatric asthma cases and controls (N=1,017) revealed subphenotypic variation associated with disease control level and the intensity of the treatment received (p=4.6e-3, p=3.3e-3; linear regression).
CONCLUSIONS:
Our results render PACA as a calibrated and well-powered new state-of-the-art tool for learning de novo subtypes that are more likely to reflect clinically significant heterogeneity.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4354 Revealing disease genetics: unveiling genetic architecture through differential copy number variation in admixed populations

Authors:

B. Carvalho[1,2], B. Henning[3], T. de Araujo[1,2], H. Tadiello[1,2], V. Almeida[1,2], A. Coan[1,2], M. Montenegro[1,2], M. Alvim[1,2], C. Yasuda[1,2], F. Cendes[1,2], I. Lopes-Cendes[1,2]; [1]Univ.e Estadual de Campinas (UNICAMP), Campinas, Brazil, [2]Brazilian Inst. of NeuroSci. and Neurotechnology (BRAINN), Campinas, Brazil, [3]Univ. of Washington, Seattle, WA

Changes in the DNA may lead to changes at the protein level, which may ultimately be transferred to organisms and diagnosed as a disease. Understanding the genetic landscape of epilepsy will significantly improve patient lives and healthcare delivery. We have seen that copy number variants (CNVs) have contributed to elucidating the etiology in several epilepsy patients. CNVs are genomic intervals (ranging from 1 kilobase to an entire chromosome) that deviate from the normal diploid state of a reference genome. Recent technological and analytical advances revealed the abundance of CNVs in human genomes, showing thatCNVs account for a large amount of human genetic variation and contribute to interindividual variation in complex phenotypes such as disease susceptibility. We propose a statistical learning-based method for CNV detection from whole exome sequencing (WES) data, with extensions to whole genome sequencing(WGS) data. Our focus is on Brazilians, an admixed population often underrepresented in genomic studies, and this approach allows us to investigate population-specific associations between CNVs and phenotypes. In the algorithm's training phase, we use CNV calls from SNP arrays (109 epilepsy cases and 237 controls) as gold-standard to improve the method's sensitivity. When predicting CNV calls on new samples, the method uses only allele dosage and overall abundance derived from site-specific depths. Investigating the inheritance patterns and the genes affected by the CNVs we found in high frequency in case samples should shed light on their influence on the disease phenotype. Complex traits are associated with large and de novo CNVs, while small CNVs that disrupt haploinsufficient genes also implicate the development of diseases. The CNV calls we detected will be combined with statistical learning approaches to provide the community with a well-calibrated CNV calling strategy for NGS data. We expect this approach to play an essential role in untangling the complexity of traits in samples of the Brazilian population, withan easy way of generalizing it to other underrepresented populations.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4355 Revealing variable genetic association patterns in COVID-19 patient subgroups: A machine learning approach in genome wide association studies

Authors:

N. Avramidis[1], S. Sohan[2], N. Pirastu[3], E-P. Castineira[1], K. Rawlik[1], K. Baillie[1,4,5]; [1]Pandemic Sci. Hub, Ctr. for Inflammation Res. and Roslin Inst., The Univ. of Edinburgh, Edinburgh, United Kingdom, [2]4Inst. for Adaptive and Neural Computation, Sch. of Informatics, The Univ. of Edinburgh, Edinburgh, United Kingdom, [3]Human Technopole, Milan, Italy, [4]MRC Human Genetics Unit, Inst. of Genetics, Edinburgh, United Kingdom, [5]Intensive Care Unit, Royal Infirmary of Edinburgh, Edinburgh, United Kingdom

There is evidence that critically ill COVID-19 patients exhibit a range of clinical phenotypes which can affect the patient's response to treatment. Understanding the underlying biological reasons for this variation can facilitate the development of advanced therapeutic interventions. Using combined linked electronic health care record data and genotypes for individuals who were recruited to both the Genetics of Mortality in Critical Care (GenOMICC) and the ISARIC Coronavirus Clinical Characterisation Consortium, (ISARIC4C) studies, we aimed to identify and genetically characterize subgroups of critical illness in COVID-19. We stratified 6,996 critically ill COVID-19 patients into distinct subgroups, by applying latent class analysis on the presence of 25 symptoms at hospitalization. Furthermore, we conducted a Genome-Wide Association Study (GWAS) in a one-vs-rest design, where each cluster in turn is treated as a case group with the rest of the clusters acting as control group. The resulting clusters are stable for a number of subgroups from 2 to 9. As the model progresses to higher level solutions, new clusters are formed by partitioning of one of the existing clusters, with minimal sample movement amongst already existing clusters. Most subgroups exhibit at least two of three core symptoms: fever, cough, and shortness of breath. The subgroup's distinction is based on the presence of less common symptoms, such as myalgia, loss of taste and smell, and gastrointestinal symptoms. A model with more than 6 clusters results in an underpowered study design for a one-vs-rest GWAS. Accounting also for cluster stability and phenotypic differentiation, an efficient study design for one-vs-rest GWAS lies between 4 and 6 clusters.The results from the cluster specific GWAS are indicative for genetic heterogeneity amongst the identified subgroups. Examining 49 previously identified genetic susceptibility loci, we found that cluster specific effect sizes exhibit variability in the association patterns. We are currently exploring these differences with the aim of identifying cluster-specific genetic informed drug targets.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4356 Revisiting statistical modeling of sQTLs leads to significant improvement in identification and characterization of RNA splicing associated variants.

Authors:

D. Wang, M. Gazzara, S. Jewell, C. Brown, Y. Barash; Univ. of Pennsylvania, Philadelphia, PA

Identification and characterization of splicing quantitative trait loci (sQTLs) has emerged as a critical component in understanding the function of noncoding genetic variants implicated in disease. However, a significant number of sQTLs remain undiscovered due to limitations in both splicing quantification and statistical methods. Here we present a sQTL mapping framework that identifies thousands of novel variants that have been recurrently omitted in recent studies. Our method combines event and transcript level quantifications to identify variants associated with a more comprehensive set of splicing phenotypes that includes intron retention and alternative transcript start/end which are not considered by existing approaches. Using GTEX as a case study, we show that existing pipelines using Leafcutter (Li et al. 2017) fail to report over 25% of complex slicing variation across multiple tissues compared to our hybrid approach using MAJIQ (Vaquero et al. 2016, 2023).

We also develop robust statistical methods to handle discrete and highly correlated multivariate splicing phenotypes which have more power to detect sQTLs while reducing false discoveries. In brief, our method leverages an optimized beta-binomial regression implementation to account for uncertainty in splicing quantifications computed from sparse junction spanning reads. Through further modeling of overdispersed count data, phenotype correlation, missing values, and heteroscedasticity, our model outperforms current methods which were adapted 'as is' from eQTL studies but are still the standard in the field.

To handle multiple hypothesis testing across highly correlated splicing phenotypes, we introduce a permutation free approach based on a multivariate null resampling model. This is followed by optimization of the cis-window size around splice sites which maximizes sQTL discovery relative to the testing burden incurred by a larger search space. The method is many orders of magnitude faster than permutations while still controlling FWER.

Finally, to facilitate downstream variant interpretation, we also introduce improved visualization tools which are implemented as part of the Voila package in MAJIQ. Some key features of this tool include visualization of coordinate splicing changes and a combined transcript-splice junction view of sQTLs. Our pipeline discovers novel intron retention associated variants in the Alzheimer's CASS4 gene and variants in NAGNAG motifs which provide novel insight into the functional role of genetic variants in splicing regulation.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4357 Robust inference with summary statistics knockoffs in genome-wide association studies

Authors:

X. Qi[1], M. E. Belloy[1], X. Liu[1], H. Tang[2], Z. He[1]; [1]Stanford Univ., Palo Alto, CA, [2]Stanford Univ. Sch. of Med., Stanford, CA

Genome-wide association studies (GWASs) are widely used to identify genetic variants that impact diseases by meta-analyzing genotyped deoxyribonucleic acid (DNA) sequences across populations and cohorts. Motivated by the limitations of GWASs to capture weaker associations of susceptibility variants and to construct polygenic architecture for complex diseases, we extensively evaluate the robustness of a summary statistics knockoff inference method, GhostKnockoff, and develop a pipeline for the genome-wide discovery of putative causal variants and downstream function interpretations. With empirical simulation studies, we show that GhostKnockoff based on mixed effect model score test control false discovery rate (FDR) and achieve high power for data with related samples. GhostKnockoff also demonstrates robustness to the source of input Z-scores, e.g. Z-scores from different types of marginal association tests or from a meta-analysis for data with independent samples. We apply GhostKnockoff to the meta summary statistics aggregated over nine European ancestral genome-wide association studies and whole exome/genome sequencing studies to select potentially causal genetic variants of the Alzheimer's disease (AD) and to identify functionally informed genes. The downstream analysis of single-cell transcriptomics data is performed to validate genes' biological signals between AD cases and controls.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4358 Role of Epigenetics in the Oral Health Status of Monozygotic and Dizygotic Twins - A Comparative Study

Authors:

D. Francis; Saveetha Dental Coll. and Hosp., SIMATS, Chennai, India

Background: For years, twin research has been at the center of the nature vs. nurture argument. They have been a great source of knowledge on the genetic basis of complex traits, as well as a means of investigating genetic and environmental influences on behavioral and medical characteristics. Twin studies are an intriguing research tool because of its capacity to accurately isolate a feature and measure its influence. The aim was to assess the oral health status and concordance between monozygotic and dizygotic twin pairs. Materials and Methods: After obtaining prior consent, a cross-sectional descriptive study was conducted among 9 monozygotic and 21 dizygotic twin pairs who were reared together. Perception towards oral health practices was assessed using a pre-tested questionnaire. The WHO oral health assessment form (1997) was employed to assess the oral health status. Zygosity determination was determined using the medical records, dermatoglyphics and details about chorionicity and number of placental cords. Pearson's correlation was calculated to determine the correlation among the monozygotic and dizygotic twin pairs. Results: The monozygotic twin pairs showed a greater correlation compared to the dizygotic twin pairs in dental caries, periodontal disease and malocclusion. Conclusion: In the present study, monozygotic twin pairs showed a higher correlation rate than the dizygotic twin pairs, suggesting considerable evidence that genes play a significant role in the etiology of dental caries, periodontal disease and malocclusion.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4359 Safety evaluation of elevated fetal hemoglobin level: Evidence from human genetics

Authors:

T. Zou, L. Barrera, M. Parker; Beam Therapeutics, Cambridge, MA

Background: Elevation of fetal hemoglobin (HbF) is a promising therapeutic strategy to treat thalassemia and sickle cell disease. However, the side effects of elevated HbF are unclear. We aimed to evaluate the safety of elevated HbF level through large-scale association analysis of HbF-increasing genetic variants in UKBB.
Methods: From previous literature, we identified 4 independent common variants (rs7482144 in *HBG2*, rs1427407 and rs11886868 in *BCL11A*, and rs9399137 in *HBS1L/MYB*) and three rare variants (11:5250055:A:G, 11:5249971:G:A, and 11:5254983:G:C) significantly associated with increased HbF levels. For common variants, we calculated the total score of HbF-increasing variants and conducted phenome-wide association studies (PheWAS) of this score with 297 continuous traits, 916 ICD10 diagnoses, and 3 pregnancy outcomes (offspring birth weight, small for gestational age (y/n), and live birth status (y/n)). Analyses were performed using PLINK in unrelated European ancestry individuals in UKBB (n=337,420) adjusting for age, sex, and 20 genetic principal components. For rare variant carriers identified from the 150k whole genome sequencing data, we characterized their demographic factors, pregnancy outcomes, and disease diagnoses.
Results: A total of 255,773 individuals in UKBB carried at least 1 common HbF-increasing variant and 31,100 carried ≥ 6 HbF-increasing common variants. On average, each common variant increased HbF by 0.5%, thus the total score could elevate HbF ~0-4% assumably. PheWAS revealed significant associations between HbF-increasing alleles and measured blood cell traits in UKBB (most significant association with mean corpuscular hemoglobin, Beta=0.04, $P= 1.7x10^{-184}$). There was no association between common HbF-increasing alleles and any ICD10 diagnoses, pregnancy outcomes, or death in UKBB ($P> 5.4x10^{-5}$). We identified 69 heterozygous carriers of known HbF-increasing rare variants in the *HBG1/HBG2* promoter, which are expected to increase HbF to 2-24%. We did not observe abnormal blood cell traits, disease patterns, or pregnancy outcomes in rare variant carriers. Together, these analyses in UKBB suggest lifelong moderate elevation of HbF levels in the general population is not associated with adverse outcomes. Additional analysis is needed to understand the effect of HbF elevation beyond 24% and the effect in thalassemia and sickle cell patients.
Conclusions: We leveraged large-scale human genetics to evaluate the safety of elevated HbF. HbF-increasing alleles were significantly associated with measured blood cell traits, but not any ICD10 disease diagnoses, pregnancy outcomes, or death in UKBB.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4360 SaseR: a novel tool forfast and Scalable Aberrant Splicing and Expression Retrieval in R

Authors:

A. Segers[1], J. Gilis[1], M. Van Heetvelde[1], E. De Baere[2], L. Clement[1]; [1]UGent, Ghent, Belgium, [2]Ghent Univ. and Ghent Univ. Hosp., Ghent, Belgium

Millions of patients suffer from rare Mendelian diseases, for whom a diagnostic rate of their pathogenic variants of 15-75% is currently achieved with whole exome sequencing (WES) and whole genome sequencing (WGS). There is growing evidence that the diagnostic rate can be further improved by discovering mutations in intronic and in other non-coding regions that contribute to disease by disrupting transcriptional regulation. Therefore, WGS is increasingly complemented with RNA-seq profiling to boost the diagnostic rate by identifying aberrant expression (AE) and aberrant splicing (AS).

The AE and AS discovery, however, is not possible with default bulk RNA-seq workflows. With this respect, Outrider and Fraser have disrupted the field by providing formal count-based outlier tests to pick up AE and AS, respectively, while automatically controlling for latent confounders. But, their approach is slow and Fraser is discarding a lot of information as it only uses junction reads. We argue that juggling with input counts and offsets can effectively unlock conventional bulk RNA-seq workflows for fast and scalable AE and AS analyses.

We introduce a single workflow for differential expression, differential splicing (DS), AE and AS based on conventual bulk RNA-seq workflows. Moreover, our workflow is very flexible and provides a simple and unified workflow that can be extended to many applications, e.g. modelling allele specific expression. Our framework also remains future proof to novel sequencing-based technologies and applications, such as transcript counts with long read sequencing, as long as the quantification can be recasted in specific feature counts in conjunction with proper offsets.

In this study, we show how different input counts and offset combinations can be used to assess many research questions in our framework, with a particular focus on aberrant expression and splicing. We further discuss the performance of our workflow to detect AE and AS analyses in simulation and real case studies. We show how our workflows vastly outperform existing state-of-the-art tools in terms of computational speed and scalability, while also dramatically boost the performance to detect ASwhile maintaining a similar performance for AE detection.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4361 Scalable knockoff-based framework for summary statistics analysis

Authors:

b. chu; Stanford Univ., MENLO PARK, CA

We propose GGhostKnockoff for analyzing GWAS (genome-wide association study) summary statistics data. GGhostKnockoff empirically clusters tightly linked variants into groups, runs a joint Lasso regression on marginal Z-scores, and returns a set of conditionally independent discoveries with guaranteed False Discovery Rate control. The methodology is based on recent advances in statistical knockoff theory. Using GGhostKnockoff, we re-analyzed 400+ GWAS results based on publicly available data. We show that GGhostKnockoff makes 10-30% additional (independent) discoveries compared to conventional GWAS, that the discoveries explain additional heritability of the traits, that they are consistent with known genetic correlation across phenotypes, and that the discoveries overlap with real functional experiments. Each phenotype takes ~1 hour utilizing a personal laptop, making GGhostKnockoff a practical approach to analyze the largest summary statistics datasets today.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4362 Scalable summary statistics-based heritability estimation method with individual genotype level accuracy

Authors:

M. Jeong, A. Pazokitoroudi, S. Sankararaman; UCLA, Los Angeles, CA

Heritability estimation of complex traits on large-scale biobanks has allowed us to better understand the genetic architecture underlying various diseases and traits. While numerous methods have been developed to accommodate the increasing size of data and address privacy concerns, most approaches have inherent limitations. For instance, the widely used LD-score regression (LDSC) allows for easy sharing and fast estimation based on summary statistics but lacks the accuracy of methods utilizing individual genotypes. On the other hand, methods that directly use individual genotypes either fail to scale to modern biobanks with nearly millions of samples or are limited in their applicability due to privacy constraints. Often, analyzing individual genotype data such as the Whole Genome Sequencing (WGS) data in the UK Biobank (UKBB) is restricted to its cloud server, highlighting the need for a scalable summary statistics-based approach with high accuracy to reduce its cost. Here we present SUMmary statistics-based Randomized Haseman-Elston regression (SUM-RHE), an extension of a previously proposed method-of-moments heritability estimator to the setting where only summary statistics are available. Our method estimates heritabilities of complex phenotypes with accuracies comparable to approaches that require individual genotypes, while exclusively relying on summary statistics. SUM-RHE employs the same summary statistics (GWAS and LD scores) as LDSC, and an additional population parameter obtained on a reference population, which can be efficiently estimated and readily shared for public use. To validate our approach, we conduct a comprehensive benchmark using simulated phenotypes from the UKBB genotype data with diverse genetic architectures. Our results demonstrate that SUM-RHE achieved a substantial improvement in the Mean Squared Error (MSE) by nearly four-fold compared to LDSC across various genetic architectures while being calibrated (effectively increasing the sample size by a factor of four). We apply our method to real phenotypes for additional validation.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4363 SCiMS: Sex calling for metagenomic sequences.

Authors:

E. Davenport, K. Kirven, H. Tran; Pennsylvania State Univ., University Park, PA

There is growing interest in the bacteria, archaea, and other microbial species that live in our bodies, otherwise known as the microbiome. To characterize it, metagenomics targets all of the DNA in a sample for sequencing in order to gain a complete picture of what microbial taxa and functions exist in a sample. Host characteristics like age and sex are often associated with microbiome composition and are therefore important metadata variables to consider during analysis. However, these metadata are not always collected during sampling. Additionally, sample swaps can occur during processing prior to sequencing and are often challenging to detect. Metagenomic processing pipelines currently lack the ability to confidently call host sex from sequence data alone, either when that metadata is not available or as a QC check. To address this gap, here we present an easy to implement bioinformatic tool that calls sex from shotgun metagenomic data called Sex Calling in Metagenomic Sequences, or SCiMS. In short, SCiMS takes in an alignment file of reads mapped to the human genome, which is typically generated as one of the first steps during metagenomic data analysis. It uses a statistic based on the ratio of reads mapping to the autosomes and to the X (Rx) and Y (Ry) chromosomes to assign biological sex. First, to assess accuracy and depth requirements of the method, we simulated metagenomic datasets with increasing amounts of host reads, from 100 host reads/sample to 1 million host reads per sample using wgsim (v.0.3.1-r13). SCiMS accurately calls sex with as low as 1000 simulated host reads total, demonstrating its potential utility on metagenomic samples with low host:high microbial read content. Second, we applied SCiMS to 639 metagenomic samples from 15 tissues from the Human Microbiome Project. These samples varied in host read content anywhere from 194 host reads to 3,311,380 host reads per sample. SCiMS correctly called sex from 90% of the samples, even from body sites with low host read sequence content such as stool and left retroauricular crease. Finally, we applied SCiMS to 107 mouse fecal and cecal samples and also demonstrated high accuracy in sex calls (95% called correctly compared to metadata). In conclusion, we present SCiMS as the first tool for calling sex from metagenomic reads. SCiMS can be applied to data collected from any species with a heterogametic sex system and is freely available online.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4364 scTIE: data integration and inference of gene regulation using single-cell temporal multimodal data

Authors:

Y. Lin[1], T-Y. Wu[2], X. Chen[2], S. Wan[3], B. Chao[2], J. Xin[2], J. Yang[4], W. H. Wong[2], R. Wang[4]; [1]Yale Univ., New Haven, CT, [2]Stanford Univ., Stanford, CA, [3]Natl. Yang Ming Chiao Tung Univ., Hsinchu, Taiwan, [4]The Univ. of Sydney, Sydney, Australia

Single-cell technologies offer unprecedented opportunities to dissect gene regulatory mechanisms in context-specific ways. Although there are computational methods for extracting gene regulatory relationships from scRNA-seq and scATAC-seq data, the data integration problem, essential for accurate cell type identification, has been mostly treated as a standalone challenge. Here we present scTIE, a unified method that integrates temporal multimodal data and infers regulatory relationships predictive of cellular state changes. scTIE uses an autoencoder to embed cells from all time points into a common space using iterative optimal transport, followed by extracting interpretable information to predict cell trajectories. Using a variety of synthetic and real temporal multimodal datasets, we demonstrate scTIE achieves effective data integration while preserving more biological signals than existing methods, particularly in the presence of batch effects and noise. Furthermore, on the exemplar multiome dataset we generated from differentiating mouse embryonic stem cells over time, we demonstrate scTIE captures regulatory elements highly predictive of cell transition probabilities, providing new potentials to understand the regulatory landscape driving developmental processes.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4365 Selecting Covariates for Genome-Wide Association Studies

Authors:

N. Rappoport[1], E. Dor[2], I. Margaliot[3], N. Brandes[4], O. Zuk[3], M. Linial[5]; [1]Ben-Gurion Univ. of the Negev, Beer-sheva, Israel, [2]Ben-Gurion Univ. of the Negev, Beer-Sheva, Israel, [3]The Hebrew Univ. of Jerusalem, Jerusalem, Israel, [4]UCSF, San Francisco, CA, [5]The Hebrew Univ., Jerusalem, Israel

A central goal in performing Genome-Wide Association Study (GWAS) is to identify statistically significant associations between genetic variations and phenotype, thus pointing to the possible biological mechanisms underlying the studied phenotype. However, GWAS is often prone to multiple uncontrolled confounders and biases (e.g., selection bias and population structure). The standard practice in GWAS is to test each SNP independently for association with the trait, which may lead to a high rate of false positives when confounders that are correlated with both the trait and the variant are not included in the model as covariates. The routine GWAS protocol suggests including covariates whose purpose is to control for the indirect effects unrelated to the phenotype of interest and eliminate the influence of confounders. These covariates include technical components (e.g., the genomic center and SNP-chip technology) but also covariates of biological and medical importance, such as the sex and age of the individual that may directly affect the phenotype and Principal Components used to adjust for population structure.The choice of which covariates to include in a GWAS is important since it affects the ability to detect true association signals of variants, to correct for confounders and avoid false positives, and the running time of the analysis. Despite the importance of this issue, there is no consensus or clear guidelines for the right choice of covariates. Therefore, studies typically employ heuristics for their selection with no clear justification. Here, we explore the dependence of the GWAS analysis results on the choice of covariates for a wide range of quantitative and binary human phenotypes. We propose guidelines for the choice of covariates based on the phenotype's type (quantitative vs. disease), heritability, and disease prevalence, aiming to maximize the statistical power to detect true associations and fit accurate polygenic scores while avoiding spurious associations and minimizing computation time.We analyze 36 traits in the UK-Biobank dataset. We show that the genotype batch and assessment center can be safely removed as covariates, thus significantly reducing the GWAS computational burden for these traits.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4366 Sensitivity of MR analyses to pQTL instrumental variable selection contributes to lack of replication in heart failure

Authors:

K. Catalano[1], C. Li[1], A. Hukku[1], E. Kvikstad[2], S. Vasquez-Grinnell[2], J. Szustakowski[2], E. Holzinger[1], J. Maranville[1]; [1]Bristol Myers Squibb, Cambridge, MA, [2]Bristol Myers Squibb, Lawrenceville, NJ

Heart failure (HF) has become increasingly prevalent, and there is a need to identify druggable targets to expand treatment options. Mendelian Randomization (MR) analysis of HF using circulating protein data can identify proteins found to enhance or reduce disease risk that may be targeted with inhibition or agonist therapies. MR is a power tool for causal inference, but the method relies on assumptions that are easily violated by methods of instrumental variable (IV) selection, potentially affecting different MR methods to an heterogenous extent. Recent HF MR work using proteomic data highlights this challenge, because despite common proteomic and GWAS data, there is little overlap in the proteins identified as putatively causal and previous loci implicated in HF GWAS[1,2,3,4]. Here, we use pQTL data from the UK Biobank Pharma Proteomics Project Consortium[5] and publicly available HF GWAS summary data[6] to identify the sensitivity of MR analysis to pQTL IV selection across commonly implemented MR methods. We focused on two methodology decisions that can introduce bias, identifying pQTLs that are 1) independent signals (LD pruning), and 2) not confounded by horizontal pleiotropy (inclusion or exclusion of *trans*-pQTLs). We iteratively filtered pQTL instruments across a range of r^2 thresholds used in the literature to assess the sensitivity of MR to IV independence. We then conducted MR analysis with *cis*-only instruments, defined by a range of distances from a protein encoding gene used in the literature, and tested the sensitivity of the analysis to including *trans*-pQTL instruments in concert with varying definitions of *cis*-acting instruments. We applied three commonly used MR analyses to understand the extent to which the results of each method were affected, IVW[7], MR-Egger[8], and GSMR[9]. We found all MR analysis methods were sensitive to IV selection methods to varying extents. Protein significance increased when multiple non-independent instruments were included for the gene. This effect was amplified when considering only *cis*-pQTLs, especially when the number of IVs used was limited. This sensitivity analysis reveals avenues for consensus-building in methodology that may improve accuracy and replication success, thus enriching the possibility successful drug target identification. Acknowledgements: UKBB-PPP Consortium, Paradigm4 Team, and Pharmalex Team Refs:1. Preprint-https://doi.org/10.1101/2022.04.14.22273877. PMID: 353005233. PMID: 368602794. PMID: 359640125. Preprint: https://doi.org/10.1101/2022.06.17.496443 6. PMID: 319194187. PMID: 298461718. PMID: 29335400 9. PMID: 28527048

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4367 Serum IgA and Gd-IgA1 levels associate with pulmonary phenotypes. The Multi-Ethnic Study of Atherosclerosis (MESA) Lung Study

Authors:

T. Pottinger[1], Z. Liu[1], L. Liu[2], K. Buschur[1], A. Manichaikul[3], J. L. Curtis[4], S. Rich[5], E. A. Hoffman[6], B. M. Smith[1], K. Kiryluk[7], R. Barr[1]; [1]Columbia Univ., New York, NY, [2]Columbia Univ., New York City, NY, [3]Univ Virginia, Charlottesville, VA, [4]Univ. of Michigan, Ann Arbor, MI, [5]Univ. of Virginia, Scottsville, VA, [6]Univ. of Iowa, Iowa City, IA, [7]COLUMBIA Univ. HEALTH Sci.S, New York, NY

Immunoglobulin A (IgA) protects mucosal surfaces against pathogens. IgA deficiency is associated with impaired lung function, and small airway remodeling; and is highly heritable. Subclinical levels of IgA in plasma could have similar correlations. IgA1, an isotype of IgA, is susceptible to inheritable deficits in O-glycosylation; the resulting galactose-deficient IgA1 (Gd-IgA1) is pathogenic in IgA-related vasculitis and nephropathy. Plasma Gd-IgA1 levels have been less investigated in pulmonary diseases. We hypothesized that there was a causal relationship between serum IgA (all isotypes) and lung function, airway wall thickness, and emphysema. **Methods:** The Multi-Ethnic Study on Atherosclerosis (MESA) recruited participants free of cardiovascular disease in 2000-02. IgA was measured in serum using ELISA assays. The MESA Lung Study performed spirometry and chest CT, upon which percent emphysema (-950 HU) and airway dimensions were measured. Linear regression models were used to analyze percent emphysema controlling for age, sex, genetic ancestry, smoking, pack-years, height, weight, and education. Generalized estimating equations were used to account for multiple airway measures per participant controlling for the covariates previously described. Mendelian randomization analyses were also conducted using MR MiSTERI. Variants associated with serum total IgA and Gd-IgA1 levels were extracted from the Trans-Omics for Precision Medicine (TOPMed) database for analysis. **Results:** Among 1,296 participants of diverse genetic ancestries with serum IgA and serum Gd-IgA1 levels and segmental CT measures, there were significant positive associations between log normalized serum Gd-IgA1 levels and segmental airway average minor inner diameter and average wall thickness (β=0.067; p = 0.034 and β=0.12; p = 0.012, respectively). In addition, there was a statistically significant negative association of percent emphysema with log normalized serum IgA levels (n=5,397; β=-0.08; p = 0.0045) and log normalized serum Gd-IgA1 levels (n=2,811; β=-0.089; p = 0.019). Lastly, MR analyses of emphysema and serum total IgA showed a causal association of the biomarker on the phenotype (n=4,153; β=-0.08; p = 0.01). **Conclusions:** In one of the first cohort studies analyzing serum Gd-IgA1 levels, we identified associations between reduced serum IgA concentration with percent emphysema (both total IgA and Gd-IgA1) and lower airway abnormalities (Gd-IgA1). In addition, MR analyses indicated a causal impact of serum total IgA on emphysema but not serum Gd-IgA1. These findings support the protective role of IgA and Gd-IgA1 in lung host defense and COPD pathogenesis.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4368 Sex, onset-age, and APOE-stratified genetic association studies of dementia reveal novel risk loci in African Americans

Authors:

R. Sherva[1], C. Zhu[2], R. Zhang[3], R. Hauger[4], MVP Cognitive Decline and Dementia During Aging Working Group, G. Schellenberg[5], M. Pericak-Vance[6], R. Mayeux[7], J. Haines[8], L-S. Wang[9], L. Farrer[10], M. Logue[11]; [1]Boston Univ. Sch. of Med., Boston, MA, [2]Boston Univ., Natick, MA, [3]VA Boston, Boston, MA, [4]VA San Diego, San Diego, CA, [5]Univ Pennsylvania Sch Med, Philadelphia, PA, [6]Univ. of Miami Miller Sch. of Med., Miami, FL, [7]Columbia Univ, New York, NY, [8]Case Western Reserve Univ, Cleveland, OH, [9]UPENN, Philadelphia, PA, [10]Boston Univ Sch Med, Boston, MA, [11]Natl. Ctr. for PTSD at VA Boston, Jamaica Plain, MA

Alzheimer's disease (AD) risk variants have been identified in European ancestry cohorts that only manifest in or have stronger effects at certain disease stages or in individuals with a specific sex or Apolipoprotein-E (*APOE*) genotype. Until recently, sample sizes in non-European ancestry cohorts have been underpowered to perform stratified analyses to detect these associations. We generated African ancestry (AA) genome-wide association study datasets stratified by sex, age at onset (<75 vs ≥75), and *APOE*-ε4 carrier status in participants from the Million Veterans Program (MVP) and the Alzheimer's Disease Genetics Consortium (ADGC). Outcomes in MVP were Alzheimer's Disease and related dementias (ADRD; n=4073 cases and 19,648 controls) and proxy dementia (i.e., reported dementia in a parent, n=6216 cases and 21,566 controls) while ADGC analyses examined AD (n=2425 cases and 5069 controls). Results from the proxy dementia GWAS were included the corresponding sex-stratified results based on the sex of the affected parent. Analyses were performed using logistic regression models adjusted for age, sex, and ancestry principal components. MVP and ADGC results were combined via sample size-weighted meta-analysis. Genome-wide significant associations were observed in an intergenic region near ephrin receptor A5 (*EPHA5*) (rs141838133, $p=2.19x10^{-8}$) in individuals with onset < 75 years, in glutamate ionotropic receptor NMDA type subunit 3B (*GRIN3B*) near the known AD risk gene *ABCA7* (rs115882880, $p=3.83x10^{-8}$) in females, and near thrombospondin type laminin G domain and EAR repeats (*TSPEAR*) (rs139130053, $p=4.27x10^{-8}$) in *APOE*-ε4 non-carriers. *EPHA5* has two possible AD-related functions: ephrin receptors modify the strength of existing synapses in the brain an in pancreatic islets EPHA5 affects basal and glucose-stimulated insulin secretion to improve glucose homeostasis. Another ephrin family member, *EPHA1*, is a well-established AD risk gene. It is unclear whether *GRIN3B* represents a locus distinct from *ABCA7*. The peak *GRIN3B* SNP is in weak to moderate LD with previously reported *ABCA7* SNPs. *GRIN3B* acts as an inhibitory subunit of CNS ion channels that decreases the current amplitude and calcium permeability of motor neurons. It has been linked to amyotrophic lateral sclerosis and schizophrenia. *TSPEAR* regulates Notch signaling but has not been linked to neuronal function. In summary, age, sex, and *APOE*-stratified analyses of dementia in AA participants from two cohorts revealed new associations. Stratified analyses may yield critical information about the genetic heterogeneity underlying dementia risk and lead to advances in precision medicine.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4369 Sex-specific association between anthropomorphic traits and intraocular pressure in the Canadian Longitudinal Study on Aging: A bidirectional Mendelian randomization study.

Authors:

J. Bastasic[1], E. Freeman[1], J. Little[1], P. G. Hysi[2], K. M. Burkett[3], M-H. Roy-Gagnon[1]; [1]Sch. of Epidemiology and Publ. Hlth., Univ. of Ottawa, Ottawa, ON, Canada, [2]Section of Ophthalmology, Sch. of Life Course Sci., King's Coll. London, London, United Kingdom, [3]Dept. of Mathematics and Statistics, Univ. of Ottawa, Ottawa, ON, Canada

Background: Observational studies have found evidence of association between anthropomorphic traits and intraocular pressure (IOP). It is important to determine whether these associations are causal and/or sex-specific in order to develop prevention and treatment approaches. Using Mendelian randomization (MR), we aimed to determine if anthropomorphic traits are causally related to IOP in the Canadian Longitudinal Study on Aging (CLSA).
Methods: We used the baseline data from the European descent subgroup of the CLSA participants with phenotype and genotype data (n=23,004 individuals after sample quality control exclusion). Anthropomorphic traits were based on measured values and included body mass index (BMI), waist-to-hip ratio, and waist-to-hip ratio adjusted for BMI. The outcome was defined as the treatment-adjusted corneal-compensated IOP. We used an MR two-stage least-squares regression analysis with instrumental variables defined as sex-specific polygenic risk scores for the three anthropomorphic traits. Polygenic risk scores were constructed based on published genome-wide association results.
Results: We observed significant associations ($p<0.05$) between all anthropomorphic traits and IOP after adjustment for age, sex, province of recruitment, education, total household income, marital status, smoking and alcohol consumption status. Using the same adjustment variables in the MR analysis, we did not find evidence of a causal association between the anthropomorphic traits and IOP at the 5% significance level. The estimated causal effect of BMI on IOP in males was close to significant ($p=0.0518$).
Conclusion: MR analyses found borderline significant evidence of a causal male-specific association between BMI and IOP. Further investigation of sex-specific associations with better defined instrumental variables would be important.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4370 Shared genetic associations between asthma and COPD stratified by age of onset of asthma.

Authors:

Y. Li[1], Y. Yang[1], D. Cornejo-Sanchez[2], G. Wang[2], S. M. Leal[2], A. T. Dewan[1]; [1]Yale Sch. of Publ. Hlth., New Haven, CT, [2]Columbia Univ., New York, NY

Asthma is the leading chronic disease for children and is also common among adults. Chronic obstructive pulmonary disease (COPD) is a leading cause of death and disability worldwide, primarily affecting individuals > age 65. Studies have suggested that patients with asthma and COPD experience more symptoms, frequent exacerbations, poorer quality of life, a more rapid decline in lung function, higher mortality, and greater use of healthcare resources, compared to those with either condition alone. Shared genetic risk factors may contribute to both asthma and COPD, and these factors may differ between adult and childhood onset asthma. This research was conducted using the UK Biobank. We included subjects who self-defined as White British and have similar genetic ancestry based on a principal components (PCs) analysis. Asthma was defined by either ICD-10 code or diagnosis by a doctor which was self-reported by the study subject. Childhood asthma was defined as asthma with age of onset ≤10 years (N=8,134) while adolescent/adult asthma (hereafter referred to as adult asthma) had onset >10 years (N=29,369). Controls (N=290,722) for childhood and adult asthma, did not include individuals with autoimmune disease. A proxy COPD status as previously described in PMID 32477647 was determined by the ratio of the best measures for forced expiratory volume in 1-second (FEV1) to forced vital capacity (FVC). Cases had a FEV1/FEV ratio < 0.70 (N=44,323) while controls had an FEV1/FEV ratio >0.70 (N=232,489). Imputed (INFO>0.8) and directly genotyped variants were included in the generalized linear mixed model as implemented in REGENIE. Age at recruitment, genetic sex, and 10 PCs were included in all the models and baseline smoking status was adjusted for in the COPD proxy model. FUMA 1.5.4 was used to annotate SNVs to genes and only SNVs with P < 5?10-8 were used to investigate the genetic overlap between traits. The number of significant SNVs were 3,186 for adult asthma, 5,685 for childhood asthma, and 7,594 for proxy COPD with a subset of these SNVs that mapping to 445, 624, and 990 genes, respectively. There was overlap of significant SNVs/genes between phenotypes: 1,875 SNVs (256 genes) for adult and childhood asthma; 177 SNVs (62 genes) for adult asthma and proxy COPD; 101 SNVs (67 genes) for childhood asthma and proxy COPD, and 89 SNVs (46 genes) were shared between all three phenotypes. For genes shared between all phenotypes, the number is almost double what is expected to be shared by chance (46 vs. 26.1). These data confirm that there is overlap in the genetic architecture of asthma and COPD, and this shared genetic etiology is largely irrespective of the age of onset of asthma.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4371 Shared genetic factors between type 1 diabetes and co-occurring autoimmune diseases.

Authors:

A. Nwagbata[1], A. Iakovliev[1], S. McGurnaghan[2], H. Colhoun[2], A. Spiliopoulou[1,2], P. McKeigue[1]; [1]Usher Inst., Univ. of Edinburgh, Edinburgh, United Kingdom, [2]Inst. of Genetics and Cancer, Univ. of Edinburgh, Edinburgh, United Kingdom

Background: People with type 1 diabetes (T1D) have a higher frequency of other autoimmune diseases (AIDs). Which causal mechanisms underlie this co-occurrence is unclear. Most studies investigating the genetic overlap between T1D and other AIDs have used independent cases for each trait. The aim of this study was to investigate if genetic factors associated with T1D are also associated with co-occurrence of other AIDs among people with T1D.
Methods: Among 4964 people with T1D in the Scottish Diabetes Research Network Type 1 Bioresource, we ascertained who also had hypothyroidism (915), pernicious anaemia (355), celiac disease (189), psoriasis (193), inflammatory bowel disease (187) and rheumatoid arthritis (184). Genetic markers for T1D comprised 23 genotypic scores computed by genome-wide aggregation of *trans*-effects (GATE) of common SNPs on gene expression in whole blood (9) and on circulating protein level (14). We previously showed that the GATE scores for these genes and proteins were strongly associated with T1D and supported an "omnigenic" model for its genetic architecture, where *trans*-effects of many SNPs coalesce on relatively few core genes. Logistic regression was used to test the association of these 23 GATE scores with case-control status for each co-occurring AID and the remaining T1D cohort as controls. Proteins were mapped to their corresponding gene names using Uniprot-ids.
Results: The GATE scores for the expression or protein level of 8 and 4 T1D core genes were associated with the risk of hypothyroidism and of pernicious anaemia, respectively ($P < 5.5 \times 10^{-3}$ for expression scores, $P < 3.5 \times 10^{-3}$ for protein scores). Genes *FOXP3*, *CD5* and *CD247* were associated with both hypothyroidism and pernicious anaemia. Genes *LGALS3BP*, *STAT1*, *CRTAM*, *CCL19* and *CD5L* were associated only with hypothyroidism, and *CCL15* was associated only with pernicious anaemia. These significant associations had the same effect direction on co-occurring AIDs as on T1D. Besides *CD247* and *LGALS3BP*, which play other significant roles in the immune system, these genes are involved in the induction and activity of CD4+ regulatory T cells (Tregs). There were no significant associations for the remaining AIDs, which can be attributed to moderate sample sizes.
Conclusion: We provide evidence of association between GATE scores for the expression or protein level of T1D core genes and the risk of co-occurring hypothyroidism and pernicious anaemia among people with T1D. This suggests shared immune system pathways between these AIDs and highlights possible causal mechanisms underlying this co-occurrence; the detected genes could serve as potential targets for drug development.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4372 Shared mutational burden within human Spina Bifida cohort and genetic similarities with comorbidities

Authors:

M. Strain; Duke Univ., Durham, NC

Neural tube defects (NTDs) are the most common congenital malformation of the brain and spinal column, affecting approximately 1/1000 births in the US annually. Despite strong evidence for a complex genetic architecture, the specific genes and inheritance models underlying NTD risk have not been fully elucidated. Most risk variants identified thus far have been private mutations in individual NTD families. In an effort to further identify genetic factors influencing the most prevalent NTD subtype, the Spina Bifida (SB) Sequencing Consortium has performed whole genome and exome sequencing of 139 patient-parent trios from the Duke NTD cohort. To identify genes that may harbor an excess of rare variants in SB, we compared the aggregate burden of rare variants (MAF ≤ 0.005) within gene bodies from our 139 SB probands against 776 non-Finnish European controls from the gnomAD data base. While controlling for population stratification, we identified three genes with increased mutational burden in the SB cases ($q < 0.05$). The most notable of these three was tubulin beta 3 (*TUBB3*; $p = 0.0003$), for which clinical variants linked to other brain malformations have been identified. We were also interested in whether there was a potential overlap between the genetic basis of hydrocephalus (HC) and SB. HC can occur idiopathically, but is also a common comorbidity of SB (31% of our SB cases; NTD+HC). We used summary statistics from a publicly available idiopathic HC genome-wide association study (GWAS) to generate polygenic risk scores for HC within our SB cohort. Using 18 SNPs surpassing a p-value threshold of 5e-4 from the HC GWAS, the resulting PRS explained nearly 30% of the variability in our SB cohort ($p = 0.003$). These data support the hypothesis that specific genes with a burden of rare variants contribute to NTD, as well as the overlap between the genetic basis of idiopathic HC and comorbid NTD+HC.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4373 Single Nuclei Transcriptome-Wide Association Study Atlas of the Human Brain Uncovers Cell-Type-Specific Contributions to Schizophrenia and Alzheimer's Disease

Authors:

S. Venkatesh[1], G. Hoffman[1], D. Lee[1], D. Mathur[1], P. NM[1], B. Zeng[1], A. Hong[1], C. Casey[1], M. Alvia[1], Z. Shao[1], S. Argyriou[1], K. Therrien[1], P. Auluck[2], D. Bennett[3], S. Marenco[2], V. Haroutunian[1], J. Bendl[1], J. Fullard[1], G. Voloudakis[1], P. Roussos[1]; [1]Icahn Sch. of Med. at Mount Sinai, New York, NY, [2]Natl. Inst. of Mental Hlth., Bethesda, MD, [3]Rush Univ. Med. Ctr., Chicago, IL

Transcriptome-Wide Association Studies (**TWAS**) complement Genome-Wide Association Studies (**GWAS**) by integrating tissue specific variation, uncovering mechanistic effects of genomic loci (especially in non-coding regions), and by adding power to discover associations between genomic features and phenotypes. However, until now, TWAS efforts have been mostly limited to tissue homogenate specimens, from e.g. dorsolateral prefrontal cortex (**DLPFC**) or heart. To address this, we built cell-type-specific transcriptomic imputation models (**TIMs**) using DLPFC single-nuclei RNA sequencing (**snRNAseq**) from genotyped individuals to perform TWASs with single cell resolution for schizophrenia (**SCZ**) and Alzheimer's Disease (**AD**).

We utilize the psychAD cohort that includes snRNAseq in more than 6 million nuclei from the DLPFC of 1,495 unique donors. We use PrediXcan to build TIMs, S-PrediXcan to perform summary-level TWAS, and custom scripts to perform Phenome-Wide Association Studies (**PheWAS**) in the Million Veteran Program (**MVP**).

Cell-type-specific TIMs predict several thousand genes in 36 cell types. Using the latest Psychiatric Genetics Consortium GWAS for SCZ, we find over 300 significant associations across over 200 unique genes. Of note, the majority of these significant associations came from excitatory neurons and oligodendrocytes. We then compare SCZ TWAS of cell-type-specific TIMs with a previously established TIM for DLPFC. Among genes with a prediction performance R^2 of 0.01 or greater in both TIMs, TWASs of excitatory neurons and oligodendrocytes have the highest correlation with tissue homogenate (0.55 and 0.40 respectively). Using AD summary statistics, we find over 120 significant associations across over 80 significant genes. Moreover microglia TWAS of AD correlates strongly with a fresh sorted microglia TWAS of AD (0.82). Finally, we perform PheWAS analysis on critical genes selected from summary-level TWAS in MVP.

Cell-type-specific TIMs find unique contributions of particular cell types to SCZ and AD. Of particular interest are genes that have opposite effect sizes in different cell types. For example, *ASPHD1* has significant but opposite effects in association with SCZ between astrocytes and oligodendrocytes. This is consistent with the literature, which finds that *ASPHD1* overexpression and knock-out are associated with SCZ. This suggests our approach can identify associations that would otherwise be lost in bulk tissue analyses. Finally, the high correlation between our frozen snRNAseq pseudobulk microglia AD TWAS and fresh bulk RNAseq sorted microglia AD TWAS lends credibility to our TIMs.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4374 † Single nucleus multi-omics profiling reveals epigenomic and transcriptional dynamics in Alzheimer's disease

Authors:

Y. Gong[1], X. Zhang[1], C. Qiu[1], K. Su[1], Q. Tian[2], H. Shen[3], H-W. Deng[4]; [1]Tulane Univ., New Orleans, LA, [2]Tulane Univ., Mandeville, LA, [3]Tulane Univ, New Orleans, LA, [4]Tulane Univ, Mandeville, LA

Alzheimer’s disease (AD) is a debilitating neurodegenerative disorder characterized by progressive cognitive decline and memory loss. The intricate molecular mechanisms underlying AD pathology remain poorly understood, necessitating innovative approaches to elucidate key regulatory factors. In this study, we employed single nucleus Assay for Transposase-Accessible Chromatin sequencing (snATAC-seq) and single nucleus RNA sequencing (snRNA-seq) to investigate the epigenomic and transcriptomic landscapes in AD-affected human prefrontal cortex. Our analysis focused on the snATAC and RNA -seq from 18 AD cases (10 female and eight male) ranging from 77 to 90 and 15 age-matched controls. Following the integration analysis of chromatin accessibility and transcriptional profiles, we identified seven distinct cell clusters corresponding to astrocytes, oligodendrocytes, excitatory neurons, inhibitory neurons, oligodendrocyte progenitor cells, endothelial cells, and microglia. Delving deeper through differential expression analysis, we unveiled unique epigenetic and transcriptional signatures tied to AD. These encompassed genes involved in synaptic function, neuroinflammation, and amyloid-beta metabolism. Intriguingly, we noticed substantial dysregulation of transcription factors and CD44 in astrocytes, hinting at an intricate nexus of epigenetic and transcriptional regulatory networks implicated in AD pathogenesis. In summary, our study provides a comprehensive single-nucleus multi-omics perspective on the epigenomic and transcriptomic alterations in brains afflicted by AD. This insight significantly enhances our understanding of the molecular pathways implicated in AD pathogenesis, offering potential targets for future therapeutic intervention

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4375 Sixteen years' experience of *SMA* carrier screening and prenatal diagnosis in Iranian populations.

Authors:

A. Khanbazi[1,2], M. Beheshtian[1,2], M. Azad[1], M. Akbari[1], A. Kariminejad[1], K. Kahrizi[2], H. Najmabadi[1,2]; [1]Kariminejad - Najmabadi Pathology & Genetics Ctr., Tehran, Iran, Islamic Republic of, [2]Univ. of Social Welfare and Rehabilitation Sci., Tehran, Iran, Islamic Republic of

Spinal muscular atrophy (SMA), generally characterized by muscle weakness and degeneration, is caused by the deterioration of motor neurons in the spinal cord due to the deletion or mutation in the *SMN1* gene on chromosome 5. To promote genetic counseling and enable families to make more informed decisions, determining SMA carrier status in multi-ethnic populations with a high rate of consanguineous marriage such as Iran is crucial. Therefore, in this study we aimed to investigate the prevalence of SMA carriers among individuals referred to Kariminejad - Najmabadi Pathology & Genetics Centre for carrier screening, between 2006-2022, by physicians, healthcare centers, and genetic laboratories from across Iran. Additionally, we report the results of prenatal diagnosis (PND) tests conducted in this Centre.We calculated the carrier frequency of SMA among unrelated individuals who did not have a confirmed or probable family history of the disease. Ninety-six percent (n=569) of tests were done by Multiplex ligation-dependent probe amplification (MLPA) (MRC Holland) and the remaining (n=23; 4%) were conducted by quantitative RT-PCR.A total of 592 individuals met our criteria. The mean age (SD) of investigated cases was 30 (±7) years. The carrier frequency of SMA was estimated to be 4.5% (CI=0.95%, 3.1-6.4). Among normal individuals, 87% (n=516) had two copies, 7% (n=43) had three copies, and 1% (n=6) had four copies of *SMN1*. PND was performed on 331 fetuses; of which 21% (n= 70) were affected, 28% (n=91) were normal, and 51% (n=170) were carriers. 188 out of 257 families (73%), had a history of at least one affected offspring, and parents were considered as obligate carriers. However, in 12 families (5%), one of the parents had a normal copy number of *SMN1* that might be suggestive of silent carriers.Considering the historical influx and the subsequent genetic diversity in Middle east countries including Iran, our result showed a higher rate of carrier frequency than that reported from European or East Asian populations (~1.8%) and aligned more closely with nations possessing similar cultural backgrounds including consanguineous marriage, specifically Morocco (4%) and Saudi Arabia (2.6%). Therefore, providing genetic counseling and premarital screening for couples in such countries along with prenatal diagnosis among at-risk couples plays a pivotal role in preventing the births of children affected by SMA. Furthermore, our result emphasizes the necessity for considering the risk of a silent carrier state among individuals with normal results.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4376 Statistical challenges with GWAS on ratio traits obscure the interpretation of many published associations

Authors:

T. Soare[1], Z. McCaw[1], H. Somineni[1], S. Mukherjee[1], K. Sandor[1], Insitro Research Team, P. Tagari[1], D. Koller[1], G. Davey Smith[2], D. MacArthur[3,4], C. O'Dushlaine[1]; [1]Insitro, South San Francisco, CA, [2]Univ. of Bristol, Bristol, United Kingdom, [3]Garvan Inst., Sydney, Australia, [4]Murdoch Children's Res. Inst., Melbourne, Australia

Ratio traits are metrics defined by a numerator trait and a denominator trait, and are widely used clinically to assess health (e.g. BMI, waist-to-hip ratio). Such traits have been a target for many genome-wide association studies (GWAS), generally with the implicit or explicit goal of uncovering the genetic underpinnings of the numerator while controlling for the denominator. In the GWAS catalog, ratio traits represent 3% of traits analyzed and 8% of reported associations.

However, there are conceptual problems with the ratio approach. Using a naive ratio as the outcome does not actually test the direct effect of genotype on the numerator while holding the denominator constant, and the genetic effect estimate in the ratio model is difficult to interpret: as the association may be due to the numerator alone, the denominator alone, or both the numerator and the denominator. Further, the genetic effect estimate is likely to be biased. Such issues could be avoided by simply testing the numerator trait as the outcome and using the denominator trait as an additional covariate ("adjusted model").

To explore the impact of issues with the ratio model on published GWAS results, we use waist circumference, hip circumference, and their ratio (WHR) to assess central adiposity. Permuting waist circumference (the numerator) while leaving hip circumference unchanged, we find the ratio model has a 5-10% false positive rate, identifying 77 spurious genome-wide significant loci in the UKBB (N=350k). In contrast, the adjusted model effectively controls type I error (0 genome-wide significant loci). False positive variants are enriched for denominator associations, suggesting the false positive rate is driven by the heritability of the denominator. A properly adjusted model identifies 20% more loci than the ratio model, and variants identified by both models have stronger associations in the adjusted model, showing the adjusted model is better powered to detect direct effects on the numerator.

We find similar effects in other ratio traits, suggesting that published GWAS findings are substantially affected by this statistical issue. We recommend re-analysis of ratio-based GWAS results using a covariate adjustment model in order to reduce false positive associations, maximize power, and aid in biological interpretation.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4377 Statistically robust familial variant analysis

Authors:

A. Janiszewski, I. Protopopov, J. Behr, X. Bonilla, N. Arrigo, Z. Xu; SOPHiA GENETICS, Rolle, Switzerland

Familial variant analysis (FVA) can be a powerful tool for the identification of disease-causing germline variants. FVA investigates variant segregation patterns and excludes variants that do not follow the inheritance patterns of interest, reducing the number of candidate disease-causing variants that need to be further assessed. However, the sequencing noise present in large-panel sequencing may cause errors in variant zygosity estimation, hindering the filtering ability of FVA. This effect is expected to grow with increasing pedigree size, as sequencing errors in one individual can negatively affect the analysis of the entire family. Here, we develop an FVA algorithm that combines sequencing data and affection status of family members in a statistically robust way and guarantees full control over the false negative rate. Our approach is based on characterizing the consistency of sequencing data with the expected variant zygosity, in each family member, for all inheritance modes available, and representing it as a set of p-values with the subsequent combination of the relevant scores using Fishers p-value combination procedure. This approach supports the analysis of pedigrees of any size, and data missingness: it leverages on affection status of a family member even if sequencing data is unavailable, extending the use of FVA in clinical sample analysis. Using in-silico simulations of family data, we demonstrate the controlled false negative rate of our algorithm and estimate its efficiency defined as the fraction of variants that do not follow the tested segregation pattern that are correctly filtered out. The experiments cover autosomal recessive, autosomal dominant, X-linked and compound heterozygous variants propagating in pedigrees with up to 3 generations and siblings. For each case, we simulated 100 families with average read coverage of 30 and 100 and defined noise model. We show that the filtering efficiency increases with increasing availability of affection status and/or sequencing data of individuals. For example, filtering power for autosomal dominant variants increased by 16,2% if grandparents' affection status was added to a trio analysis, and by 45% if grandparents' sequencing data were also added. Altogether, our FVA algorithm provides enhanced variant filtering capabilities for identification of disease-causing variants from large NGS data such as whole-exome or whole-genome sequencing. Combined with the ability to analyse pedigrees of any size and to include individuals for which only affection status is available, our algorithm provides means to improve the efficiency and accuracy of variant analysis and interpretation.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4378 Studies of Linkage and Polygenic Risk to Improve Genetic Testing for Cardiovascular Disease Across Populations

Authors:

G. Tietz[1], A. Ellison[1], T. Tan[1], N. Shah[1], S. Behera[1], J. Belyeu[2], R. Gibbs[3], F. Sedlazeck[3], E. Atkinson[1]; [1]Baylor Coll. of Med., Houston, TX, [2]Illumina, Inc., San Diego, CA, [3]Baylor Coll. Med., Houston, TX

Cardiovascular disease (CVD) is the leading cause of death in the United States. Genetics can influence CVD risk through regulating lipids like lipoprotein A (Lp(a)), which can double an individual's risk for heart attack at elevated levels. Given its high heritability of 90%, Lp(a) is especially promising for genetic testing and several variants within the gene *LPA* are integrated into a clinical gene panel to identify those at risk of CVD. However, recent findings have identified poor predictive transferability of these variants to Hispanic individuals, which may be explained by different patterns of genetic variation and linkage across ancestries impacting tagging of a shared causal locus.

To combat this inequity, we aim to refine genetic testing for Lp(a) through ancestry-informed analyses of linkage structure in *LPA*, combined with the development of a novel polygenic scoring model accounting for both a single locus of large effect with polygenic variation. Considering both with an ancestry-aware approach, we observed improved prediction for populations in which genetic testing for CVD currently underperforms.

Here, we start by investigating linkage in Latin American (LAT) and European (EUR) individuals of the 1000 Genomes Project. We identified a recombination event within *LPA* with population-specific patterns where recombination rates decreased with the proportion of EUR ancestry present in LAT individuals. This recombination event differentially disrupts linkage of a clinically used *LPA* variant to KIV-2, a critical locus in *LPA* that is infrequently called given its repetitive nature. In the context of the variants within *LPA* that are clinically tested, this suggests that their decreased predictiveness may be explained by more frequent recombination events decreasing their tagging of KIV-2.

We then evaluated whether considering polygenic burden with a KIV-2 proxy locus would improve predictive performance across populations. First building an Lp(a) PRS on EUR individuals of the UK Biobank, we then added consideration of variants associating well with KIV-2 across populations to generate an extended PRS model. Through testing and benchmarking, we identified improved prediction accuracy of Lp(a) levels across diverse ancestries with our new extended PRS model.

Collectively, our work proposes new approaches to improve transferability of genetic testing for CVD across populations through analyses of structure, and integration of polygenic burden and a locus of large effect into a novel model. Building from this work will help narrow a current gap in genetic testing and ensure that the promise of precision medicine is extended to all populations.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4379 † Subcontinental admixture in Europeans and European Americans can affect GWAS findings and polygenic score performance.

Authors:

M. Gouveia[1], T. Leal[2], K. Meeks[1], E. Tarazona-Santos[3], C. Bustamante[4], A. Adeyemo[1], C. Rotimi[1], D. Shriner[1]; [1]NIH, Bethesda, MD, [2]NIH, Cleveland Clinic, OH, [3]Univ.e Federal de Minas Gerais, Belo Horizonte, Brazil, [4]Stanford Univ. Sch. of Med., Stanford, CA

Failing to account for local ancestry, which is a result of admixture, can introduce residual confounding and thereby affect studies of association, polygenic adaptation, and polygenic risk scores. However, admixed populations are often considered as those resulting from recent interbreeding between continentally separated populations. Thus, European-ancestry populations are recognized as stratified but not as admixed. We recently showed that European-ancestry individuals are three-way admixed at the European subcontinental level. Using the classical *LCT*-height false positive association, we demonstrated that adjusting for both global and local ancestry enables the identification of spurious associations. In the present study, we evaluated the impact of ancestry adjustment in genome-wide association studies (GWAS) findings and polygenic score (PGS) performance. We performed a GWAS of height in 17,684 European Americans (EA) with different levels of ancestry correction, namely unadjusted, global, and global + local (full). Previously reported GWAS associations ($p < 5x10^{-8}$) that were significant in our analysis unadjusted for ancestry but not significant after full adjustment were considered likely false positives. Following this approach, we interrogated 1,147 independent associations for height reported in the GWAS Catalog and identified 75 (6.5%) likely false positives due to uncorrected ancestry stratification. Two associations (at rs2857693 and rs10838798) showed strong residual confounding by local ancestry, i.e., they were replicated in analyses unadjusted and adjusted for global ancestry, but lost significance after full ancestry adjustment. To assess the utility of our approach in identifying spurious associations, we calculated PGS for height in our EA dataset using GWAS summary statistics from the GIANT consortium (> 4 million European ancestry individuals). The PGS estimated from loci in GIANT that we identified as likely false positives had 5-fold lower predictive value than PGS estimated from randomly selected sets of GIANT height associations ($p < 0.001$). Overall, our findings provide evidence that spurious associations in GWAS resulting from failure to fully account for population stratification can have substantial downstream consequences, including poorer PGS performance. Additionally, accounting for subcontinental admixture minimizes the risk of follow-up functional analysis of novel but false positive findings.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4380 Synthetic Slope Analysis Empowers GWAS of Physiological Decline

Authors:

Z. McCaw, D. Amar, C. O'Dushlaine, T. Soare; Insitro, South San Francisco, CA

Certain measures of physiological function are known to deteriorate with advancing age. Examples include grip strength, lung function, and cognitive performance. Accelerated decline of these variables is often a hallmark of disease, such as amyotrophic lateral sclerosis (ALS). Understanding the genetic basis of accelerated physiological decline may be key to patient risk stratification and drug target identification.

Performing genome-wide association studies (GWAS) for the rate of physiological decline is hindered by the paucity of genotype-phenotype cohorts with extended longitudinal follow-up. For example, in the UK Biobank, only ~10% of subjects have follow-up measurements. Machine learning (ML) models can be trained to predict a subject's rate of decline from baseline covariates. However, the accuracy of such models is fundamentally limited because predicting a rate of change from a single time point is an ill-posed problem. Limited predictability poses a significant challenge for imputation or proxy GWAS.

Recently, synthetic surrogate analysis (SynSurr) was developed for leveraging an ML-derived surrogate phenotype to improve GWAS on a partially missing target phenotype. Here, we use the predicted rate of decline as a synthetic surrogate for the incompletely observed empirical rate of decline, an application of SynSurr we described as Synthetic Slope Analysis (SSA). SynSurr has several advantages over imputation-based GWAS. Unlike imputation, SynSurr does not assume that the predicted and observed rates of decline come from the same distribution. Rather, SynSurr jointly models the observed and predicted rates within a bivariate regression framework. SynSurr estimates the same effect size as standard GWAS of the observed rate of decline; properly controls the type I error, even when the predicted rate of decline is unreliable; and improves power in proportion to the correlation between the observed and predicted rates of decline.

We apply SSA to empower GWAS on the rates of decline for several traits known to deteriorate in ALS. Compared with standard GWAS on the observed rate of decline, SSA identifies more genome-wide significant associations. We assess the biological relevance of the associations identified by SSA to ALS and neurological disease more broadly.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4381 Systematic Integration of Multi-omics Data for the Study of Coronary Artery Disease and Subclinical Atherosclerosis

Authors:

C. Yang[1], F. Aguet[2], K. Ardlie[2], W. Post[3], R. Gerszten[4], H. Wheeler[5], K. Taylor[6], C. McNamara[7], C. Miller[1], S. Bekiranov[1], S. Kasela[8], T. Lappalainen[9], H. Im[10], P. Durda[11], C. Johnson[12], X. Guo[6], Y. Liu[13], J. Polak[14], D. Herrington[15], C. Clish[2], D. Van Den Berg[16], R. Tracy[11], E. Cornell[11], T. Blackwell[17], G. Papanicolaou[18], J. Rotter[19], S. Rich[1], A. Manichaikul[20]; [1]Univ. of Virginia, Charlottesville, VA, [2]Broad Inst. of Harvard and MIT, Cambridge, MA, [3]Johns Hopkins Univ, Baltimore, MD, [4]Div. of Cardiovascular Med., Beth Israel Deaconess Med. Ctr., Boston, MA, [5]Loyola Univ Chicago, Chicago, IL, [6]Dept. of Pediatrics, The Inst. for Translational Genomics and Population Sci., The Lundquist Inst. for BioMed. Innovation at Harbor-UCLA Med. Ctr., Torrance, CA, [7]Beirne B. Carter Ctr. for Immunology Res., Cardiovascular Div., Dept. of Med., Charlottesville, VA, [8]New York Genome Ctr., New York, NY, [9]SciLifeLab & NY Genome Ctr., New York, NY, [10]Univ. of Chicago, Chicago, IL, [11]Dept. of Pathology and Lab. Med., The Robert Larner, M.D. Coll. of Med., Univ. of Vermont, Burlington, VT, [12]Univ. of Washington, Seattle, WA, [13]Duke Univ., Chapel Hill, NC, [14]Dept. of Radiology, Tufts Univ. Sch. of Med. and Boston Univ. Sch. of Med., Boston, MA, [15]Div. of Cardiology, Dept. of Internal Med., Wake Forest Univ. Hlth.Sci., Winston-Salem, NC, [16]Univ. of Southern California, Los Angeles, CA, [17]Univ. of Michigan, Ann Arbor, MI, [18]NHLBI, Bethesda, MD, [19]Lundquist Inst., Harbor-UCLA Med Ctr, Torrance, CA, [20]Univ Virginia, Charlottesville, VA

Coronary artery disease (CAD) is a leading cause of death and disability worldwide. Prior genome-wide association studies (GWAS) of CAD have identified over 400 independent loci. However, the molecular consequences of the GWAS variants and their relevance to subclinical atherosclerosis have not been explored comprehensively in human cohorts. In this study, we aimed to prioritize the CAD-related molecular targets and investigate their relationship with subclinical atherosclerosis using molecular 'omics data from TOPMed Multi-Ethnic Study of Atherosclerosis (MESA). We performed Bayesian colocalization to identify the genes underlying genetic loci from a multi-ancestry GWAS of CAD (Tcheandjieu et al. 2022) using transcriptomic data from peripheral blood mononuclear cells (PBMCs) in MESA, a multi-ethnic longitudinal genetic epidemiology cohort. To prioritize CAD-related genes, we performed additional follow-up of the colocalized genes using (a) Colocalization with GWAS of subclinical atherosclerosis (CAC and cIMT), (b) regression analysis of subclinical atherosclerosis on measured expression in MESA, (c) examination of protein QTL (pQTL) and methylation QTL (mQTL) evidence for the CAD causal variants and (d) phenotypic consequences of gene knockouts in mice. We also performed weighted gene co-expression network analyses (WGCNA) on the MESA transcriptomics data to identify modules of highly correlated genes and their association with subclinical atherosclerosis in MESA. Based on our Bayesian colocalization, we identified 54 putative causal genes underlying CAD-assocatied loci. Among these, we observed 10 genes showing at least three forms of evidence in follow-up analyses,
including *PLEKHJ1*, *LAYN*, *BICC1*, *SCARB1*, *PAN2*, *CCDC30* and *CARCAL*. For example, *PLEKHJ1* shows association with CAC and cIMT in MESA. *BICC1*, *SCARB1* and *PAN2* show significant association with heart/cardiovascular traits in mice. WGCNA identified four modules associated with subclinical atherosclerosis in MESA. Overlapping our 54 colocalized genes with the modules from WGCNA, we found one module was significantly enriched for colocalized genes. Pathway analysis of genes from this module highlighted common features among the identified genes, including encoding components of the complement system and inflammatory response. Our study identified multiple candidate genes at GWAS loci that are supported by molecular QTL, and additional follow-up analyses predicted effects of these genes on CAD and subclinical atherosclerosis in both humans and mice. Additionally, our study showed the value of network analysis further characterize pathways implicated by GWAS genes.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4382 TargPred: a web-based platform to uncover SNP-gene associations within genomic regulatory blocks

Authors:

F. Uljanic, K. Mandic, D. Hrsak, A. Baresic; Ruder Boskovic Inst., Zagreb, Croatia

The quest to pinpoint genetic variants tied to complex diseases and traits has led to an extensive number of genome-wide association studies (GWAS). These studies have uncovered countless genetic variants, but the task of connecting non-coding single nucleotide polymorphisms (SNPs) to their affected protein-coding genes remains challenging. Within the genome, segments known as Genomic Regulatory Blocks (GRBs) are populated with highly conserved non-coding elements. These elements exhibit preserved gene order, or synteny, across a range of vertebrate species, suggesting that the genes situated inside the GRBs are under long-range regulation.
In this research, we introduce TargPred, an online tool that unlocks the potential of linking regulatory SNPs within GRB regions to their putative target genes. This innovative approach leverages previously established associations between enhancer and promoter transcriptional activity within the GRB, and likely target genes for each GRB. TargPred, an open-access tool, provides a comprehensive view of genomic landscapes and highlights enhancer-promoter associations in genes that are under long-range regulation for 2544 traits from the GWAS catalog. An application of TargPred is showcased in a schizophrenia population study where the target gene, initially identified based on its proximity to a GWAS-identified SNP, was reassessed through the lens of our GRB-focused methodology.
TargPred aims to empower researchers to explore patterns of heritability in complex traits and diseases, prioritize drug targets, and uncover genetic features that are primarily governed by long-range regulation. This tool, we hope, will offer new insights into disease and trait patterns, further illuminating the complex nature of genetic regulation. Targpred is available at: targpred.irb.hr.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4383 Testing for differences in Hardy-Weinberg Disequilibrium between groups

Authors:

E. Garg[1], J. Romain[2], L. Sun[2], A. Paterson[1]; [1]Hosp. for Sick Children, Toronto, ON, Canada, [2]Univ. of Toronto, Toronto, ON, Canada

Testing for deviation from HWE (Hardy-Weinberg Equilibrium) is a standard quality control approach before conducting genetic analyses. However, deviation from HWE can be caused either by technical biases or violation of numerous assumptions related to population structure. Therefore, assessing HWE differences between groups is important. In our pilot work, we assessed HWE across groups in the Illumina 2.5M array data from the 1000 Genomes Project phase 3 data. Further, we combined group results by using inverse-variance and second-order meta-analysis. Additionally, we compared group results using Cochran's Q to test for variability among the groups. We identified 26 autosomal SNPs with MAF>5% with significant HWE (p<E-8) in one or more population or sex groups, and further with heterogeneity in effect sizes between ancestry- or sex-based groups. The most common basis for heterogeneity was evidence for deviation in males, but not females. BLAST of the sequence around such SNPs typically identified similarity to the recently released T2T chromosome Y sequence, which contains 30Mb of sequence missing from previous references, implying that they are paralogous sequence variations (PSV). We aim to implement similar analysis using the T2T (telomere-to-telomere v2) aligned high-coverage 1000 Genomes Project data. We hypothesize that the T2T data will have less technical biases, and thus allow us to better investigate heterogeneity across groups arising due to population structure.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4384 The association between Copy number variations, Congenital Heart disease, and the risk of Mental Illness.

Authors:

S. Sabtiy[1], s. vaez[2], S. Montalbano[1], T. Werge[1], A. Ingason[3], D. Helenius[1]; [1]Inst. of Biological Psychiatry, Roskilde, Denmark, [2]Inst. of biological psychiatry, Roskilde, Denmark, [3]Inst. for biological psychiatry, Roskilde, Denmark

<Objective: This study aims to elucidate an unbiased, population-based association between recurrent copy number variations (CNVs), congenital heart disease (CHD), and the risk of developing a psychiatric disorder (PD). Furthermore, it seeks to determine whether the presence of CNVs confounds this association, accounting for other relevant birth factors. **Background:** The number of adult survivors with CHD is steadily increasing. Thus, there is a growing awareness of the significant risk of neurodevelopmental issues in children and adults affected by CHD. Recent studies have highlighted the role of CNVs, a well-known genetic risk factor, in CHDs and PDs, including attention-deficit/hyperactivity disorder (ADHD), schizophrenia (SCZ), SCZ spectrum disorder, affective disorder, autism (ASD), and bipolar disorder (BDP). Despite previous studies investigating the association between CNVs and CHD, there remains a significant gap in obtaining unbiased, population-based estimates for the true extent of this association. Also, there is a notable knowledge gap regarding the underlying genetics contributing to the association between CHD and PDs. **Method:** We used a Danish population-based case-cohort study, utilizing the extensive Lundbeck Foundation Initiative for Integrative Psychiatric Research (iPSYCH) 2015 database. The iPSYCH2015 case-cohort study encompassed a study base of N=1,657,449 singleton births between 1981 and 2008, with known maternal residence in Denmark and a follow-up period extending until 2015. A population cohort consisting of N=50,615 individuals was randomly selected, and it overlapped with cases identified with ADHD (N=29 668), schizophrenia (SCZ) (N=8 113), SCZ spectrum disorder (N=16 008), ASD (N=24 975), BDP (N=3 819), and affective disorder (3 819). CNVs had been previously identified from the iPSYCH2015 database, and relevant clinical information was integrated into the analysis. The associations were analyzed through rigorous statistical analysis using Weighted Cox proportional hazards models to estimate the risks of each PD associated with CHD and each CNV associated with CHD while accounting for other relevant birth factors. **Findings:** The findings of this study validate the increased risk of CHD in individuals carrying CNVs at 22q11.2 and 2q13 loci. Additionally, this study demonstrates a statistically significant association between CHD and an increased risk of ADHD and ASD within an unbiased random population cohort. Our initial results indicate that the two risk factors of PDs, CHD and carrier status, act independently.>

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4385 † The expected polygenic risk score (ePRS) framework: an equitable metric to quantifying polygenetic risk via modeling of ancestral makeup.

Authors:

Y-J. Huang[1,2,3], N. Kurniansyah[3], M. Goodman[3], X. Guo[4], P. de Vries[5], H. Chen[6], E. Boerwinkle[6], N. Smith[7], L. Raffield[8], G. Peloso[9], L. Lange[10], B. Psaty[11], J. Bis[12], J. Brody[7], S. Zoellner[13], J. Wang[14], A. Stilp[7], C. Laurie[7], N. Min[15], J. Smith[14], W. Zhao[14], C. Kooperberg[16], S. Redline[3], S. Rich[17], M. Fornage[6], R. Kaplan[18], N. Franceschini[19], D. Levy[20], A. Morrison[6], J. Rotter[21], T. Sofer[1,2,3,22]; [1]Harvard Med. Sch., Boston, MA, [2]Beth Israel Deaconess Med. Ctr., Boston, MA, [3]Brigham and Women's Hosp., Boston, MA, [4]The Lundquist Inst., Torrance, CA, [5]Univ. of Texas Hlth.Sci. Ctr., Human genetics center, Houston, TX, [6]Univ. of Texas Hlth.Sci. Ctr. at Houston, Houston, TX, [7]Univ. of Washington, Seattle, WA, [8]UNC - Chapel Hill, CHAPEL HILL, NC, [9]Boston Univ., Boston, MA, [10]Univ Colorado Denver, Aurora, CO, [11]Dept. of Epidemiology, Univ. of Washington, Seattle, WA, [12]Cardiovascular Hlth.Res., Univ. of Washington, Seattle, WA, [13]Dept.s of Biostatistics and Psychiatry, Univ. of Michigan, Ann Arbor, MI, [14]Univ. of Michigan, Ann Arbor, MI, [15]Univ. of Mississippi Med. Ctr., Jackson, MS, [16]Fred Hutchinson Cancer Ctr., Seattle, WA, [17]Univ. of Virginia, Scottsville, VA, [18]Albert Einstein Coll. of Med., Bronx, NY, [19]Univ North Carolina at Chapel Hill, Chapel Hill, NC, [20]NHLBI/NIH, Framingham, MA, [21]Lundquist Inst., Harbor-UCLA Med Ctr, Torrance, CA, [22]Harvard T.H. Chan Sch. of Publ. Hlth., Boston, MA

Introduction: Polygenic risk scores (PRS) are useful for predicting and quantifying genetic risk. However, PRS depends on genetic ancestry due to differences in allele frequencies between ancestral populations. This leads to implementation challenges when analyzing diverse populations. First, because race and ethnicity-defined groups sometimes have different mixes of genetic ancestry proportions, distributions of PRS may differ across such groups, potentially leading to wrong interpretations when stratifying individuals into risk groups. Second, while adjustment for principal components (PCs) is an effective way to account for population stratification when estimating PRS effects, PCs do not capture equivalent information when constructed using different datasets. We aim to propose a comprehensive framework that can calibrate PRS based on one's ancestral makeup, protect from population stratification in association analysis, and provide an equitable way to quantify genetic risk across diverse populations.
Methods: We propose an individual-level metric called "expected PRS (ePRS)", defined as the expected value of a PRS based on one's global or local admixture patterns. Further, we define the "residual PRS (rPRS)", as measuring the deviation of the PRS from its expected mean (the ePRS) according to admixture patterns. Because ePRS captures the ancestral composition of individuals, we hypothesize that adjusting for ePRS in PRS-outcome association analysis will account for population stratification.
Results: Simulation studies confirm that it suffices to adjust for ePRS to obtain nearly unbiased estimates of the PRS-outcome associations, without further adjusting for PCs. Using TOPMed dataset, the estimated effect size of the rPRS adjusting for the ePRS is similar to the estimated effect of the PRS adjusting for 11 genetic PCs. For diastolic blood pressure, African American individuals had higher PRS values than individuals self-reported from other race/ethnic groups. By calibrating these PRSs via ePRS, rPRS values have similar distributions across self-report race/ethnic groups.
Conclusions: The ePRS framework provides a strategy to differentiate individual genetic risk from differences in PRS distributions due to ancestral makeup. Unlike PCs, the interpretation of ePRS is intuitive, and the meaning of ePRS is identical across datasets while protecting from population stratification bias in association analysis.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4386 The extent to which augmenting extant reference panels with population-specific sequences improves imputation quality

Authors:

J. Carlson[1], M. Krishnan[1], S. Liu[1], K. Anderson[1], J. Zhang[1], H. Cheng[2], T. Naseri[3], M. Reupena[4], S. Viali[5], R. Deka[2], N. Hawley[6], S. McGarvey[7], D. E. Weeks[8], R. L. Minster[1]; [1]Univ. of Pittsburgh, Pittsburgh, PA, [2]Univ. of Cincinnati Coll. of Med., Cincinnati, OH, [3]Naseri & Associates Publ. Hlth.Consultancy Firm and Family Hlth.Clinic, Apia, Samoa, [4]Lutia i Puava ae Mapu i Fagalele, Apia, Samoa, [5]Natl. Univ. of Samoa, Apia, Samoa, [6]Yale Sch. of Publ. Hlth., New Haven, CT, [7]Brown Univ., providence, RI, [8]Univ of Pittsburgh, Pittsburgh, PA

Genotype imputation is fundamental to association studies, and even gold standard panels like TOPMed have limitations to the populations and variants for which they yield good imputation.
To quantify the impact that varying the number of population-specific sequences in the reference panel has on imputation quality, we constructed 6 in-house reference panels from 2,504 1000G samples plus varying numbers of Samoan samples (4, 24, 48, 96, 384, and 1,285) from whole-genome sequencing and compared them to the 1000G Phase III and TOPMed imputation panels. Each reference panel was used to impute genotype data for 1,897 Samoan participants who were not part of any reference panel. We examined average imputation quality (r^2) and the number of well-imputed variants ($r^2 \geq 0.8$) on chromosomes 5 and 21 to assess performance and compared them to two gold-standard reference panels: TOPMed and 1000G Phase III. To further characterize variants that might gain the most in imputation accuracy, we also calculated LD scores split into low and high strata at the median value within MAF bins.
The 1000G + 1285 Samoan panel yielded > 200,000 more high-quality variants on chromosome 5 than the TOPMed panel, with 48,374 of these having a MAF $\geq$ 0.01. The largest gains were seen for lower-frequency variants with an up to 125% increase in well-imputed variants with MAF < 0.01 compared to the TOPMed imputation. Imputation quality increased as the number of Samoans represented in the panel increased. Panels with 48 or more Samoans included outperformed the TOPMed panel for all variants with MAF $\geq$ 0.001. The gains in imputation quality for the 1000G + 1285 Samoan reference panel compared to the TOPMed panel were greatest for low LD score variants. For rs200884524, a variant on chromosome 5 associated with dyslipidemia and enriched in Polynesians, the imputation quality was highest (r^2 = 0.89-0.95) for the reference panels that included Samoan haplotypes. Additionally, the imputed MAF from the reference panels with Samoans (0.207-0.222) was much closer to what is expected via targeted genotyping (0.202-0.233).
While not necessarily prescriptive for future studies, in this study we showed that as few as 48 population-specific participants added to 1000G yielded superior imputation quality to TOPMed. Our findings also demonstrated that panels containing Samoan-specific haplotypes improve the imputation of population-specific variants located in small LD blocks the most. These findings provide a framework to help future studies construct reference panels of their own to obtain high-quality imputation for genetic association studies.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4387 The genetic basis of covariance between blood and urine biomarkers

Authors:

C. Boetto[1], L. Henches[1], A. Auvergne[1], A. Frouin[1], H. Aschard[1,2]; [1]Inst. Pasteur, Paris, France, [2]Harvard TH Chan Sch. of Publ. Hlth., Boston, MA

Our comprehension of the genetic basis of individual human metabolites has improved significantly over the past years, providing critical insights into the biochemical pathways and potential functional implications of disease-associated variants. Conversely, the mechanism driving co-regulation and interaction between metabolites remains poorly understood. Metabolites often display high correlation and an increasing number of studies show that multivariate methods, including simple approaches such as biomarker ratios, can offer a promising path towards better disease prognosis. However, screening methods for predictors of correlation effect are lacking and real data application have therefore been very limited in scope. Here we conducted a genome-wide association screening (GWAS) for covariance effect between 28 urine and blood biomarkers in 150K unrelated participants from the UK Biobank, using MANOCCA (Multivariate Analysis of Conditional CovAriance), a novel, robust and powerful method we recently developed. Briefly, the concept of the MANOCCA approach to capture effects on the covariance consists in applying a multivariate test to the products of original variables. We identified 75 independent loci associated at genome-wide significance level ($P<10^{-8}$) with effects on covariance between the biomarkers. The strongest signal was observed with rs887829 ($P<10^{-300}$), within a region of nine UDP-glucuronosyltransferase genes. The variant, highly associated with the mean of albumin and alkaline phosphatase, displays a strong effect on the covariance between those two biomarkers and several other biomarkers. Although the proposed test is by construction orthogonal to tests based on the mean, most loci identified by the MANOCCA also display mean effect as measured by either univariate (standard linear regression) or multivariate (MANOVA) methods. Nevertheless, across the 75 signals, we observed a range of patterns, including cases where the covariance between a pair of biomarkers was associated with a variant not displaying any marginal mean effect on those biomarkers. Altogether, this study shows there exists a substantial genetic component of biomarker covariance, providing novel working hypotheses that can improve our understanding of the physiological processes and pathways involved in health and diseases.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4388 The genetically regulated transcriptome better predicts plasma protein levels than the observed transcriptome and reveals mechanisms underlying autoimmune diseases

Authors:

H. Wittich[1], S. Rich[2], J. Rotter[3], H. Im[4], A. Manichaikul[5], H. Wheeler[6], NHLBI TOPMed Consortium; [1]Loyola Univ. Chicago, Chicago, IL, [2]Univ. of Virginia, Scottsville, VA, [3]Lundquist Inst., Harbor-UCLA Med Ctr, Torrance, CA, [4]Univ. of Chicago, Chicago, IL, [5]Univ Virginia, Charlottesville, VA, [6]Loyola Univ Chicago, Chicago, IL

Genetic regulation of transcription and translation are important mechanisms through which noncoding genetic variants affect complex traits. While many studies have been successful at identifying cis-eQTL (within 1 Mb of the transcription start site), low effect sizes and a high multiple testing burden have made it challenging to detect trans-eQTL. Furthermore, eQTL studies do not consider translation and post-translational modifications that affect the function of proteins. Here, we leverage the TWAS method, PrediXcan, to test the cis component of gene expression for association with observed plasma protein levels to identify cis- and trans-acting genes that regulate protein levels. We performed a TWAS for 3,622 plasma proteins measured in 3,301 individuals with European ancestry from the INTERVAL study and replicated our results for 1,039 proteins measured in 971 individuals from the Trans-omics for Precision Medicine (TOPMed) Multi-Ethnic Study of Atherosclerosis (MESA). We used gene expression prediction models from 49 tissues trained with GTEx Project genotype and RNA-Seq data to predict the cis component of gene expression (GReX) for every gene in each tissue. We then tested the GReX of each gene for association with the observed abundance of each protein measured via SOMAscan assay. We found 1,210 trans-acting associations for 239 target proteins that replicated in TOPMed (FDR<0.05) with a median expected true positive rate (pi1) across tissues of 0.390 and 1,168 cis-acting associations for 218 target proteins that replicated in TOPMed (FDR<0.05) with a median pi1 of 0.806. Gene set enrichment analysis revealed that both the cis- and trans-acting targets were enriched for associations with blood protein levels and autoimmune diseases in the GWAS catalog. We found the highest pi1 for replicated associations between predicted RNA levels and protein levels of the same underlying gene (cis-same), with a median of 0.888 across tissues. This contrasts the findings of many studies that RNA and protein levels do not correlate. Using RNA-Seq data from the TOPMed MESA study, we found that the mean Pearson correlation of expression and protein levels for significant cis-same genes was higher in predicted tissues (R=0.17) than in observed (R=0.10, p=7.50e-11). We found higher correlations between predicted expression and protein levels in tissues closely involved in regulating circulating protein levels like liver and whole blood than in the brain tissues, perhaps due to the blood-brain barrier. These results indicate that predicted expression levels may be a better proxy for plasma protein levels because they exclude variation due to environmental factors.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4389 The identification of high-confidence disease-gene associations at the intersection of genetic and genomic evidence.

Authors:

C. Mordaunt, V. Kumar, A. Rupp, E. Laing; Computational Biology, Genomic Sci., GSK, Upper Providence, PA

Identifying high-confidence disease-gene associations from GWAS studies remains a difficult task. The application of orthogonal genomic datasets at scale in an indication-aware manner promises to improve gene prioritization. Here, we present an approach to integrate and score multiple lines of genomic evidence for genes at GWAS loci to identify high-confidence disease genes. We show that these genes are enriched for previous clinical success and apply this method to multiple sclerosis as an example disease.

Genes were evaluated across multiple types of genomic evidence. Expression specificity was calculated by combining disease-tissue and -cell-type associations from text-mining with mRNA expression levels from bulk and single-cell RNA-seq and protein expression from mass spectrometry. Human disease and mouse model phenotype data were compiled and phenotypic similarity was assessed. Gene characteristics such as protein class, mouse orthology, and loss-of-function essentiality were also included. Each piece of evidence was benchmarked to estimate the enrichment of previous clinical success for high scoring genes, and then an overall score was calculated as a weighted sum of the benchmarking ratios relative to the maximum possible score.

An integrated table was prepared compiling genomic scores across 20,226 genes and 6,863 diseases (137 million gene-disease pairs). Through benchmarking, each evidence type was found to identify genes with significantly higher odds of clinical success. After summarizing genomic evidence into overall scores, 6 million high-scoring gene-disease pairs were identified, with benchmarking indicating a 2.1 fold enrichment for clinical success (p = 5.5E-91). This approach was applied to GWAS for multiple sclerosis by intersecting genes near GWAS-associated loci with high-scoring genes based on genomic evidence and appending multiple sclerosis differential gene expression results. The intersection of these criteria resulted in the identification of 194 genes for further investigation, including genes with a well-known immune function.

High-confidence gene-disease pairs were identified by integrating multiple genomic data types and overlapping with genetic evidence. The framework used to score and benchmark these data types could be expanded to other types of genomic data. The combination of genetic and other genomic evidence matching a gene to a disease will likely result in the discovery of more robust disease-associated genes.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4390 The local portability of polygenic scores across populations varies significantly along the genome.

Authors:

A. Lupi, A. Vazquez, G. de los Campos; Michigan State Univ., East Lansing, MI

In the last fifteen years, thousands of Genome-Wide Association Studies (GWAS) have been published. These findings are increasingly used to develop polygenic risk scores (PGS). However, the overwhelming majority of participants in GWAS have European (EU) ancestry, and PGS derived from EUs have poor predictive performance in other populations. Many studies have evaluated the portability of whole-genome PGS across populations. However, no previous research has quantified how much the portability may vary along the genome. We hypothesize that the portability of SNP effects between populations varies significantly over the genome because of varying levels of genome differentiation. We developed a Monte Carlo-ANOVA (MC-ANOVA) method to estimate the portability of local PGS between populations. The method quantifies the impact of genome differentiation (allele frequency differences and imperfect linkage disequilibrium) on the portability of local PGS. We applied the methodology to data from the UK-Biobank to develop maps of the predicted portability of EU-derived marker effects in African, Caribbean, East Asian, and South Asian populations. Averaged over the entire genome, we estimate that genome differentiation alone leads to relative accuracies (RA, the ratio of PGS R-squared in non-EU relative to within EU prediction) ranging from 0.268 (African) to 0.771 (South Asian). However, we report a sizable variability in RA's along the genome, suggesting that even in populations with low overall portability of EU-derived SNP effects (*e.g.*, African) there are still chromosome segments with high portability of EU-derived PGS. We validated our findings using real PGS for six complex traits (height, serum urate, LDL and HDL cholesterols, BMI, and glucose). Our results show that EU-derived SNP effects from SNPs in regions where MC-ANOVA predicts RA have relatively high empirical RA of real PGS. We provide R-software that implements the proposed method and maps of the predicted RA (for 10 Kbp chromosome segments) for individuals of African, Caribbean, South Asian, and East Asian descent. The information in these maps can be used to prioritize variants to develop more accurate PGS for cross-population prediction.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4391 The pitfalls of overly parameterized polygenic risk scores and new diagnostics to combat overfitting

Authors:

A. Aw[1], J. McRae[2], E. Rahmani[3], Y. S. Song[1]; [1]UC Berkeley, Berkeley, CA, [2]Illumina, Inc., Foster City, CA, [3]UCLA, Los Angeles, CA

Polygenic risk scores (PRSs) are increasingly advocated as useful genetic instruments in clinical settings. Previous work has shown that predictability of PRSs is often improved by including thousands or even tens of thousands of variants, much more than the number of genome-wide significant variants based on GWAS.
Here, we challenge a commonly-held view that insignificant variants contributing to PRS predictability are causal yet insignificant at a genome-wide level owing to small effects. We argue, using statistical principles and empirical validation on the UK Biobank (UKB), that high-dimensional PRSs are in fact overparameterized, with the overparameterization capturing population stratification. We first show mathematically that (1) PRSs trained on a cohort can be represented as a linear combination of principal components (PCs), and (2) PRSs with random effect sizes (as one would get for insignificant non-causal variants; “random projection”) tend to capture leading axes of variation in the data. These facts imply concrete desirable checks of stratification-driven overfitting, including the evaluation of PRS predictability inflation and sensitivity to effect perturbations.
Applying our checks to 500,000 participants from the UKB across 141 quantitative traits, we find that 89% of all phenotypes have inflated predictability (incremental R^2) from PRSs constructed by random projections. Crucially, the degree of inflated predictability is significantly driven by the degree of phenotypic stratification along PCs.
Applying sensitivity analyses to PRSs constructed for 103 traits, we find that the performances (percentile-prevalence rank correlations) of more than 70% of PRSs are insensitive to shuffling or sign flipping of effects of weakly significant target variants. This suggests that the effects of insignificant variants typically do not correspond to true causal signals and therefore many PRSs are likely overparameterized. Further, the degree of insensitivity is not driven by homogeneity of the effects of these target variants; rather it is significantly driven by the strength of PRS stratification along PCs of the training cohort, the sum of effects of target variants, and the number of target variants. Focusing on 8 Lipoprotein(a) PRSs trained using various PRS methodology (source: PGS Catalog), we find that no PRS is uniformly sensitive, across 11 evaluation metrics considered, to effect shuffles or sign flips. Different PRSs also emerge as most sensitive depending on the metric of choice.
In summary, including insignificant GWAS variants often leads to overfitted PRSs, which can be avoided through inflation and sensitivity checks.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4392 The relationship between major psychiatric disorders, substance use behaviors, and longevity: A multivariable Mendelian randomization and multi-omics study

Authors:

A. Hamandi[1], D. Rosoff[2,1], A. Bell[1], L. Mavromatis[1], L. Park[1], J. Jung[1], J. Wagner[1], F. Lohoff[1]; [1]NIH, Bethesda, MD, [2]Univ. of Oxford, Oxford, United Kingdom

Observational studies suggest major psychiatric disorders, including bipolar disorder (BD), schizophrenia (SCZ) and major depressive disorder (MDD) reduce longevity; however, frequently co-occurring alcohol consumption and smoking have also been shown to impact longevity, making it difficult to disentangle which among the co-morbid disorders is driving this reduction. Here we deploy Mendelian randomization (MR) and transcriptomic imputation to parse the associations these disorders have to longevity. This study sourced existing summary-level genome-wide association study (GWAS) data from European-ancestry cohorts on psychiatric disorders, substance use behaviors, cardiometabolic diseases, multivariate longevity, and epigenetic aging acceleration (EAA).

MR analysis shows MDD, weekly alcohol consumption, and smoking negatively associate with longevity in a single-variable model; while only smoking negatively associates in a multivariable one. MR also shows a corresponding positive association between smoking and PhenoAge-measured EAA, and that these smoking associations are cardiometabolic disease-independent.

Transcriptome-wide association studies on smoking across several tissue types identified 118 novel gene-smoking associations not captured by the original GWAS. Colocalization analysis revealed several molecular features which share a causal variant with both lifetime smoking index placement and longevity, including PRMT6 transcripts and those for five genes in locus 17q21.31. Broadly, novel and colocalized genes were involved in cellular proliferation, learning, and neurodevelopment. Drug-target MR analysis identified genes upstream of smoking behavior whose protein products are amenable to drug targeting and involved in reward processing and nicotine metabolism.

Using genomic methods that enable the simultaneous assessment of major psychiatric disorders and substance use behaviors, we provide evidence that smoking—but not drinking, MDD, BD, nor SCZ—has a negative independent association with longevity that is recapitulated in the transcriptome. Our findings highlight the importance of developing smoking cessation therapies for the longevity of psychiatric populations, and identify prime drug-target candidates for such future translational research.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4393 The SNP rs6859 in NECTIN2 gene is associated with underlying heterogenous trajectories of cognitive changes in older adults.

Authors:

A. Lathika Rajendrakumar, K. G. Arbeev, O. Bagley, A. I. Yashin, S. Ukraintseva; Duke Univ., Durham, NC

Background: Functional decline associated with dementia, including in Alzheimer's disease (AD), is not uniform across individuals, and respective heterogeneity is not yet fully explained. Such heterogeneity may in part be related to genetic variability among individuals. In this study, we investigated whether the SNP rs6859 in *NECTIN2* gene (a major risk factor for AD) influences trajectories of cognitive decline in older participants from the Alzheimer's Disease Neuroimaging Initiative (ADNI). **Method:** We retrospectively analyzed records on 1,312 participants from ADNIMERGE subset of the ADNI database. We used longitudinal measures of Mini-Mental State Examination (MMSE) scores in participants, who were cognitively normal, or having AD, or other cognitive deficits. Multiple linear regression analysis was conducted to investigate the association of SNP rs6859 with normalized lowest MMSE score recorded per participant. For trajectory analysis, the optimal number of latent classes was determined by Bayesian Information Criterion (BIC). We fitted a multivariate latent class regression model with SNP rs6859 both as fixed effects and as a mixture term with age, sex, education and ethnicity, as covariates for the optimal class model. **Results:** The SNP rs6859 was independently associated with MMSE in multiple linear regression analysis (-2.009, 95% CI: -3.596, -0.422, $p<0.05$). Three distinct subgroups performed best relative to the other latent groups during model selection (BIC: 56271.20). There was clear difference of MMSE trajectories across latent classes which was particularly evident for Class 1 (highest MMSE group with sharp decline). In the posterior classification, 54.57% (n=716), 43.75% (n=574), 1.68% (n=22) were classified as class 1 (highest MMSE group), class 2 (intermediate MMSE) and class 3 (low MMSE). In the heterogenous linear mixed model, the rs6859 - A risk allele was significantly associated with MMSE class membership and related decline; Class 1 (-4.633, 95% CI: -6.556, -2.711, $p<0.001$), Class 2 (2.573, 95% CI:0.121,5.026, $p<0.05$) and Class 3 (-3.974, 95% CI: -5.663, -2.286, $p<0.001$) respectively. **Conclusion:** This study found statistical evidence supporting the classification of three latent subclass groups representing complex MMSE trajectories in the ADNI cohort. The SNP rs6859 was associated with decreasing MMSE. It can, therefore, be suggested as a candidate genetic predictor of variations in MMSE trajectories between individuals. Functional studies may help further elucidate this relationship.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4394 The use of deep learning to alter and analyze facial expressions in syndromic facial diagnostics

Authors:

T. Patel[1], A. Othman[1], O. Sumer[2], F. Hellman[2], P. Krawitz[3], A. Elisabeth[2], P. Hu[4], C. Ngongang[1], L. Suzanna[1], R. Waikel[1], D. Duong[1], B. Solomon[1]; [1]NIH, Bethesda, MD, [2]Inst. of Computer Sci., Augsburg Univ., Augsburg, Germany, [3]Inst. for Genomic Statistics and Bioinformatics, Univ. of Bonn, Bonn, Germany, [4]NHGRI (NIH), Bethesda, MD

Deep learning is increasingly being used to examine genetic diseases. Aspects of the underlying datasets can impact results, therefore confounders should be studied and mitigated where possible. Stereotypically, Williams (WS) and Angelman (AS) syndromes are associated with a "happy demeanor," which includes a smiling face. Thus, clinical geneticists may be more likely to identify these conditions in images of smiling individuals.In our previous study we explored, aging as a confounder in syndromic facial diagnostics, (PMID: 35480315) and additionally found that clinical geneticists could more accurately classify WS in smiling individuals. To more fully study the impact of facial expression in facial diagnostics, we curated publicly available facial images of over 3500 individuals with genetic syndromes. We utilized ResNet, facial expression recognition categorization to determine whether or not that individual was classified as smiling. For WS and AS, images with neutral expressions had significantly lower prediction probabilities for the correct syndrome designation than those with smiling expressions. This was not seen for images of individuals with two conditions selected for comparison, 22q11.2 deletion and Noonan syndromes. To further explore this impact, we altered the facial expressions to compare the implications of a person smiling or not smiling. We trained Hyperstyle, the GAN-inversion technique to determine the vector representations of our images that are compatible with StyleGAN2. Then, using InterfaceGAN, we edited these vectors to make them capable of recreating the identical original images but with a new facial expression. Using surveys, eye-tracking, and DL methods, we examined how altering facial expressions affects predictions by human experts as well as neural net classifiers. Our findings show that altering the facial expression in pictures of people with WS and AS syndromes from smiling to neutral affects syndrome classification.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4395 The utility of polygenic risks scores in predicting age of onset of End Stage Renal Disease

Authors:

K. Collins[1,2,3], E. Gilbert[1,2], V. Mauduit[4], K. Benson[1,2], E. Elhassan[5,1], C. Hill[6], A. McKnight[7], A. Maxwell[6], P. van der Most[8], M. De Borst[8], W. Guan[9], P. Jacobson[9], A. Israni[10], B. Keating[11], G. Lord[12], C. O'Seaghda[5], S. Markkinen[13], I. Helanterä[14], K. Hyvärinen[13], J. Partanen[15], S. Madden[1], S. Limou[4], G. Cavalleri[1,2], P. Conlon[5,1]; [1]Royal Coll. of Surgeons in Ireland, Dublin, Ireland, [2]The Sci. Fndn. Ireland FutureNeuro Ctr. of Excellence, Dublin, Ireland, [3]SFI Ctr. for Res. Training in Genomics Data Sci., Galway, Ireland, [4]Nantes Univ., Nantes, France, [5]Beaumont Hosp., Dublin, Ireland, [6]Queen's Univ. Belfast, Belfast, United Kingdom, [7]Queen s Univ. of Belfast, Belfast, United Kingdom, [8]Univ. of Groningen, Groningen, Netherlands, [9]Univ Minnesota, Minneapolis, MN, [10]Hennepin County Med Ctr, Minneapolis, MN, [11]UPenn, Philadelphia, PA, [12]Univ. of Manchester, Manchester, United Kingdom, [13]Finnish Red Cross Blood Service, Helsinki, Finland, [14]Helsinki Univ. Hosp., Helsinki, Finland, [15]Finnish Red Cross Blood Serv, Helsinki, Finland

Background

End Stage Renal Disease (ESRD) is a condition where an individual’s kidneys stop functioning on a permanent basis, thus requiring either transplant or dialysis. ESRD affects 800,00 individuals in the US alone, with 71% on dialysis and 29% with a kidney transplant. Multiple mechanisms have been implicated in the aetiology of kidney disease including monogenic and acquired conditions as well as polygenic factors. In cases of monogenic kidney disease, if the pathogenic mutation is known, some of the variation in age of onset of ESRD can be explained by pathogenic variant type, such as in the case of Polycystic Kidney Disease. However, most of the variation in age of onset is still largely unexplained. Here we test the hypothesis that polygenic risk scores (PRS) for a range of renal-related traits can predict age of onset of ESRD.

Methods

Using a dataset of 10,121 patients with ESRD from 9 European ancestry cohorts, PRS were calculated based on large GWAS for the following traits: hypertension, albuminuria, kidney volume, eGFR, and decline in eGFR. First, we investigated whether there was a difference in polygenic burden for each trait between individuals with ESRD and healthy controls. Next, we investigated the impact of PRS on age of onset of ESRD using a Cox PH mixed effects model. Finally, we compared the median survival between those in the highest and lowest tertiles of polygenic risk for each trait.

Results

Individuals with ESRD were found to have an elevated burden for albuminuria and for decreased eGFR in comparison to healthy controls (p-values 0.03, 0.04 respectively). Polygenic burden for decreased eGFR was also associated with earlier onset of ESRD (HR: 1.03 P = 0.02). Those in the lowest tertile of risk for reduced eGFR have a median survival of 49.1 years, compared to 48.3 years for those in the highest tertile (p=0.009).

Conclusions

These findings suggest that polygenic burden for albuminuria and reduced eGFR could explain some of the variation in age of onset of ESRD. Further investigation is required to combine this common variant burden with pathogenic variants and rare variant burden in order to build a more complete model to predict age of onset of ESRD.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4396 † Theoretical and empirical interpretation of heritability estimation in admixed populations.

Authors:

J. Huang[1], M. Shriver[2], A. Zaidi[3]; [1]Penn State Univ., University Park, PA, [2]Penn State Univ, University Park, PA, [3]Univ. of Minnesota, Minneapolis, MN

Estimating heritability in admixed populations is challenging due to their complex genetic diversity and population structure. Heritability is commonly estimated using genome-wide restricted maximum likelihood (GREML), which models the phenotypic resemblance between individuals as a function of genetic similarity. Despite the simple intuition, the accuracy and interpretation of GREML estimates remain a matter of debate, particularly in the presence of population structure and assortative mating. While much of the discussion has focused on the inflation of GREML estimates due to confounding effects of genetic and environmental stratification, there is comparatively little appreciation of the fact that population structure not only confounds but also contributes to heritability. This lack of clarity limits our understanding of the applicability of statistical genetics methods to diverse cohorts.
Here, we evaluated the behavior of GREML estimation in recently admixed populations using quantitative genetic theory and simulations. We showed that GREML estimates of heritability are biased in admixed populations, even in the absence of confounding. We traced the source of this bias to two assumptions: (i) that the effect sizes of causal variants are independent, and (ii) that the population mates randomly. These sources of bias can either inflate or deflate heritability estimates depending on the trait architecture and the degree of admixture structure in the population.
We sought to empirically understand the extent of this bias in admixed Americans using genetic data from the African Americans in the South West from the 1000 Genomes Project. We estimated the bias in the genetic variance explained by GWAS SNPs for 26 traits, including anthropometric traits, blood pressure, lipid levels, and blood cell counts using summary statistics from the GWAS catalog. We found that, despite incorrect assumptions, GREML estimates of heritability should, in practice, be close to the true parameter values, at least for the traits studied. However, for skin pigmentation and cholesterol (HDL and LDL), GREML should underestimate heritability in admixed populations because it does not take into account the fact that the effects of causal variants are correlated. As such, GREML estimates of heritability in admixed cohorts should be interpreted with caution.
We provide a clear interpretation of the GREML estimator of heritability in admixed populations and discuss its implications for genome-wide association studies and polygenic risk prediction in diverse cohorts.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4397 Three open questions regarding polygenic score portability

Authors:

J. Wang[1], M. Zietz[2], J. Mares[3], P. J. Rathouz[4], V. M. Narasimhan[1], M. F. Przeworski[3], A. Harpak[1]; [1]The Univ. of Texas at Austin, Austin, TX, [2]Columbia Univ. Irving Med. Ctr., New York City, NY, [3]Columbia Univ., New York City, NY, [4]Dell Med. Sch., Austin, TX

A major obstacle hindering the broad adoption of polygenic scores (PGS) is their lack of "portability" to people that differ—in genetic ancestry, environmental exposures or other characteristics—from the GWAS samples in which genetic effects were estimated. First principle considerations suggest some of the portability problem arises from differences in genetic ancestry; how much, however, remains unknown.

Here, we use the UK Biobank (UKB) to measure the change in individual-level PGS prediction accuracy as a continuous function of genetic dissimilarity to the GWAS sample from the White British subset of UKB. Through specific case studies, we highlight three major gaps in our understanding of PGS portability.

First, while previous studies have shown an average trend of decay in prediction accuracy in genetic ancestry groups other than that of the GWAS sample, such groupings confound differences in genetic similarity to the GWAS sample with many other factors, including distinct environmental and social exposures. We show that for many cases, continuous variation in prediction accuracy can be explained comparably well by socio-economic measures and by genetic ancestry.

Second, we show that trends of portability can be highly distinct across traits. While portability decays linearly with genetic dissimilarity for some traits such as LDL levels, this is not true across all traits, in contrast with previous reports. For BMI, portability is not monotonically decreasing with genetic dissimilarity. For several immune-related traits, prediction accuracy drops near zero quickly even at intermediate levels of genetic dissimilarity to the GWAS sample. We show that this quick drop is largely due to immune associations in the White British subset of the UKB being poorly reflective of genetic effects away from the GWAS sample.

Third, we show that even qualitative trends of portability can depend on the measure of prediction accuracy used. For some traits such as Hemoglobin levels (MCH), prediction accuracy at the group level (e.g., R^2) decays—while some measures of individual level prediction accuracy (e.g., reduction in mean squared error) increase with genetic dissimilarity.

Together, these results highlight the importance of considering characteristics specific to the trait, the GWAS sample and prediction samples, as well as the most relevant metric of prediction accuracy for the application of a given polygenic score, in order to understand and tackle the urgent problem of portability.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4398 Thyroid function and metabolic syndrome: a two-sample bidirectional Mendelian randomization study.

Authors:

N. Pleic, T. Zemunik, I. Gunjača, M. Babić Leko; Univ. of Split, Sch. of Med., Split, Croatia

Background: In clinical practice, metabolic syndrome (MetS) is often observed alongside thyroid dysfunction. MetS is defined as a cluster of cardiometabolic abnormalities including obesity, hypertension, hyperglycaemia and dyslipidemia. Multiple studies have suggested an association between thyroid dysfunction and MetS or its components, but the causality and direction of these associations remain yet to be proven. The aim of this study was to examine the causal effects of thyroid function on metabolic syndrome risk and its components, and vice versa, using a mendelian randomization (MR) approach. **Methods:** We performed a two-sample bidirectional MR study using summary statistics from the most comprehensive genome-wide association studies (GWAS) of thyroid-stimulating hormone (TSH, n=119,715), free thyroxine (fT4, n=49,269), MetS (n=291,107), as well as components of MetS: waist circumference (n=462,166), fasting blood glucose (n=281,416), hypertension (n=463,010), triglycerides (TG, n=441,016) and high-density lipoprotein cholesterol (HDL-C, n=403,943). We chose the multiplicative random-effects inverse variance weighted (IVW) method as the main analysis. Sensitivity analysis included weighted median and mode analysis, as well as MR-Egger, MR-PRESSO and Causal Analysis Using Summary Effect estimates (CAUSE). **Results:** Our results suggest that higher fT4 levels lower the risk of developing MetS (OR=0.96, P=0.037). Genetically predicted fT4 was also positively associated with HDL-C (β=0.02, P=0.008), while genetically predicted TSH was positively associated with TG (β=0.01, P=0.044). These effects were consistent across different MR analyses and confirmed with the CAUSE analysis. In the reverse direction MR analysis, genetically predicted HDL-C was negatively associated with TSH (β=-0.03, P=0.046) in the main IVW analysis. **Conclusions:** Our study suggests that variations in normal-range thyroid function are causally associated with the diagnosis of MetS and with lipid profile, while in the reverse direction, HDL-C has a plausible causal effect on reference-range TSH levels.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4399 Title:Deconvoluting Bulk RNA-seq Data of 1200 samples of Dorsolateral Prefrontal Cortex tissue using reference single nucleus RNAseq data of 400 samples for Cell-Type-Specific TWAS of AD Dementia

Authors:

S. Tang[1], A. S. Buchman[2], P. L. Jager[3], D. Bennett[2], S. Tasaki[2], J. Hu[1], J. Yang[1]; [1]Emory Univ., Atlanta, GA, [2]Rush Univ. Med. Ctr., Chicago, IL, [3]Columbia Univ., New York, NY

Recent transcriptome-wide association studies (TWAS) of Alzheimer's disease (AD) dementia have identified dozens of risk genes by integrating transcriptomic data with GWAS summary data. The potential genetic effects of the identified risk genes on AD dementia can be examined as mediated via transcriptome. Although cell-type-specific TWAS has been recently done for AD dementia using single nucleus RNAseq (snRNA-seq) data (n=400) of dorsolateral prefrontal cortex (DLPFC) tissue, most TWAS studies use transcriptomic data generated by bulk RNA-seq.

This study will compare three methods (TCA, bMIND, and BayesPrism) to identify the best one to deconvolute bulk RNA-seq data. Specifically, TCA utilizes a Tensor Composition Analysis (TCA) approach; bMIND is based on a Bayesian multivariable regression model with Gaussian priors; and BayesPrism assumes a Bayesian multinomial model. Both Bayesian methods use priors derived from reference snRNA-seq data. By using snRNA-seq data of 400 DLPFC samples which also have bulk RNA-seq data profiled, we will randomly select snRNA-seq data of 200 samples as reference data to deconvolute bulk RNA-seq data, and use the remain for testing the deconvolution accuracy with respect to cell-type-specific Pearson's correlation and mean squared errors.

We will then apply the best deconvolution method to a large-scale bulk RNA-Seq dataset of DLPFC tissue (n=1200), with reference snRNA-seq data of 400 DLPFC samples. The obtained sample-level cell-type-specific gene expression levels and their corresponding whole genome sequencing data (n=1200) will be used to train cell-type-specific gene expression imputation models, using the TIGAR tool developed by our lab (Nagpal et. al, AJHG, 2019). These trained expression imputation models will then be used with the most recent GWAS summary data (n=762K) of AD dementia to conduct cell-type-specific TWAS. We will use both penalized regression methods with Elastic-Net penalty and the nonparametric Bayesian Dirichlet process regression (both implemented in TIGAR) to train expression imputation models. For each test gene, the TWAS p-values based on expression imputation models trained using both regression methods will be combined by the aggregated Cauchy association method to produce a final omnibus TWAS p-value.

To summarize, by deconvoluting the bulk RNA-Seq data with available reference scRNAseq data, we expect to increase the sample size for training cell-type-specific expression imputation models to increase the TWAS power for studying AD dementia.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4400 TIVAN-indel a computational framework for annotating and predicting non-coding regulatory small insertions and deletions

Authors:

L. Chen[1], A. Agarwal[2], F. Zhao[1], Y. Jiang[3]; [1]Univ. of Florida, Gainesville, FL, [2]Indiana Univ., Bloomington, IN, [3]Univ North Carolina at Chapel Hill, Chapel Hill, NC

Small insertion and deletion (sindel) of human genome has an important implication for human disease. One important mechanism for non-coding sindel (nc-sindel) to have an impact on human diseases and phenotypes is through the regulation of gene expression. Nevertheless, current sequencing experiments may lack statistical power and resolution to pinpoint the functional sindel due to lower minor allele frequency or small effect size. As an alternative strategy, a supervised machine learning method can identify the otherwise masked functional sindels by predicting their regulatory potential directly. However, computational methods for annotating and predicting the regulatory sindels, especially in the non-coding regions, are underdeveloped. By leveraging labeled nc-sindels identified by cis-expression quantitative trait loci analyses across 44 tissues in Genotype-Tissue Expression (GTEx), and a compilation of both generic functional annotations and large-scale epigenomic profiles, we develop TIssue-specific Variant Annotation for Non-coding indel (TIVAN-indel), which is a supervised computational framework for predicting non-coding regulatory sindels. As a result, we demonstrate that TIVAN-indel achieves the best prediction performance in both with-tissue prediction and cross-tissue prediction. As an independent evaluation, we train TIVAN-indel from the 'Whole Blood' tissue in GTEx and test the model using 15 immune cell types from an independent study named Database of Immune Cell Expression. Lastly, we perform an enrichment analysis for both true and predicted sindels in key regulatory regions such as chromatin interactions, open chromatin regions and histone modification sites, and find biologically meaningful enrichment patterns.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4401 Towards fully automated polygenic risk score estimation from noisy data using a machine learning approach

Authors:

S. Squires[1], D. P. Fraser[1], S. Sharp[2], M. Weedon[1], R. Oram[1]; [1]Univ. of Exeter, Exeter, United Kingdom, [2]Stanford Univ., Palo Alto, CA

Introduction: Polygenic risk scores (PRSs) enable the summary of genetic data into a single number for research and clinical uses. A PRS often combines the number of effect alleles of single nucleotide polymorphisms (SNPs) in a weighted sum, but more complicated functions are also used. An impediment to the further use of PRSs is the difficulty in curating the data because if any of the SNPs are missing it can make the PRS less accurate, difficult to compare between datasets or unusable. Various methods including quality control, imputation and the use of proxy SNPs can be utilised to mitigate these issues, but they require technical expertise and may still produce poor results. As PRSs increasingly use more complicated functions these problems are likely to become more acute. Our aim is to produce a method that can produce an estimate of the PRS without substantial data curation even if the data has poor quality or missing SNPs.
Methods: We design and demonstrate a novel machine learning methodology (M-MLP) to estimate PRSs. We train a deep learning model with inputs that include proxy SNPs in addition to the PRS SNPs. We add in a masking and corruption procedure into the training to simulate poor quality input SNPs forcing the model to learn a robust representation which can handle noise in its inputs. We primarily use a 67-SNP PRS for type-1 diabetes (T1D) which includes pairwise interaction terms creating a complex nonlinear function between input SNPs and output PRS. We use UK Biobank (UKBB) imputed genetic data for training and testing our models, including keeping 387 T1D cases in our test set.
Results: We test our model and the PRS by simulating corruption of the input SNPs. By corrupting the input SNPs with uniform probability ranging from 0 to 0.25 the PRS sees its T1D cases to population AUC fall from 0.92 (95% CI 0.90-0.93) to 0.71 (CI 0.68-0.73) while the M-MLP sees a smaller fall from 0.92 (CI 0.90-0.93) to 0.89 (CI 0.88-0.91). We also altered specific numbers of non-proxy SNPs leaving the rest uncorrupted. With 12 of 67 SNPs corrupted (at a probability of 0.2) the PRS and M-MLP produces AUCs of 0.89 (CI 0.88-0.91) and 0.92 (CI 0.90-0.93), respectively. With 61 SNPs corrupted the PRS and M-MLP AUCs are 0.77 (CI 0.75-0.79) and 0.91 (CI 0.90-0.93), respectively. We further demonstrate the method using two other T1D PRSs and a Celiac PRS.
Conclusions: Our M-MLP method produces equivalent performance on uncorrupted data and improved performance on simulated corrupted data compared to the PRS. As the M-MLP model is resilient to corrupted or missing data it has the potential to be used with minimal data curation allowing wider use of PRSs and more confidence in the outputs.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4402 Towards improving access to autism clinical genetic testing: Estimating personalized genetic probabilities in over 3,000 autistic children

Authors:

S. Kuo[1,2], C. van der Merwe[1,2], J. M. Fu[1,2], M. E. Talkowski[2,1], A. Thurm[3], S. J. Sanders[4,5], S. L. Bishop[4], E. B. Robinson[1,2,6]; [1]Broad Inst. of MIT and Harvard, Cambridge, MA, [2]Massachusetts Gen. Hosp., Boston, MA, [3]Natl. Inst. of Mental Hlth., Bethesda, MD, [4]Univ. of California San Francisco, San Francisco, CA, [5]Oxford Univ., Oxford, United Kingdom, [6]Harvard Sch. of Publ. Hlth., Boston, MA

Introduction: With 373 genes now associated with neurodevelopmental disorders (NDDs) like autism and intellectual disability, clinical genetic testing yields among autistic children will continue to rise. Despite the promise of clinical genetics for autism and the universal autism genetic testing guidelines from professional medical associations, clinical genetic testing remains vastly underutilized in autism. To help clinicians and families consider clinical genetic testing for autistic children, we aimed to develop an algorithm for predicting pretest probabilities of carrying a NDD-associated variant (NDAV).
Methods: Genetic probability prediction algorithms were developed in a large autism sample with whole exome sequencing and harmonized deep phenotyping data (N=3440 children aged 4-17 years from two independent cohorts). NDAVs were defined as de novo protein truncating variants, copy number deletions, missense variants (MPC>2), or genomic disorders in any NDD-associated gene. We first built logistic regression models predicting NDAV carrier status from LASSO-selected phenotypes, including demographic, clinical, developmental, and cognitive information. We stratified genetic probability predictions by >60,000 possible combinations of phenotypes, then compared each of these stratified probabilities to the average genetic probability in the general population.
Results: Genetic probabilities for carrying a NDAV were highly variable across autistic children, ranging from ~1% in children without intellectual disability or congenital heart problems who began walking by 12 months, to ~55% in children with co-occurring intellectual disability and congenital heart problems who began walking after 18 months. Because NDAVs overlap with clinically returnable variants and ~90% of the sample had estimated genetic probabilities greater than the average genetic probability in the general population, these data underscore the importance of recommending clinical genetic testing for autistic children. Our algorithm provides personalized genetic estimates with minimal burden and can easily be completed in <5 minutes.
Conclusions: The probability that an autistic child carries a potentially returnable genetic variant can be estimated from readily accessible phenotypes and used by clinicians and families in making decisions about clinical genetic testing. By working with diverse stakeholders in the autism community to translate our prediction algorithm into a public web resource, we hope to reduce barriers to accessing clinical genetic testing and to understanding autism's genetic influences and heterogeneity.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4403 Towards precision medicine: Factor analysis-driven phenome-wide data integration for enhanced polygenic score performance

Authors:

Y. Wang[1], Y. He[1], C. Fadil[2], K. Tsuo[3], S. Koyama[3], K. Paruchuri[1], Z. Yu[3], H. E. Leland[1], Z. Zheng[3], B. Truong[1], C. E. Carey[3], R. K. Walters[1], B. N. Neale[1], E. B. Robinson[1], P. Natarajan[1], A. R. Martin[1]; [1]Massachusetts Gen. Hosp., Boston, MA, [2]Harvard Univ., Boston, MA, [3]Broad Inst. of MIT and Harvard, Cambridge, MA

Polygenic scores (PGS) hold promise for stratifying patient disease risk, but they currently face critical limitations that preclude maximal accuracy and generalizability. Most PGS are developed from single-trait and single-ancestry genome-wide association studies (GWAS), missing opportunities to harness the phenome-wide breadth of modern multi-ancestry biobanks. We capitalized on factor analysis, a principled multidimensional distillation approach, to comprehensively model the interconnected landscape of human health.

Factor analysis unveiled 35 latent factors from 2,772 phenotypes in 361,144 White British individuals from the UK Biobank (UKB). Using GWAS on each of the 35 factors, we developed PGS, conducted PGS-pheWAS, and assessed predictive performance stratified by ancestry in three biobanks: UKB, Mass General Brigham Biobank (MGBB), and All of Us (AoU). Our analysis encompassed 52,946 African, 39,015 American, 10,809 Central and South Asian, 8,691 East-Asian, 158,413 European and 2,214 Middle Eastern individuals.

We identified a total of 142, 1,031, and 4,699 Bonferroni-significant associations out of 43,120, 115,235, and 190,785 comparisons in the UKB holdout, MGBB, and AoU, respectively. Notably, factors related to self-reported health and medical history showed the highest number of associations (2,203), followed by socio-demographics (1,192) and physical measurements (1,062). Importantly, directions of significant associations remained consistent across ancestries and biobanks, regardless of differences in recruitment strategies and geographic regions, highlighting the generalizability of factor-based PGS. Moreover, factor-based PGS demonstrated equal or superior accuracy and portability compared to PGS derived from several-fold larger case/control studies. For example, PGS based on a respiratory function factor (N = 360,406), specifically associated with diseases of the respiratory system among various phecode categories, outperformed an asthma PGS derived from multi-ancestry GWAS ($N_{Effective}$ = 715,234), with improvements of 19% in relative accuracy for African individuals.

Our findings highlight the power of modeling phenome-wide data from deeply phenotyped biobanks for accurate and generalizable prediction of patient outcomes in datasets lacking comparable depth or harmonization. We are integrating factor-based PGS with social determinants of health to develop comprehensive risk prediction models. These models are intended to further enhance performance, elucidate the most crucial factors underlying health outcomes, and identify key targets to improve health and reduce disparities.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4404 TraitScan: a trait subset scanning method for multi-trait GWAS

Authors:

R. Cao[1], E. Olawsky[1], E. McFowland, III[2], E. Marcotte[1], L. Spector[1], T. Yang[1]; [1]Univ. of Minnesota, Minneapolis, MN, [2]Harvard Business Sch., Boston, MA

Multi-trait analysis has been shown to have greater statistical power than single-trait analysis. Most of the existing multi-trait analysis methods only work with a limited number of traits and usually target achieving high statistical power instead of identifying relevant traits. We developed TraitScan, a powerful, summary-statistic-based, and fast algorithm that selects and tests a subset of traits from a moderate or large number of traits. We evaluated TraitScan using extensive simulations and found that it outperformed existing methods in terms of both testing power and trait selection under certain sparsity conditions. To demonstrate the utility of TraitScan, we applied it to search for traits associated with Ewing Sarcoma, a rare bone tumor in children, among 706 traits in UK Biobank. Our analysis revealed a few promising traits worthy of further investigation, highlighting the potential of TraitScan in enabling more effective multi-trait analysis as the deeply-phenotyped genome-wide association study databases emerge. Our algorithm is implemented in an R package ‘TraitScan’ available online.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4405 † Transcriptome-wide association study of whole blood lipid traits in a diverse population.

Authors:

J. Lundin[1], U. Peters[1], Y. Hu[1], J. Haessler[1], Y. Cai[1], K. North[2], L. Raffield[2], M. Graff[2], K. Young[2], S. Buyske[3], A. Reiner[4], T. Ye[4], R. A. Smit[5], Z. Wang[5], D. Levy[6], R. Joehanes[6], N. Heard-Costa[7], G. Peloso[8], Y. Wang[8], C. Kooperberg[1], PAGE Study; [1]Fred Hutchinson Cancer Ctr., Seattle, WA, [2]Univ. of North Carolina, Chapel Hill, NC, [3]Rutgers Univ, Piscataway, NJ, [4]Univ. of Washington, Seattle, WA, [5]Icahn Sch. of Med. at Mount Sinai, New York, NY, [6]Population Sci. Branch, Natl. Heart, Lung, and Blood Inst. of the NIH, Bethesda, MD, [7]Boston Univ. Chobanian & Avedisian Sch. of Med., Boston, MA, [8]Boston Univ. Sch. of Publ. Hlth., Boston, MA

Background. Blood concentration of lipids including triglycerides (TRIG), high and low density lipoprotein cholesterol (HDL and LDL, respectively), and total cholesterol (TC) are well-established risk factors for cardiovascular disease and diabetes. In this study we evaluated whole blood gene expression associated with blood lipid levels to identify genetic regulatory mechanisms. This study, to our knowledge, is the first large-scale gene expression analysis of blood lipids.
Methods. We conducted and meta-analyzed transcriptome-wide association studies of four lipid measures (TRIG, HDL, LDL, TC) from peripheral whole blood among self-reported racially and ethnically diverse cohorts. The discovery participants were from the Cameron County Hispanic Cohort, two Framingham Heart Study (FHS) substudies, and one Women's Health Initiative (WHI) substudy [n=4,684; 43% European American (EA), 24% Black or African American (AA), 33% Hispanic or Latino/a (HL)]. The replication participants were from additional FHS and WHI substudies (n=2,405; 82% EA, 14% AA, 4% HL). After quality control, we performed mixed effects linear regression analyses on ~20,000 genes adjusted for age, sex, body-mass index, smoking, population structure, cell type composition, and technical covariates. Genes with $p<0.05/\sim20{,}000$ genes in the discovery stage were carried forward to an independent sample for replication ($p<0.05$/number of discovery genes). Gene Ontology for Biological Process was assessed for each lipid trait to gain mechanistic insight of differentially expressed genes. Two-sample bidirectional Mendelian randomization (MR) analysis was performed to determine whether differential expression was causal or consequential to blood lipid level.
Results. From the discovery meta-analyses, we replicated 1,419 genes significantly associated with TRIG, 511 with HDL, 5 with LDL, and 139 with TC. Functional enrichment categories for TRIG, HDL, and LDL included protein targeting to endoplasmic reticulum, oxidative phosphorylation, and cholesterol metabolic process. *FADS1*, a gene that regulates fatty acids and was enriched in the lipid metabolic process pathway for LDL, was significant in MR analyses identifying *FADS1* gene expression as causal for LDL, HDL, and TC blood lipid levels.
Conclusion. Overall, we identified multiple lipid-related genes in self-reported racially and ethnically diverse individuals that provide new insights into biological pathways and genetic regulatory mechanisms involved in modulating blood lipid concentrations. On-going work includes sensitivity analyses to evaluate clinical comorbidities and other potential confounders.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4406 Transcriptome-wide Causal Inference Identifies Potential Causal Genes for Male-pattern Baldness

Authors:

B. Kim[1,2], J. Shim[3], D. Kim[1], S-H. JUNG[4], M. Song[1,5], I. Shim[1], B. Jang[1,6], M. Woojae[2], C. Jeong[7], J. Lee[3], J-W. Kim[3], W-Y. Park[3,1], D. Lee[3], H-H. Won[1,3]; [1]Sungkyunkwan Univ., Seoul, Korea, Republic of, [2]Seoul Natl. Univ. Bundang Hosp., Seongnam, Korea, Republic of, [3]Samsung Med. Ctr., Seoul, Korea, Republic of, [4]Univ. of Pennsylvania, Philadelphia, PA, [5]Washington Univ. Sch. of Med., St. Louis, MO, [6]Icahn Sch. of Med. at Mount Sinai, New York, NY, [7]Seoul Natl. Univ., Seoul, Korea, Republic of

Male-pattern baldness (MPB) is a polygenic phenotype with subtype-specific causes. Although previous genome-wide association studies (GWASs) identified hundreds of genetic loci associated with MPB, it is challenging to prioritize potential causal genes for MPB. Here, we conducted a GWAS and transcriptome-wide Mendelian randomization (TWMR) of MPB subtypes to identify directly associated genes. A total of 186,093 male individuals of European ancestry from the UK Biobank were included in this study: no progression of hair loss (Pattern 1, n = 59,168), M-shaped hair loss (Pattern 2, n = 42,923), O-shaped hair loss (Pattern 3, n = 49,798), and severe hair loss (Pattern 4, n = 34,204). Phenotypes for MPB were defined as 1) severity of MPB (MPB0, continuous from the Pattern 1 to 4), 2) M-shaped baldness (MPB2, Pattern 1 versus 2), 3) O-shaped baldness (MPB3, Pattern 1 versus 3), and 4) severe baldness (MPB4, Pattern 1 versus 4). GWAS of each MPB subtype was performed using SAIGE adjusted for age, age^2, genotype batch, and the first 10 principal components. TWMR of each MPB subtype was conducted using SMR with expression quantitative loci (eQTL) in seven relevant tissues from the Genotype-Tissue Expression project: cell cultured fibroblasts, lower leg skin, suprapubic skin, skeletal muscle, prostate, testis, and thyroid tissues. For more reliable evidence of direct association, we performed a colocalization analysis between eQTLs on significant genes from TWMR. To investigate the potential pleiotropy of candidate genes, we conducted phenome-wide Mendelian randomization with 332 phenotypes from the pan-ancestry genetic analysis of the UK Biobank. A total of 539, 110, 176, and 379 lead variants that reached a genome-wide significant level ($P < 5.0 \times 10^{-8}$) were identified in GWAS of MPB0, MPB2, MPB3, and MPB4, respectively. We identified 61 directly associated genes that were significant in both TWMR and colocalization analyses; 52, 2, 18, and 37 genes for MPB0, MPB2, MPB3, and MPB4, respectively. Of the identified genes, 36.1% showed no significant potential pleiotropy. These candidate genes include Wnt/β-catenin signaling pathway-related genes such as *CTNNB1* (β-catenin), *HIC1*, *FTX*, *MSI1*, *IRF1*, *FAM53B*, *IQGAP1*, and *ZBTB38*. Notably, *EFEMP1*, a non-pleiotropic gene identified in this study, was directly associated with MPB4 and has been reported to be related to hair growth, but there is no medication targeting this gene. Our study provides directly associated genes for MPB that might be potentially causal for MPB and have less pleiotropic effects on other phenotypes. These candidate genes need to be further validated functionally as therapeutic targets.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4407 Transcript-specific rare-variant analysis identifies novel transcript-trait associations in the UK Biobank

Authors:

L. Ward[1], R. Hoffing[2], A. Deaton[3], A. Holleman[3], L. Krohn[4], P. LoGerfo[3], M. Plekan[3], S. Akle[3], P. Nioi[1]; [1]Alnylam, Cambridge, MA, [2]Alnylam Pharmaceuticals, Somerville, MA, [3]Alnylam Pharmaceuticals, Cambridge, MA, [4]Alnylam Pharmaceuticals, Montreal, QC, Canada

A single gene can produce multiple transcripts with distinct molecular functions. Rare-variant association tests often aggregate all coding variants across individual genes, without accounting for their presence or consequence in resulting transcript isoforms. To test the utility of transcript-aware variant masks, rare predicted loss-of-function (pLOF) variants were aggregated for 17,057 protein coding genes using 56,262 distinct transcript-specific variant masks. These masks were tested for their association with 318 quantitative phenotypes across 406,922 individuals in the UK Biobank. The transcript-specific approach resulted in 796 transcript-phenotype associations that were more statistically significant than the gene-based alternative, including *LDLR* transcript ENST00000252444 and medication-adjusted LDL (transcript-specific statistic, $p = 5.7x10^{-66}$, effect = 1.75 SD increase; gene-based statistic, $p = 1.1x10^{-27}$, effect = 0.80 SD increase). Of those associations, 256 were significant using the transcript-specific approach but not the gene-based approach, including *PCSK5* transcript ENST00000376752 and standing height (transcript-specific statistic, $p = 1.3x10^{-16}$, effect = 0.72 SD decrease; gene-based statistic, p = 0.02, effect = 0.04 SD decrease). This approach demonstrates the importance of considering the effect of pLOFs on specific transcript isoforms when performing rare-variant association studies.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4408 TransTWAS: A Multi-tissue Transcriptome-wide Association Studies with High Dimensional Transfer Learning

Authors:

D. Lai[1], H. Wang[1], T. Gu[2], Y. Zhang[1,3]; [1]Dept. of Statistics and Actuarial Sci., Faculty of Sci., The Univ. of Hong Kong, Hong Kong, China, [2]Dept. of Biostatistics, Columbia Univ., New York, NY, [3]Ctr. for PanorOmic Sci., Li Ka Shing Faculty of Med., The Univ. of Hong Kong, Hong Kong, China

Transcriptome-wide association study (TWAS) utilizes gene expression to examine the association between a patient's genotype and corresponding complex trait. TWAS imputes gene expression levels from genotypes through samples with matched genotypes and gene expression levels in a fixed human tissue; thus, the genetic architecture of complex traits is estimated by the imputed gene expression level. However, a major challenge is building a robust and accurate imputation model for tissues with a limited sample size. This paper introduced an imputation approach based on an efficient transfer learning algorithm (TransTWAS). TransTWAS leveraged information from external tissues to improve the prediction performance in the target tissue. It explicitly considered the concordance between external and target tissues; therefore, information from similar external tissues will contribute more to the information borrowing. Imputation models generated from GTEx and GEUVADIS panels are provided. Extensive simulations and analysis on multiple GWAS summary statistics datasets showed considerable improvements in statistical power and replication rate compared with existing single- (e.g., PrediXcan) or multiple-tissue methods (e.g., UTMOST).

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4409 Tree-based testing procedure for phenome-wide association studies

Authors:

D. Posner[1], E. D. Kolaczyk[2], K. Cho[1], P. W. Wilson[3], J. Dupuis[2]; [1]VA Boston Hlth.care System, Boston, MA, [2]McGill Univ., Montreal, QC, Canada, [3]VA Atlanta Hlth.care System, Decatur, GA

Biobanks have broadened the scope of epidemiological inquiries, enabling the discovery of shared genetic architectures between thousands of heritable traits. However, the increase in scope is accompanied by a commensurate increase in multiple testing. We propose a tree-based testing procedure (TreePCO) to lessen the multiple testing burden of large-scale biobank studies. Association of a biomarker with a broad trait group is tested at a less stringent significance level and, if rejected, the level is passed to tests of more specific trait associations. Extensive simulations for familywise error rate (FWER) and statistical power were conducted to test one genetic variant for association with each of 2000 traits. The traits were partitioned into 20 broad groups, then further into 5, 50, and 100 subgroups of increasingly specific hypotheses. FWER was controlled across all tree levels and various trait correlation structures. Simulations for power used an additive model with causal variant of allele frequency of 0.1, odds ratio of 1.29, case prevalence of 0.2, and sample size of 500,000. Power of TreePCO was greater than Bonferroni correction in all scenarios except weak pleiotropy (<0.5% of associated traits) with non-clustered causal associations. To validate TreePCO in practice, we re-analyzed UK Biobank (n=400,000) summary statistics for the associations between GCKR variant rs1260326 and 1380 phenotypes derived from electronic health records (PMID: 32504056). Using TreePCO, we found 6 new associations which were not identified by Bonferroni correction: substance addiction and disorders (β=-0.08, p = 1.9 x 10^{-4}), alcohol disorders (β=-0.05, p = 1.1 x 10^{-4}), noninfectious gastroenteritis (β=0.04, p = 2.4 x 10^{-4}), chronic liver disease and cirrhosis (β=0.08, p = 2.8 x 10^{-4}), cholelithiasis with acute cholecystitis (β=-0.15, p = 4.1 x 10^{-5}), and urinary calculus (β=0.07, p = 5.3 x 10^{-4}). All Bonferroni-significant associations were also identified by TreePCO. The associations for alcohol disorders (β=-0.04, p = 6.41 x 10^{-17}), chronic liver disease (β=0.09, p = 3.1 x 10^{-17}), and urinary calculus (β=0.04, p = 3.1 x 10^{-11}) were successfully replicated in European-ancestry participants of the Million Veteran Program (n=442,734). The associations with alcohol consumption and liver disease are well-documented, but less is known about rs1260326 and urinary calculus or kidney stones. Mediators of the association likely include triglyceride-glucose index (PMID: 34616176) and other metabolic biomarkers. In summary, TreePCO is a powerful testing procedure for large-scale biobank studies, including but not limited to phenome-wide association.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4410 Two-sample Mendelian Randomization of Major Depressive Disorder and Inflammatory Bowel Disease

Authors:

C. Simpson, S. R. Brant; Univ. of Tennessee Hlth.Sci. Ctr., Memphis, TN

Inflammatory bowel disease (IBD), including the subtypes Crohn's disease and ulcerative colitis, is a chronic, immune-mediated disease that is typically diagnosed in the late teens or twenties. However, it can begin at any age. Symptoms include persistent diarrhea, abdominal pain, cramping, loss of appetite, weight loss, and fever. A significantly lower quality of life has been reported in many patients. This effect appears to go beyond the simple burden of dealing with the disease as the complex relationship between the gut and the brain may increase the risk to patients with IBD of developing mental illness. Whether the effect is unidirectional, bidirectional, or caused by common external risk factors is unknown. This study aims to explore the causal relationship between IBD and major depressive disorder. Using publicly available genome-wide association study summary datasets, we selected SNPs from GWAS of Major Depressive Disorder (MDD) (N=480,359), Crohn's Disease (CD)(N=51,874), Ulcerative Colitis (UC)(N=47,745) and Inflammatory Bowel Disease (IBD)(N=65,642). We performed inverse-variance weighted (IVW) two-sample Mendelian randomization (TSMR). We observed an inverse association with the outcomes CD (beta=2.23E-02, P=0.04) and UC (beta=0.37, P=0.04) with MDD, but not with IBD (beta=-0.14, P=0.87). Analyses are ongoing and will be presented.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4411 Unbiased functional genomics-driven discovery of Alzheimer's Disease mechanisms via analysis of 382 independent GWAS loci

Authors:

Y. Leung, P. Kuksa, J. Cifello, O. Valladares, G. Schellenberg, L-S. Wang; Univ. of Pennsylvania, Philadelphia, PA

Most genomic loci identified by genome-wide association studies (GWAS) are in strong linkage disequilibrium (LD) with many nearby variants. Although all of these variants could be functional, they are non-coding and thus any underlying gene regulatory mechanisms must be context-dependent and hypothesized. This phenomenon is particularly true for complex diseases such as Alzheimer's Disease (AD).

We investigated non-coding regulatory mechanisms underlying AD GWAS variants (Bellenguez *et al* 2022) in enhancer regions using SparkINFERNO (Kuksa *et al* 2020), a recent non-coding variant analysis pipeline, and FILER (Kuksa *et al* 2022), an extensive functional genomics (FG) catalogue. SparkINFERNO performs direct analysis of full GWAS summary statistics in a genome-wide manner. This automated analysis is hypothesis-free, and provides a systematic report of all candidate causal variants at the SNP, LD block, locus, enhancer or eQTL (expression quantitative trait loci). This analysis identifies and aligns evidence for the nominated causal variants, context-specific regulatory mechanisms, and target genes underlying observed GWAS signals then links them to disease mechanisms.

SparkINFERNO facilitates biologists' ability to identify the tissue context for biological validation: a total of 7863 candidate variants (382 loci) were analyzed (compared to 75 variant/loci reported). Of the top 20% of the enhancer enrichments, 87% are immune/brain (known AD mechanisms). We also found digestive and adipose tissue enrichments (17% of all enhancer enrichments), indicating gut-microbiome or other metabolism-related mechanisms could play a role in AD pathophysiology.

SparkINFERNO can also pinpoint cell type/tissue specificity at the locus level: *CLNK*, *ERCC1*, *MME*, *PLCG2* and *RASA1* are blood/brain-specific. Two are novel plausible genes: 1) *ERCC1* is involved in multiple DNA repair pathways and is also a significant eQTL in the AMP-AD brain study; 2) *RASA1* is expressed highly in neurons (The Human Protein Atlas) and is also a brain eQTL. We also nominated *HS3ST5* and *MARK4* to be digestive-specific and related to the gut-microbiome. Independently, there is metabolomic support from AMP-AD on these two genes.

By integrating different FG datasets profiled in a variety of tissue contexts with GWAS signals, SparkINFERNO revealed relevant tissue contexts for AD. Insights yielded by SparkINFERNO in a hypothesis-free manner can help researchers design context-specific experiments on divergent/novel cell types to understand AD and other phenotypes from different perspectives.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4412 Unlocking missed structural variants: enhancing short-read data with haplotype-informed methods

Authors:

S. Rubinacci[1,2,3], P-R. Loh[1,2,3]; [1]Div. of Genetics, Dept. of Med., Brigham and Women's Hosp. and Harvard Med. Sch., Boston, MA, [2]Ctr. for Data Sci., Brigham and Women's Hosp., Boston, MA, [3]Program in Med. and Population Genetics, Broad Inst. of MIT and Harvard, Cambridge, MA

Structural variants (SVs) play a crucial role in genetic diversity and disease susceptibility. Current SV calling methods based on short-read whole-genome sequencing (WGS) have limitations in capturing the full spectrum of SVs, and recent analyses of long-read sequencing in the Human Pangenome Reference Consortium (HPRC) dataset have confirmed that only one-third of SVs are accessible with existing short-read-based methods. However, recent work (Mukamel et al. 2021, Science; Hujoel et al. 2022, Cell) suggests that it may be possible to unlock many more detectable SVs from short-read WGS by developing haplotype-informed methods that amplify read-level signals (e.g. read-depth, discordant read pairs and split reads) through data aggregation across distantly related samples. We therefore evaluated the performance of current SV calling pipelines by analysing 39 samples present in both the high-coverage 1000 Genomes SV callset and the HPRC dataset. Our focus was on identifying read-level SV signatures in short-read WGS for SVs called exclusively within the HPRC dataset. We utilised annotations from the Genome in a Bottle consortium, restricting our examination to unambiguously mappable regions.

Within the 1000 Genomes dataset, we observed an average of 1412 deletions per sample in the considered region (in contrast to 1931 in HPRC), with 76% of them exhibiting similar calls in the HPRC dataset. Notably, the use of long reads in HPRC enabled the detection of short deletions (average size of 325bp), whereas short-read sequencing primarily captured larger deletion events (average size of 727bp). We also identified an average of 1887 insertions (vs. 7047 in HPRC); however, only 42% of these insertions had a similar call in the HPRC dataset, highlighting the challenge of detecting and reporting accurate breakpoint insertions with short-reads.

Examining unique calls in the HPRC dataset, we found that approximately 20% of the identified insertions were mobile element insertions (MEI). These insertions showed discordant read pair signals in short-read sequencing data. Interestingly, among the MEIs present in the 1000 Genomes SV callset, 87% were found in HPRC, indicating the high precision of current methods in detecting MEIs, yet at the cost of lower recall. Furthermore, variable number tandem repeats, which accounted for nearly 40% of SVs in the HPRC dataset, were rarely found in the 1000 Genomes SV callset.

These findings motivate further work we will pursue to develop novel software capable of efficiently capturing these signals and integrating them across shared haplotypes within large biobank cohorts.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4413 † Unsupervised clustering revealed metabolic syndrome endophenotypes with differing phenotypic and genotypic traits in UK and Taiwan Biobanks

Authors:

A. Lim[1], E. Lim[2], P-L. Chen[3], C. Fann[4]; [1]Academia Sinica, Taipei, China, [2]Academia Sinica, Taipei, Taiwan, [3]Natl. Taiwan Univ. Hosp., Taipei, Taiwan, [4]Inst Biomed Sci, Acad Sinica, Taipei, Taiwan

Background: High heterogeneity of complex diseases impedes effective translational research from genetic findings.
Methods: As a proof-of concept, clinically-relevant endophenotypes were identified through unsupervised clustering in UK Biobank (UKB) metabolic syndrome (MetS) individuals. GWAS and associated post-GWAS analysis were performed to identify genotypic traits associated with each endophenotypes. Potential drug repurposing targets were established through unbiased identification of drugs targeting associated genes. Heritability and polygenic risk score models for each endophenotypes were calculated. MetS endophenotypes were further evaluated and validated in Taiwan Biobank (TWB).
Results: Five MetS endophenotypes were identified in UKB and TWB with similar proportion, comparable phenotypic traits and associated clinical outcomes: Cluster 1 (C1) non-descriptive, Cluster 2 (C2) hypertensive, Cluster 3 (C3) obese, Cluster 4 (C4) lipodystrophy-like, and Cluster 5 (C5) hyperglycaemic. *LPCAT2*, *NUDT21* and *OGFOD1* were associated with all clusters in UKB except C2, this finding was validated in C1 and C3 of TWB. When analysing the cluster-specific genes in UKB, C1 had 156 distinct genes while C2 had 16, C3 had 98, C4 had 133 and C5 had 8. C1 GWAS revealed cardiac-specific *TRIM63* and *MYBPC3*, skeletal muscle-specific *MYLPF* and *RAPSN*. C5 GWAS identified known T2D genes such as *TCF7L2*, *IRS1*, *BBIP1* and *GIN1*. C1, C3 and C4 were associated with genes highly expressed in brain tissues such as *CN1H2*, *TMEM151A*, *MT3* and *C1QTNF4*. Comparing genotypic traits between TWB and UKB, C1 of TWB overlapped at 43 genes, C3 overlapped at 21 genes and C4 overlapped at 112 genes with UKB. Example of overlapped genes for C1: *LPL*, *TOMM40*, C3: *FTO*, *SLC6A2* and C4: *LPL*, *APOA1*, *APOC1*, *PCSK7*, *GCKR*. C1-associated genes were enriched in gene-set targeted by cardiovascular system drugs such as diuretics targeting *SLC12A3* and *SLC12A4*; C3 by anti-obesity drugs targeting *SLC6A2* and antithrombotic drugs targeting *F2*, *TFPI* and *VEGFA*; C4 by lipid-modifying drugs targeting *APOB* and *LPL*; C5 by drugs of alimentary tract targeting *CES1*. With full sample size, our first-of-its-kind PRS for C4 performed the best at lower p-value thresholds. Based on subset of 3800 cases (as C5 with n=3869 was the smallest cluster), three out of five PRS models for MetS endophenotypes performed better than that of all MetS across p-value thresholds.
Conclusion: The novel combination of approaches from machine learning and genetic epidemiology allowed for identification of clinically important MetS endophenotypes across two populations, representing a key step towards precision medicine in complex diseases.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4414 Unveiling genetic insights into glaucoma subtypes: genome-wide multi-ethnic meta-analysis identifies 22 independent risk loci for normal-tension glaucoma and elucidates genetic similarities and distinctions with high-tension glaucoma.

Authors:

P. Gharahkhani[1], S. Diaz-Torres[1], W. He[1], R. Yu[1], A. Khawaja[2], C. Hammond[3], P. Hysi[4], L. Pasquale[5], Y. Wu[1], M. Kubo[6], M. Akiyama[7], T. Aung[8], C-Y. Cheng[9], C. Khor[10], P. Kraft[11], J. Kang[12], A. Hewitt[13], D. Mackey[14], J. Craig[15], J. Wiggs[16], J. Ong[17], S. MacGregor[1]; [1]QIMR Berghofer Med. Res. Inst., Brisbane, Australia, [2]Moorfields Eye Hosp, London, United Kingdom, [3]King S Coll. London, London, United Kingdom, [4]KCL St. Thomas Hosp., London, United Kingdom, [5]Mount Sinai Hlth.System, New York, NY, [6]Riken Ctr. for Integrative Med. Sci., Yokohama, Kanagawa, Japan, [7]Kyushu Univ., Fukuoka, Japan, [8]Ophthalmology & Visual Sci. Academic Clinical Program, Duke-NUS Med. Sch., Singapore, Singapore, [9]Singapore Eye Res. Inst., Singapore, Singapore, [10]Div. of Human Genetics, Genome Inst. of Singapore, Singapore, Singapore, [11]Harvard T.H. Chan Sch. of Publ. Hlth., Cambridge, MA, [12]Channing Div. of Network Med., Brigham and Women's Hosp., Harvard Med. Sch., Boston, MA, [13]Ctr Eye Res Aust, Brunswick, Australia, [14]Lions Eye Inst., Nedlands, Australia, [15]Dept. of Ophthalmology, Flinders Univ., Flinders Med. Ctr., Adelaide, Australia, [16]Harvard Med Sch, MEEI, Boston, MA, [17]QIMR Berghofer Med. Res. Inst., Herston, Australia

Background: Primary open-angle glaucoma (POAG) is often divided into two subtypes. High-tension glaucoma (HTG) is characterized by elevated intraocular pressure (IOP), while normal-tension glaucoma (NTG) is characterized by IOP in the normal range. However, this notion is still controversial as some studies argue that different tension subtypes could be part of the same pathogenic process while other studies claim that NTG represents a different etiological process where the neurodegeneration has a higher impact. This study aimed to elucidate the shared and specific genetic architecture for NTG and HTG.
Method: To identify risk loci specific to NTG we conducted a large multi-ethnic multi-trait meta-analysis of 7,942 NTG cases and 384,431 controls from studies from the United Kingdom, Canada, Finland, US, Hong Kong, Singapore and Japan, and a structural measurement of the integrity of the optic nerve, vertical cup-to-disk ratio (VCDR, N=282,100), adjusted for IOP using the mtCOJO method. We also performed an assessment of the genetic overlap between NTG and HTG (N HTG cases = 5144, N controls = 47,997) using the GWAS pairwise method (GWAS-PW).
Findings: This study identified 22 risk loci associated with NTG. Of these, 17 loci are novel for NTG, and two loci, *MIR5580* and *KPNB1*, have not previously been associated at the genome-wide significant level with glaucoma. Examination of each locus across the genome using the GWAS-PW method indicated that risk loci are shared across NTG and HTG. The magnitude of the effect of the genome-wide significant loci tends to be lower in NTG compared to their effects on HTG, particularly for IOP-related loci. Additionally, we identified 42 drug-gene interactions, including statins, with four genes (*ABCA1*, *CDKN2A*, *CDKN2B* and *ITGB3*) that were prioritized through our gene-based analysis.
Interpretation: This work expands our understanding of the genetics of NTG and highlights a strong genetic overlap between HTG and NTG. Despite the genetic overlap, we have shown that IOP-related loci tend to have a smaller effect size in NTG when compared with HTG whereas neurodegenerative loci independent of IOP have similar effect sizes on NTG and HTG. These results indicate that while there is a significant overlap in risk loci between NTG and HTG, a precise estimation of their effect sizes on NTG using larger studies could help develop genetic risk prediction models to identify individuals at a higher risk of developing NTG. We have also identified some potential targets for neuroprotective treatment through the interaction of four genes and multiple drugs.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4415 Unveiling novel endometriosis genetic associations through unsupervised, phenotypic clustering analysis of distinct clinical subtypes

Authors:

L. Guare, S. Senapati, Penn Medicine Biobank, Regeneron Genetics Center, S. Verma; Univ. of Pennsylvania, Philadelphia, PA

Endometriosis affects 10% of reproductive-age women, and yet, it goes undiagnosed for 3.6 years on average after symptoms onset. In spite of large GWAS meta-analyses, only a few dozen causal loci have been identified. We hypothesized that the challenges in identifying causal SNPs for endometriosis stem from heterogeneity across clinical and biological factors. We extracted known endometriosis risk factors, symptoms, and concomitant conditions from the Penn Medicine Biobank (PMBB) and performed t-SNE dimensionality reduction on 4,623 women with an ICD diagnosis for endometriosis. To mitigate artificial structure created by t-SNE, we performed 200 trials and averaged the resulting distance matrices. This averaged distance matrix was then utilized for hierarchical clustering. By considering distortion and silhouette score, we determined that the optimal number of clusters is 5. The 5 clusters were characterized by utilizing additional electronic health record (EHR) variables, such as procedure history, endometriosis-related comorbidities, and long-term conditions. We have designated the five subtype clusters as reproductive-health, uterine-health, EHR-asymptomatic, pain/inflammatory, and cardiometabolic based on enriched features from each group. The subtypes were observed in the discovery PMBB cohort, and we then performed an ancestry-stratified case/control GWAS for each cluster using a combined hold-out data from PMBB and eMERGE (N cases = 2,624). Two clusters had statistically significant associations in the PMBB+eMERGE dataset: the reproductive health endometriosis subtype was associated with *LOC105374621* (lead SNP rs60132896. P = 2.4x10-8) and the cardiometabolic subtype was associated with *SNORD28B* (lead SNP rs4464815. P = 4.5x10-8). Additionally, the cardiometabolic subtype was suggestively associated with one of the known endometriosis loci, *RNLS* (lead SNP rs61855396. P = 2x10-6). The statistically suggestive hits in this study were examined for previous associations in the GWAS catalog. The reproductive health subtype exhibited significant overlap with genes related to body measurements and neurological traits, while the EHR-asymptomatic, cardiometabolic, and pain/inflammatory subtypes showed suggestive associations with genes linked to blood protein measurements and asthma, DNA methylation, type 2 diabetes and smoking status, and hematological measurements and body weight, respectively. This study significantly enhances our understanding of the clinical presentation patterns of endometriosis subtypes, showcasing the innovative approach employed to investigate these complex diseases.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4416 Use of personalized reference sets for polygenic scores negates need for arbitrary groupings

Authors:

A. Sartori[1], Y. Ding[2], S. Buyske[3], C. Gignoux[4], C. Haiman[5], E. Kenny[6], C. Koopeberg[7], K. North[8], B. Pasaniuc[2], G. Wojcik[1]; [1]Johns Hopkins Bloomberg Sch. of Publ. Hlth., Baltimore, MD, [2]UCLA, Los Angeles, CA, [3]Dept. of Statistics, Rutgers Univ., Piscataway, NJ, [4]U Colorado, Anschutz Med. Campus, Aurora, CO, [5]Dept. of Population and Publ. Hlth.Sci., Keck Sch. of Med., Univ. of Southern California, Los Angeles, CA, [6]Icahn Sch. of Med. at Mount Sinai, New York, NY, [7]Div. of Publ. Hlth.Sci., Fred Hutchinson Cancer Ctr., Seattle, WA, [8]Univ North Carolina, Wake Forest, NC

Polygenic scores (PGS) helped translating GWAS results into clinical practice but with limited portability across individuals of different genetic backgrounds, especially for admixed populations with multiple recent genetic ancestries and often relying upon subjective groupings by race/ethnicity and/or “genetic ancestry groups”. This necessitates the use of a single-aggregate population reference set to estimate PGS performance, ignoring inter-individual variability. While this is meant to control for differences in allele frequencies and LD, it often combines non-comparable individuals. A personalized PGS reference set approach can overcome these problems by tailoring to single individuals. We used PGS for body mass index (BMI) and height in 41,586 individuals self-identified as African/African American (N=15993), Asian (N=4075), Hispanic/Latino (N=18950), Native Hawaiian (N=1970), and Native American (N=598) from the PAGE Study, and trained a k-nearest neighbors model (k=1,000) on PCA-based genetic distance to assess their relative position in the genetic ancestry continuum. For each index individual, we defined new reference clusters not reflecting continental-level genetic ancestry, nor racial/ethnic groupings, thus capturing inter-individual variation. After applying PGS models, we observed that PGS accuracy performed equally well with the new reference sets approach compared to grouping by racial/ethnic group (R^2=0.092 to R^2=0.093 for PGS_{BMI} and R^2=0.543 to R^2=0.545 for PGS_{Height}). In addition, when we compared all individuals to those at the top 10th percentile of the PGS distribution, no variation in the effect size magnitude was found (B=0.08 to B=0.08 for PGS_{BMI} and B=0.53 to B=0.53 for PGS_{Height}). Finally, we computed individual PGS accuracy, which has been previously shown to be highly correlated with genetic distance along the genetic ancestry continuum, within the new reference sets and assessed performance to explore potential benefits to those on the edges of previous groupings. Combining an individualized reference set approach with individual PGS will provide a robust tool for PGS estimation in a highly personalized way, holding the potential to benefit diverse individuals and translate clinically useful information into precision medicine without the need to create borders where none exist.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4417 Using a multivariate genotype-phenotype approach to measure endocrine-mediated masculinity and femininity in human faces.

Authors:

C. Da Silva[1,2,3], H. Hoskens[1,2,3], D. J. Aponte[4], B. Hallgrimsson[1,2,3]; [1]Univ. of Calgary, Calgary, AB, Canada, [2]Alberta Children's Hosp. Res. Inst., Calgary, AB, Canada, [3]McCaig Inst. for Bone and Joint Hlth., Calgary, AB, Canada, [4]Deep Surface AI, Calgary, AB, Canada

Males and females have known differences in facial shape. This variation produces sex biases in congenital abnormalities, biting mechanics, and sleep-disordered breathing. Despite this, little is understood about the origins of these differences. The prevailing opinion is that facial sexual dimorphism results from testosterone-induced masculinization of the male face, with little consideration given to the role of estrogen. Moreover, facial femininity and masculinity are assumed to be mutually exclusive (i.e., femininity is the absence of masculinity), but this has never been tested formally. In this study, we address the current gaps by comparing patterns of endocrine-mediated facial femininity and masculinity in an independent sample of European descent. These include three-dimensional facial surface scans and genotype data for 1,000 individuals from the United States (500 males and 500 females). We begin by identifying the facial shape effects associated with the estrogen and testosterone pathways using a multivariate genotype-phenotype (MGP) approach. First, we select gene sets for the estrogen and testosterone pathways. The MGP method then identifies axes of shape variation that maximally covary with these sets of genes. Next, we use geometric morphometrics to quantify their respective shape effects and determine the dimensionality of the facial sexual dimorphism axis or axes by calculating the degree of correlation between the shape effects of estrogen-mediated femininity and testosterone-mediated masculinity. The degree of correlation indicates whether endocrine-mediated facial femininity and masculinity are on opposing sides of the same axis or on separate axes. Finally, we compare results between our male and female samples. This project presents a conceptually and methodologically novel approach to studying sexual dimorphism, combining techniques from statistical genetics and geometric morphometrics. Our method provides a biologically meaningful way of measuring morphological masculinity and femininity, which is an important first step toward understanding the developmental-genetic aetiology of sex differences in the human face.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4418 Using alternative genetic models in IPF susceptibility genome-wide association studies

Authors:

T. Hernandez Beeftink[1,2], D. Chin[1,2], L. Donoghue[3], B. Guillen-Guio[1,2], O. C. Leavy[1,2], P. Cullinan[4], C. Reynolds[4], F. Martinez[5], I. Noth[6], CleanUP-IPF Investigators of the Pulmonary TrialsCooperative, C. Flores[7,8,9], J. M. Oldham[10], D. A. Schwartz[1], B. L. Yaspan[11], R. Jenkins[4], R. J. Allen[1,2], L. V. Wain[1,2], on behalf of the International IPF Genetics Consortium, International IPF Genetics Consortium: H. L. Booth, W. A. Fahy, I. P. Hall, S. P. Hart, M. R. Hill, N. Hirani, R. B. Hubbard, T. M. Maher,R. J. McAnulty, A. B. Millar, P. L. Molyneaux, V. Navaratnam, E. Oballa, H. Parfrey, G. Saini, I. Sayers, M. D. Tobin, M. K. B. Whyte, A. Adegunsoye, N. Kaminski, SF Ma, M. E. Strek, Y. Zhang, T. E. Fingerlin, M. Molina-Molina, A. Stockwell, M. Neighbors, X. R. Sheng, M. McCarthy; [1]Univ. of Leicester, Leicester, United Kingdom, [2]NIHR Leicester BioMed. Res. Ctr., Leicester, United Kingdom, [3]Genentech, San Francisco, CA, [4]Imperial Coll. London, London, United Kingdom, [5]Weill Cornell Med., New York, NY, [6]Univ. of Nottingham, Nottingham, United Kingdom, [7]Hosp. Univ.rio N.S. de candelaria, Santa Cruz De Tenerife, Spain, [8]Inst. Tecnologico y de Energias Renovables, Santa Cruz de Tenerife, Spain, [9]CIBERES - Inst. de Salud Carlos III, Madrid, Spain, [10]Univ. of Michigan, Michigan, MI, [11]Genentech, Pacifica, CA

Rationale: Idiopathic pulmonary fibrosis (IPF) is a rare progressive illness of the respiratory system with a poor prognosis and lack of therapeutic options. Studies have focused on the identification of genetic risk factors associated with IPF using genome-wide association studies (GWAS), which have highlighted new genes and pathways of interest. Objective: All prior GWAS of IPF have assumed an additive genetic model. We hypothesised that assuming non-additive associations may identify new IPF risk signals where the underlying causal effect is consistent with a dominant or recessive genetic model. Methods: We conducted the largest IPF risk GWAS using both dominant and recessive genetic models with data from seven independent studies (involving 5,159 IPF cases and 27,476 controls), imputed using TOPMed. We used logistic regression, adjusting for 10 principal components for ancestry, and meta-analysed results using inverse-variance weighted meta-analysis. We selected signals which were present in at least two studies and met genome-wide significance (p<5x10-8). We compared significant signals from the dominant and recessive models with results from an additive model GWAS using the same data. Results: Around 55 million SNPs were analysed using the dominant model, while 9 million SNPs were studied under the recessive models. We obtained 58 genome-wide significant signals using the dominant model and 25 for the recessive model. Of these, 18 signals were not genome-wide significant under an additive genetic model. Fourteen of the 18 signals were genome-wide significant under the dominant model (p varied between 3.76x10-8 - 1.65x10-11) and four under the recessive model (p varied between 4.51x10-8 - 5.11x10-12). Two of the recessive model signals were exonic signals located in the cell-cycle gene Polyamine-Modulated Factor 1 (*PMF1*) and in Epsin 3 (*EPN3*). EPN3 was previously found to be upregulated in lung tissue of IPF patients. Of previously reported additive signals, a SNP at desmoplakin (*DSP*) was most significant under a recessive model whilst the SNP at telomere gene, *RTEL1*, was most significant under a dominant model. Conclusion: The use of alternative genetic models in GWAS of IPF susceptibility enabled discovery of new signals that were not reported at genome-wide significance under the additive model. Understanding the causal model of association can inform downstream functional genomics studies.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4419 Using external reference panel and single-variant summary statistics for rare-variant aggregation tests.

Authors:

B. Ryan[1], R. Welch[1], M. Boehnke[1], C. Fuchsberger[2]; [1]Univ. of Michigan, Ann Arbor, MI, [2]Eurac Res., Bolzano, Italy

Genome-wide association studies (GWAS) have identified hundreds of thousands of associations between common genetic variants and a range of human diseases and traits. These studies are often of low power to identify associations with rare genetic variants. To increase power, aggregation tests which pool the genetic signal across multiple variants in a gene or other region of the genome to test the cumulative effect of these variants on a disease or trait are often used. There are many publicly available GWAS summary statistics, and it has been shown that aggregation tests can be performed using only single-variant score statistics, their covariances, and the allele frequencies for each variant. Accurate estimates of single-variant test statistic covariances are required, and are easily obtained given individual-level genetic data. Due to privacy concerns, individual-level genetic data often cannot be shared. Sharing covariance matrices may also be infeasible due to data size. In this study, we apply a previously proposed method of estimating single-variant test statistic covariance from an external reference panel to perform aggregation tests. We use correlation between variants in the reference panel and allele frequencies from the original sample to estimate the test statistic covariances. We propose a two-stage approach by first filtering genes using a null covariance where covariance between all variants is set to zero to perform aggregation tests, and in stage two using an external reference panel to repeat aggregation tests only on those genes passing a p-value threshold. Using the null covariance generally overestimates association significance of genes, allowing us to confidently exclude non-significant genes in the first stage. As covariance computation is the most burdensome of the testing process, this leads to a substantial reduction in computation time. We investigate the properties of our method across ten traits from the UK Biobank, three variant annotation masks, and three different aggregation tests. When p-values on the log scale produced using individual-level data and our two-stage approach across all ten traits are compared, Pearson correlation coefficients are between 0.984 and 0.996 across all scenarios using a reference panel of 1,000 individuals. By using an African American reference panel (n=1,000) from the InPSYght study, we show that this approach is robust to misspecification of the reference panel ancestry. The two-stage approach results in substantial memory and computation savings and will be a useful tool for researchers wishing to perform rare-variant aggregation tests using publicly available summary statistics.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4420 Using F_{ST} to understand putative bias induced by population stratification in Mendelian randomization studies

Authors:

N. Khankari, P. Straub, A. Konkashbaev, F. Blostein, J. Hirbo, L. K. Davis, M. C. Aldrich, T. L. Edwards, N. J. Cox; Vanderbilt Univ. Med. Ctr., Nashville, TN

Two-sample Mendelian randomization (MR) relies upon existing GWAS summary statistics for the exposure and outcome from independent samples. A potentially causal association may be estimated via inverse-variance weighted (IVW) MR leveraging these readily available GWAS summary statistics. While the current GWAS landscape is changing with the advent of large-scale genomic studies conducted among diverse populations, many existing GWAS have focused solely on individuals of European ancestry. Given allele frequencies may vary even among individuals of a seemingly homogenous European population, resulting two-sample MR conducted in European populations may remain biased due to population stratification. We sought to understand the putative bias induced in two-sample MR studies when the existing GWAS summary statistics may be affected by population stratification. In this analysis, four scenarios were examined: (1) when neither the exposure (i.e., body mass index) nor the outcome GWAS (i.e., coronary artery disease) are affected by population stratification; (2) only the exposure GWAS (i.e., triglycerides) is affected by population stratification; (3) only the outcome (i.e., Celiac disease) GWAS is affected by population stratification; and (4) both exposure (i.e., triglycerides) and outcome (i.e., Celiac disease) GWAS are affected by population stratification. To compare how the estimated IVW MR effect varies according to population substructure in these scenarios, we conducted a fixation index (F_{ST})-weighted MR analysis to examine differences according to the four 1000 Genomes populations (i.e., FIN, GBR, IBS, and TSI). Pairwise F_{ST} was calculated comparing each European 1000 Genomes population with the CEU population, which was then used to weight the IVW MR estimate accordingly. We observed the largest differences between the estimated IVW MR [Beta (StdErr): -1.17 (0.63)] and F_{ST}-weighted population-specific MR effects (i.e., FIN: -2.22 (0.10), GBR:-3.79 (0.08), IBS: -2.71 (0.09), and TSI: -3.52 (0.10)] in the fourth scenario in which both exposure (i.e., triglycerides) and outcome (i.e., Celiac disease) were affected by population stratification bias. Using F_{ST}-weighted MR as a two-sample MR sensitivity analysis may help estimate ancestry-specific effects when GWAS summary statistics are affected by population stratification.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4421 Using information technology to optimise the differential diagnosis of rare genetic diseases

Authors:

H. Li[1], J. Yang[2], H. Duan[2], Q. Shu[1]; [1]Children's Hosp., Zhejiang Univ. Sch. of Med., Hangzhou, China, [2]Zhejiang Univ., Hangzhou, China

Rare diseases (RDs) are defined as conditions or disorders that affect a prevalence of less than 6.5 to 10 in 10,000 individuals. Despite being referred to as“rare”, there are over 7,000 RDs that collectively affect tens of millions of people in the general population. RDs are mostly genetic diseases. The diagnosis of RDs usually involves phenomic analysis and genomic analysis, the combination of which can lead to a definitive diagnosis. However, the diagnosis of RDs is of great challenge. Firstly, RDs often have complex phenotypic features. Secondly, RDs have extensive phenotypic overlap; In addition to the complex and overlapping phenotypic features of RDs, the limited knowledge and experience of clinicians, and the low quality of clinical records, also make it difficult to diagnose RDs. In response to the needs of the field, we have developed a range of new information technology methods and tools to help clinicians and researchers in the areas of knowledge presentation and phenotype retrieval, intelligent question and answer, and differential diagnosis using a variety of new information technologies:(1) RDmap (http://rdmap.nbscn.org): A rare disease map is built to visualize the knowledge of complex rare diseases and the relationships between RDs. Functions such as phenotype-based precision matching and similarity matching are provided to locate and retrieve RDs.(2)PheLR (http://phelr.nbscn.org): A Phenotype-driven Likelihood Ratio analysis approach (PheLR) assisting interpretable clinical diagnosis of RDs. With a likelihood ratio paradigm, PheLR estimates the posterior probability of candidate diseases and how much a phenotypic feature contributes to the prioritization result.(3)RDmaster (http://rdmaster.nbscn.org): a phenotype-oriented Q&A system, to gather important phenotypic features and assist doctors in the differential diagnosis of RDs.Benchmarked using simulated and realistic patients, these information tools show significant advantages over current approaches (Phenomizer, BOQA, PhenoPro, LIRICAL, Exomiser, GPT-3.5 and GTP-4) and is robust to noise and inaccuracy. The widespread dissemination of these IT tools can improve the accuracy and timeliness of diagnosis of rare genetic diseases.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4422 Using Polygenic Risk Scores to Investigate Inverse Genetic Associations Between Cancer and Cognitive Decline

Authors:

D. Chen[1], M. Glymour[2], P. Gilsanz[3], C. A. Schaefer[3], L. C. Sakoda[3], T. J. Hoffmann[1], P. Buto[1], M. Choi[1], J. S. Witte[4], R. E. Graff[1]; [1]UCSF, San Francisco, CA, [2]Boston Univ., Boston, MA, [3]Kaiser Permanente Northern California Div. of Res., Oakland, CA, [4]Stanford Univ., Stanford, CA

Epidemiological evidence suggests that there exists bidirectional inverse comorbidity between cancer and neurodegenerative diseases, including Alzheimer's disease (AD) and dementia. Biases do not fully explain the relationship, so it has become critical to explore the complex pathophysiology underlying connections between these two diseases. We thus assessed whether shared genetic risk factors partially drive the inverse association among 93,656 members of the Kaiser Permanente Medical Care Plan who had genetic information available. To do this, we constructed polygenic risk scores for all-cause cancer (PRS_{cancer}), 17 common cancer types, and AD, with (PRS_{AD}) and without ($PRS_{ADnoAPOE}$) representation of two variants comprising the APOE haplotype (rs7412 and rs429358). The study population included individuals of African (n = 3,304), Asian (n = 7,139), European (n = 74,949), and Hispanic/Latino (n = 8,264) ancestry. Within ancestry cohorts, we used Cox proportional hazards models to estimate hazard ratios (HR) and 95% confidence intervals (CI) for associations of PRS_{cancer} with dementia (all-cause, AD, and vascular) and associations of PRS_{AD} and $PRS_{ADnoAPOE}$ with each cancer type. Within Europeans, there were 16,780 incident all-cause cancer cases and 6,820 incident all-cause dementia cases. All-cause cancer and dementia cases were lower in other ancestry groups, ranging from 589 to 1,248 and 292 to 503, respectively. We did not identify any significant associations between the cancer or dementia PRS and the alternate disease after multiple comparison adjustment ($p > 2.78x10^{-3}$) in any of the ancestry groups. For example, all-cause PRS_{cancer} was not significantly associated with all-cause dementia in African (HR: 1.04, 95% CI: (0.92, 1.18)), Asian (HR: 0.97, 95% CI: (0.87, 1.07)), European (HR: 1.02, 95% CI: (0.99, 1.05)), and Hispanic/ Latino (HR: 0.94, 95% CI: (0.85, 1.03)) ancestry participants. All HRs estimating the association between individual PRS_{cancer} associations with dementia phenotypes in European ancestry participants were between 0.94 (95% CI: 0.88, 0.99) and 1.03 (95% CI: 0.99, 1.07). Associations of PRS_{AD} and $PRS_{ADnoAPOE}$ with varying cancer types were similarly non-significant. PRS_{AD} was not associated with all-cause cancer in African (HR: 0.97, 95% CI: (0.90, 1.05)), Asian (HR: 1.04, 95% CI: (0.98, 1.11)), European (HR: 1.00, 95% CI: (0.99, 1.02)), and Hispanic/ Latino (HR: 0.98, 95% CI: (0.93, 1.03)) ancestry individuals. Methods that improve cross-population polygenic prediction in diverse cohorts, such as PRS-CSx, will be explored in subsequent analyses.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4423 Using population data to assess significance of rare variants in WGS of n=1 disease cases.

Authors:

M-S. Ogloblinsky[1], O. Bocher[1,2], C. Aloui[3], A-L. Leutenegger[3], O. Ozisik[4], A. Baudot[4], E. Tournier-Lasserve[3,5], H. Madeen[6], D. Lewinsohn[6], D. Conrad[6], E. Génin[1,7], G. Marenne[1]; [1]Univ Brest, Inserm, EFS, UMR 1078, GGB, Brest, France, [2]Inst. of Translational Genomics, Helmholtz Zentrum München, Munich, Germany, [3]Université Paris Cité, Inserm, NeuroDiderot, Unité Mixte de Recherche 1141, Paris, France, [4]Aix Marseille Univ, INSERM, Marseille Med. Genetics (MMG), Marseille, France, [5]Assistance publique-Hôpitaux de Paris, Service de Génétique Moléculaire Neurovasculaire, Hôpital Saint-Louis, Paris, France, [6]Div. of Genetics, Oregon Natl. Primate Res. Ctr., Oregon Hlth.& Sci. Univ., Portland, OR, [7]Ctr. Hosp.ier Régional Univ.ire de Brest, Brest, France

High genetic heterogeneity in rare diseases poses the challenge of identifying an n=1 patient's causal variant using sequencing data and standard analysis methods. The PSAP (Population Sampling Probability) method uses gene-specific null distributions of CADD pathogenicity scores to assess the probability of observing a given genotype in a healthy population. Here, we propose an extension of the PSAP method to the non-coding genome, using as testing units predefined regions reflecting functional constraint at the scale of the whole genome. Our method broadens the spectrum of variants detectable by PSAP and improves the performance of PSAP in coding regions.

We simulated disease exomes and genomes by inserting coding and non-coding pathogenic ClinVar variants in 574 healthy exomes and in 10 whole genomes from the 1000 Genomes Project respectively, under both the autosomal dominant and recessive (AD and AR) models. Inserted variants were ranked based on their PSAP p-values using either genes or functional regions as testing units. The percentage of variants ranked at top positions was compared between the two PSAP strategies and against a ranking based only on CADD scores.

For the AR model, 89% of 2,680 coding ClinVar variants were ranked first when using PSAP on functional regions, as compared to 83% when using PSAP on genes and 0.04% when using CADD scores only. For the AD model, 38% of 4,965 coding ClinVar variants were ranked first when using PSAP on functional regions, as compared to 27% when using PSAP on genes, and 0.1% when using CADD scores only. For non-coding ClinVar variants, 54% of the 115 variants were ranked in the top 10 for the AR model and 21% of 218 variants for the AD model using PSAP on functional regions, against 3% using CADD scores only. Interestingly, among the non-coding ClinVar variants, PSAP performed best on splicing variants, which are often highly clinically relevant.

On real data from 6 patients with distinct known variants causing Cerebral Small Vessel Disease, a very heterogeneous disease, the prioritization of the causal variants was improved by using PSAP on functional regions as compared to genes and all variants were ranked within the top 100. Additional post-filtering (e.g, keeping only variants absent from the gnomAD database) improved this ranking.

The PSAP method extended to functional regions is an efficient agnostic prioritization tool, which offers promising results for the diagnosis of unresolved n=1 cases of rare diseases. Our method is implemented as a user-friendly and versatile Snakemake workflow, accessible to both researchers and clinicians.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4424 Using protein language model annotations to improve the power of exome-wide association studies.

Authors:

J. Mbatchou, C. Gillies, J. Kosmicki, M. Ferreira, G. Abecasis, L. Habbeger, J. Marchini, on behalf of the Regeneron Genetics Center; Regeneron Genetics Ctr., Tarrytown, NY

Advances in sequencing technologies have enabled the study of rare genetic variation which can play a key role in understanding complex diseases. Due to the low frequency of these variants and the need for very large sample sizes or effect sizes, gene-based tests have been developed to increase the statistical power of association tests over single variant tests by aggregating information across multiple variants. Several tools have been developed that also integrate functional annotations into the gene-based tests by applying different weighting schemes at each variant site with the goal of boosting the power of rare variant analysis (e.g., REGENIE, SAIGE, STAAR).
While many in-silico algorithms predict deleteriousness of genetic variants (e.g., Polyphen, SIFT), recent algorithms based on protein language models, which utilize many protein sequences to derive the functional impact of sequence mutations, were shown to have comparable or better performance than predictions based on supervised methods. Here we studied the use of ESM-1v, a deep learning algorithm trained on 98M protein sequences, to identify deleterious missense variants for inclusion in gene-burden tests. We generated qualitative functional groupings derived from continuous ESM-1v scores for missense variants.
We demonstrate the usefulness of using ESM-1v through real data applications in UK Biobank with ~400,000 exome sequenced samples where we run exome-wide gene-based association tests on 100 phenotypes. We compare gene-based tests derived from ESM-1v scores to those based on five in-silico algorithms. We discover 64 additional significant associations by utilizing functional groupings derived from ESM-1v scores. We show that these groupings can recover 98.7% of the signals found by functional annotations from in-silico algorithms . Thus, the use of state-of the art missense predictors to identify variants for inclusion in gene burden aggregation can improve power for exome-wide association studies.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4425 Utility of using elastic net models to impute gene expression in the anterior cingulate cortex and implications for network analysis.

Authors:

J. Drake[1,2], Z. Taylor[2], A. Denham[1,2], S-A. Bacanu[3], J. Shin[4], T. Hyde[4], V. Vladimirov[2]; [1]Texas A&M Univ., College Station, TX, [2]Univ. of Arizona, Phoenix, AZ, [3]Virginia Commonwealth Univ, Richmond, VA, [4]Lieber Inst. for Brain Dev., Johns Hopkins Univ., Baltimore, MD

Background: Given the cost and difficulty of measuring postmortem brain gene expression (GE), imputation methods have steadily gained traction in the field. However, a major drawback of GE imputation is the limited number of imputed genes, with even fewer of these imputed with sufficient accuracy, ultimately hindering the network analysis. To address these limitations, we expanded on PrediXcan's elastic net methodology by incorporating neuronal cell fractions in order to impute a higher number of genes and evaluate their utility in network analysis using weighted gene co-expression network analysis (WGCNA).
Methods: For our GE imputation, we used a large postmortem brain sample containing genotype and RNA-Seq data generated in the anterior cingulate cortex (ACC) from 196 neurotypical, 228 major depressive disorder (MDD), and 123 bipolar disorder (BIP) subjects, with a total N=547. By representing each phenotype, proportionally, we further split the sample into 70% training and 30% testing datasets to minimize overfitting. We used glmnet to create 19,948 mRNA elastic net models with 10-fold cross-validation. Considering the importance of cell-specific gene expression, we further used a single-cell reference dataset to deconvolute the bulk mRNA GE to estimate the specific brain cell fractions, which were then incorporated into our models for increased accuracy. We retained only models with an average coefficient of determination (R2) > 0, Z-score p-value < 0.05, and average correlation between the imputed and measured expression > 0.10. Genotype and cell fraction estimates of the testing dataset were used for network analysis to create modules, which were tested for preservation against the measured expression of the training dataset.
Results: At an average R2 of 0.13, 13,445 gene models survived, which were then used to generate a scale-free topology in WGCNA. WGCNA identified 18 imputed modules in the BIP subjects, one of which was associated with BIP at a nominal p-value ≤ 0.05 and 5 others, showing high preservation to the measured expression. In the MDD subjects, we identified 19 modules, with 5 imputed modules preserved between the measured and imputed expressions and 11 modules associated with MDD at a nominal p-value.
Conclusion: The inclusion of cell fractions greatly increases the number of quality models. However, the imputation of these models requires mRNA GE data to estimate cell proportions, which limits its utility.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4426 Variance of Age-specific Log Incidence Decomposition (VALID): Application to female breast cancer.

Authors:

J. Hopper[1], J. Dowty[2], T. Nguyen[3], S. Li[3], G. Dite[3], R. MacInnis[4], E. Makalic[3], D. Schmidt[5], M. Bui[3], J. Stone[6], J. Sung[7], M. Jenkins[3], G. Giles[4], M. Southey[8], J. Mathews[3]; [1]Univ. of Melbourne, Kew, Australia, [2]Univ Melbourne, Carlton, Australia, [3]Univ. of Melbourne, Carlton, Australia, [4]Cancer Council Victoria, Melbourne, Australia, [5]Monash Univ., Carlton, Australia, [6]Univ. of Western Australia, Perth, Australia, [7]Seoul Natl. Univ., Seoul, Korea, Republic of, [8]Monash Univ., Melbourne, Australia

Background: The extent to which known and unknown factors explain how much people of the same age differ in their risk of a disease is fundamental to epidemiology. Risk factors can be correlated in relatives, so familial aspects (genetic and non-genetic) must be considered. Historically, this has been addressed by estimating 'heritability' as a percentage based on a deterministic 'liability'. **Methods:** We developed a unifying model (VALID) by assuming that risk, defined as log(incidence) or logit(cumulative incidence), increases exponentially across risk scores which have standard normal distributions. The variance in risk is Δ^2, where Δ = log(OPERA) is the difference in mean between cases and controls and OPERA is the odds ratio per standard deviation. A risk score correlated r between a pair of relatives generates a familial odds ratio of $\exp(r\Delta^2)$ from which classic familial risk variance decomposition can be applied. There is an upper limit to variance in risk caused by genetic factors, determined by the familial odds ratio for genetically identical twin pairs, but no limit to variation caused by non-genetic factors. We applied VALID to published data on breast cancer risk factors. **Results:** Variance in breast cancer risk is explained by different factors to different extents at different ages. The familial risk ratio, and therefore the familial variance, is greater at younger ages. A substantial proportion of familial variance prior to, but not after, menopause is explained by high-risk mutations in the major genes *BRCA1* and *BRCA2*, a small proportion by the latest polygenic risk score, and a substantial proportion by non-genetic factors shared by sisters or undiscovered recessively inherited genetic factors. Risk factors identified from questionnaires such as reproductive factors and body composition explain little variance in risk. New mammogram risk scores explain as much variation in familial risk as the polygenic risk score and a similar amount of variation in non-familial risk. **Conclusions:** Much is unknown about the finite genetic and familial aspects of breast cancer risk, especially for young women, and very little is known about individual-specific variance in risk which is unlimited. These findings are contrary to those from applying the liability model.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4427 What underlies changes in physiological trait heritability with age?

Authors:

S. Smith, E. Tucker-Drob, **A. Harpak**; Univ. of Texas at Austin, Austin, TX

Understanding how genetic effects on clinically-relevant phenotypes vary with age may provide insights critical to the prevention, treatment and management of diseases of aging. The heritability of complex traits often varies with adult age. However, it is currently unclear whether these differences are driven by changes in genetic architecture, or alternative mechanisms, such as changes in the effects of environmental exposures or changes in the extent of environmental variance. We developed a method to evaluate competing hypotheses observed differences in heritability across age-stratified cohorts. Specifically, across birth cohorts composing the large age-heterogeneous cohorts from the UK Biobank and All of Us data sets, we identify the patterns of change in genetic and environmental variance that underlie age-related differences in the heritability of physiological traits. For instance, in traits related to body size, including height, weight and BMI, we find evidence that decreasing heritability with age is driven by consistent decreases in genetic variance. Conversely, for lipid biomarker traits we show that vanishing heritability with cohort age is primarily driven by increases in environmental variance against a backdrop of relatively stable genetic variance. Our results indicate that the patterns of change in the heritability are trait specific, but are shared within clusters of related phenotypes. Essential to the clinical management of these traits in aging individuals, we show that differences in the heritability estimates of age stratified cohorts can be due to differences in either genetic or environmental variance.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4428 When are aggregation tests more powerful than single-variant tests for rare-variant association studies?

Authors:

D. Bose[1], C. Fuchsberger[1,2], M. Boehnke[1]; [1]Dept. of Biostatistics & Ctr. for Statistical Genetics, Univ. of Michigan, Ann Arbor, MI, [2]Inst. for Biomedicine, Eurac Res., Bolzano, Italy

While single-variant tests have proven remarkably successful in identifying associations in common variant GWAS, their power is substantially less for rare variants. Aggregation tests were developed and shown to have greater power for rare variants in some situations. Despite this, most studies that have employed both types of tests have reported more associations through single-variant tests, even for rare variants. In contrast, recent studies of large biobank-scale datasets have yielded many significant results with aggregation tests. Given these findings, it is of interest to conduct a thorough investigation into the range of genetic models and sample sizes for which aggregation tests are expected to be more powerful than single-variant tests. We consider a normally distributed trait following an additive genetic model with c causal out of v total rare variants in an autosomal gene/region with region-specific heritability h^2, measured in n independent study participants. Under the assumption of independent variants, we calculate analytically the power of single-variant (SV), burden, and SKAT tests to detect associations. Assuming the same minor allele frequency (MAF) for all variants and the same effect size and direction for causal variants, we show that power for all three tests depends on nh^2, c, and v. Next, to account for linkage disequilibrium, we perform simulations based on 378,215 unrelated exome-sequenced White British UK Biobank participants. We investigate power of aggregation and SV tests to detect association for the 1,060 genes on chromosome 2 having at least 1 protein-truncating variant (PTV) over a range of probabilities of the variants being causal. Aggregation tests require us to choose which rare variants in the gene to include in the gene mask for the test with the aim to include causal variants and exclude non-causal ones. We find that power of aggregation tests is highly dependent on the choice of the mask and genetic model. For example, with n=100,000 and h^2=0.1%, when 80%, 50% and 1% of PTVs, deleterious missense, and non-deleterious missense variants respectively in a gene are causal, aggregation tests are more powerful than SV tests for at least 54% of genes with the mask containing only PTVs and deleterious missense variants, but are more powerful for only 24% and 9% of genes with the mask containing only PTVs or PTVs and all missense variants, respectively. Overall, aggregation tests are more powerful than SV tests only when a substantial proportion of variants in the gene is causal.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session III

PB4429 Whole exome sequencing of African-Brazilian extended families with essential hypertension reveals 30 candidate genes

Authors:

A. Nato[1], V. Magalhães Borges[1,2], N. Teruel[3], L. Kimura[2], A. Horimoto[4,5], E. M. Wijsman[5,6], R. Mingroni-Netto[2]; [1]Dept. of BioMed. Sci., Joan C. Edwards Sch. of Med., Marshall Univ., Huntington, WV, [2]Dept. of Genetics and Evolutionary Biology, Univ. of São Paulo, São Paulo, SP, Brazil, [3]Dept. of Pharmacology and Physiology, Faculty of Med., Université de Montréal, Montreal, QC, Canada, [4]Div. of Aging, Brigham and Women's Hosp., Boston, MA, [5]Dept. of Biostatistics, Univ. of Washington, Seattle, WA, [6]Div. of Med. Genetics, Dept. of Med., Univ. of Washington, Seattle, WA

Essential hypertension (EH) is a major risk factor for cardiovascular diseases. EH is a complex disease characterized by elevated systolic blood pressure (SBP) and/or diastolic blood pressure (DBP). In Brazil, the reference values for hypertension are SBP ≥ 140 and/or DBP ≥ 90 mmHg, with 24.3% of the population reported to have hypertension. Global estimates for heritability of EH range from 15-60%. For Brazilian partially-isolated populations of African ancestry (Quilombo populations), the estimated heritabilities of SBP and DBP are 36.1% and 42.9%, respectively. To identify genetic variants contributing to increased EH susceptibility in these Quilombo populations, we conducted whole exome sequencing (WES) analysis on 26 samples from 3 large pedigrees (167 affected, 261 unaffected, and 3 unknown). WES was performed utilizing the IDT xGen Exome Hyb Panel v2, Illumina TruSeq Exome Kit, and Illumina Nextera DNA Exome protocols. WES data obtained from different panels were curated to retain common variants and trimmed to comprise the top 5 chromosomal regions of interest (ROIs) derived from our admixture-adjusted genome-wide linkage analyses results. In this study, we present the findings from our WES analysis assuming the "common disease-common variant" hypothesis. Variant filtering was performed from the annotated VCF file using the following criteria: (1) autosomal, (2) has alternate allele, (3) within exonic/splice site, and (5) allele frequency (AF) ≤ 0.2. Variant AFs were obtained from several datasets, including gnomAD, ExAC, ESP6500, and ABraOM. Synonymous variants were excluded. Our analysis revealed the presence of 57 missense variants in 30 genes within each of the identified ROIs: 17 variants (82.1% marker completion rate (mcr)) within ROI1, 9 variants (90.4% mcr) within ROI2, 6 variants (27.5% mcr) within ROI3, 16 variants (69.2% mcr) within ROI4, and 9 variants (55% mcr) within ROI5. Subsequently, we are investigating the top variants to explore potential interactions of the translated proteins and associated mutations with certain ligands. Our approach allows us to capture and analyze the exonic regions of the genome, which may provide valuable insights about genetic factors underlying EH in these populations. Moreover, these findings not only underscore the genetic diversity and complexity within these genomic regions but also emphasize their potential role in contributing to EH susceptibility. By shedding light on the intricate interplay between genetic variants and their translated proteins, our study deepens understanding of the underlying mechanisms involved in the etiology of EH.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session I

PB4430 Whole genome sequencing interaction study identifies loci with sex-specific effects underlying platelet aggregation traits

Authors:

J. Ngwa[1], M-H. Chen[2], L. R. Yanek[3], B. Gaynor[4], K. Ryan[4], K. Kanchan[5], K. Kammers[1], L. C. Becker[3], M. Taub[1], J. P. Lewis[4], A. D. Johnson[2], N. Faraday[3], R. Mathias[3], I. Ruczinski[1]; [1]Johns Hopkins Univ. Bloomberg Sch. of Publ. Hlth., Baltimore, MD, [2]NHLBI Population Sci. Branch, Framingham Heart Study, Framingham, MA, [3]Johns Hopkins Univ. Sch. of Med., Baltimore, MD, [4]Univ. of Maryland Sch. of Med., Baltimore, MD, [5]Johns Hopkins Univ., Sch. of Med., Baltimore, MD

Platelets play a key role in both hemostasis and thrombosis, and in inflammation, atherogenesis, and cancer metastasis. Whole-genome sequencing (WGS) based association studies have successfully identified several single nucleotide polymorphisms (SNPs) associated with platelet aggregation phenotypes. However, despite strong differences in platelet aggregation between males and females, a comprehensive analysis to elucidate sex-specific genetic effects has not been conducted. In this study, we used sequencing data available through the NHLBI's Trans-Omics for Precision Medicine (TOPMed) program to analyze 19 harmonized platelet traits, in response to using ADP, epinephrine or collagen, in three family-based studies: the AMISH cohort (n=255 participants), GeneSTAR (n=857 participants of European and n=683 participants of African ancestry) and the Framingham Heart Study (n=1476 participants). Linear mixed effects models, as implemented in the GENESIS Bioconductor package, including an interaction term for bi-allelic SNPs and sex were conducted separately for each study using inverse normalized age- and sex- adjusted residuals of the platelet traits as dependent variables, followed by a fixed-effects inverse-variance weighted meta-analysis using METAL. In the primary analysis using a 1 degree of freedom test for the SNP by sex interaction we identified ten SNPs on chromosome 10 below the conventional genome wide significance threshold of $5x10^{-8}$ for two epinephrine aggregation phenotypes. The chromosome 10 peak SNP (rs116725046, $p=5.2x10^{-9}$), located in the long intergenic non-protein coding RNA gene LINC00702, had a mean minor allele frequency (MAF) of 1.5% across all studies. Association of the closest protein coding gene KLF6 with platelet phenotypes was previously reported in the GWAS catalogue, however an in-silico functional analysis based on RNA expression and chromatin accessibility implicates LINC00702 as the regulatory element. In addition, several loci on chromosomes 2, 3, 7 and 11 also showed significant interactions with sex in the genome-wide meta-analysis for epinephrine and collagen. A more targeted analysis also showed sex specific effects for a variant (rs17081713) in the Estrogen Receptor 1 (ESR1) gene on chromosome 6 for ADP aggregation traits. These results provide strong evidence for the existence of sex specific genotype effects for platelet aggregation phenotypes.

Session Title: Statistical Genetics and Genetic Epidemiology Poster Session II

PB4431 X chromosome association study of asthma

Authors:

A. Eslami[1,2], M. Deneus[1], Z. Li[1], N. Gaudreault[1], S. Thériault[1,3], Y. Bossé[1,4]; [1]Inst. universitaire de cardiologie et de pneumologie de Québec, Quebec, QC, Canada, [2]Dept. of Social and Preventive Med., Laval Univ., Qubec, QC, Canada, [3]Dept. of Molecular Biology, Med. Biochemistry and Pathology, Laval Univ., Quebec, QC, Canada, [4]Dept. of Molecular Med., Laval Univ., Quebec, QC, Canada

Context: Asthma is a common respiratory disease with both genetic and environmental risk factors. Heritability estimates for asthma range from 0.55 to 0.90. Numerous genome-wide association studies (GWASs) have been completed and have identified several single nucleotide polymorphisms (SNPs) associated with asthma. However, because of the difficulty in analyzing the X chromosome, these published asthma GWASs are mostly focused on autosomal variants (chromosomes 1-22, non-sex chromosomes) and ignore the sex chromosomes. The prevalence and course of asthma is known to be sex-specific. It is thus important to take into account the effect of the X chromosome in genetic studies. **Aim**: Our research aims to perform X chromosome ssociation in the Quebec City Case-Control Asthma Cohort (QCCCAC) which consists of 1,618 French-Canadian subjects (1,089 asthmatics [660 women and 429 men] and 519 healthy controls [330 women and 199 men]). **Methodology**: Genotyping performed using the Illumina Global Screening Array BeadChip. After filtering and imputation (TOPMed reference panel) , we tested more than one million variants in chromosome X for association analysis using SAIGE (Scalable and Accurate Implementation of GEneralized mixed model) adjusting for age, sex, and the first 20 ancestry-based principal components. **Results**: Five independent variants were identified at a suggestive association threshold of *P*-value less than 1×10^{-5} in the X chromosome analysis. The most significant variant is located in *NHS* gene. The allele frequencies in asthma cases and controls were 0.017 and 0.055, respectively. **Conclusion:** Performing genetic association analysis on the X chromosome is a valuable tool for researchers to gain a better understanding of complex diseases. We will compare several statistical methods to study the association in chromosome X.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session I

PB4432 "Points to Consider" for International Genomic Data Sharing in the Health Technologies Industries

Authors:

B. Knoppers[1], S. Chase[2], Y. Joly[3], A. Thorogood[4]; [1]McGill Univ., Westmount, QC, Canada, [2]Illumina, La Jolla, CA, [3]McGill Univ., Montreal, QC, Canada, [4]Terry Fox Res. Inst., Montréal, QC, Canada

Health Technologies Industries (HTI) contribute unique resources, technologies, and expertise to translating genomic discoveries into improvements in human health. International sharing of human genomic and health data accelerates research and innovation by HTI, strengthening statistical power and reproducibility and facilitating collaboration and creative data re-use. Yet, data sharing raises important ethical issues, such as welfare and privacy risks, and evolving norms of data protection law, AI governance, research ethics, and data sharing.

A "Points to Consider" (PtC) for responsible sharing of genomic and related-health data was developed by the Centre of Genomics and Policy at McGill University in partnership with an Industry Core Group (participant members) from 10 companies. The PtC are tailored for HTI, building on the policies of the Global Alliance for Genomics and Health, namely the Framework for Responsible Sharing of Genomic and Health-related Data.

The Framework aims to activate the human right of everyone to benefit from scientific advancement and its applications, while also recognizing the moral and material interests resulting from any scientific production. The PtC were developed through: a systematic review of legislation, guidelines, and industry practices; 8 international, thematic briefs spanning from identifiability and consent to governance and intellectual property; and a series of consensus discussions.

The PtC aim to be concise and flexible, considering the diversity of HTI sectors, e.g., DTC, pharmaceutics, sequencing platforms, AI, and clinical genomics, as well as data sharing across jurisdictions. The PtC include the following:

- Promoting transparency in consent, data sharing, use of AI, and any return of results. - Assessing and mitigating identifiability risks, while preserving utility, quality and representativeness. - Conducting cross-border transfer impact assessments and employing data localization strategies to support compliance with local norms. - Ensuring data are kept securely, respect the FAIR principles, and are representative. - Balancing data control for legitimate commercial interests with data sharing for the public good ("responsible use"). - Establishing transparent partnerships with academia that respect scientific quality and freedom. - Meaningfully engaging with communities in the development of data sharing plans.

These "Points to Consider" for international genomic data sharing are intended for voluntary adoption by HTI, but can be customized into more prescriptive sectoral best practices or enforceable codes of conduct.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session II

PB4433 "Think Rare": Harnessing electronic medical record data to shorten the rare disease diagnostic odyssey

Authors:

C. Chisholm[1], A. White-Brown[1,2], G. U. Ediae[3], J. Seymour[3], I. Terekhov[3], N. Mitsakakis[3], S. L. Sawyer[3,1], K. M. Boycott[3,1]; [1]CHEO, Ottawa, ON, Canada, [2]McGill Univ., Montreal, QC, Canada, [3]CHEO Res. Inst., Ottawa, ON, Canada

Rare genetic diseases (RGD) are often unrecognized by non-genetics health care providers, which can delay a patient's referral to a genetics clinic for appropriate evaluation and testing. These delays contribute to the diagnostic odyssey experienced by many patients with RGD. The recent availability of clinical genome-wide sequencing within Ontario has increased access to testing, but as ordering is limited to genetics providers, the current system still relies on non-genetics providers to suspect RGD and refer patients to a genetics clinic. We aimed to address this barrier to referral by developing a tool to search an electronic medical record (EMR) for patients with undiagnosed RGD.
In this retrospective study, we utilized our Epic EMR and the data warehouse at the Children's Hospital of Eastern Ontario to access the clinical health data already stored. Clinical judgement and data availability were used to identify data elements that would be suggestive of RGD: specific diagnostic codes as well as frequency of appointments with multiple subspecialty departments. The selected data elements were used in a rule-based system, to create a search algorithm (named ThinkRare) that was specific to patients with RGD. To evaluate the sensitivity, the algorithm was applied to patients previously evaluated in the genetics clinic known to meet the Ontario genome-wide sequencing eligibility criteria ("gold standard" cohort). The algorithm was then optimized through iteratively testing each version on a test dataset (patients not previously seen in genetics). A manual chart review was completed to confirm if selected patients were a true positive (appropriate selection) or false positive (inappropriate selection).
The algorithm had a sensitivity of 84% in the "gold standard" cohort (n=438). The algorithm was then applied to 262,296 unique patients, and identified 355 patients with data elements consistent with having RGD. Manual chart review revealed that 184 (52%) of those patients would be considered appropriate for a referral to the genetics clinic for evaluation and potential testing; the remaining 48% of patients had a non-genetics explanation for their phenotype. The next steps of this work include clinical evaluations for the 184 patients identified by the ThinkRare algorithm (summer 2023), further algorithm adaptations to improve specificity (ongoing), and ultimately creation of a real-time alert in the Epic EMR for providers to consider a referral to genetics if their patient meets the algorithm criteria (future). The algorithm shows promise in identifying patients suspected to have a RGD and ultimately help providers "ThinkRare".

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session III

PB4434 † 20k T1d Exomes Cohort In Ukraine

Authors:

T. Oleksyk, W. W. Wolfsberger, K. Shchubelka; Oakland Univ. - Biology, Rochester, MI

Progress has been made in understanding how common genetic variants contribute to the risk of developing type 1 diabetes or T1D. Rare and local variants of low frequency are missed in global analyses but can be strong determinants of disease. In general, rare variants, as a group, have the potential to offer public health insights, explain disease onset and incidence, and represent an under-tapped reservoir of human-based drug discovery. We already collected and sequenced exomes of 3,000 type T1D patients and 3,000 matching controls in Ukraine, and conducted a genome-wide association study with the goal of discovering new genetic factors that affect the development and pathology which could help predict disease, identify therapies to delay progression, and/or improve the life those living with T1D in the region. We describe the progress of this work in the context of the current state of availability of genome data from Eastern Europe, the progress of our research from building a map of genetic diversity in Ukraine and bordering countries, explore ethical and logistic considerations of collecting DNA and consider challenges of big data analysis and sharing.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session I

PB4435 A compendiumof manually annotated genetic variants for Alkaptonuria-AKUHub

Authors:

V. Radha Krishnan[1], A. Shanmuga Sundaram[1], S. Suresh[1], S. SOUCERADJOU[1], A. Chandramohan[1], R. Mehta[2], J. Jagamoorthy[1]; [1]Zifo Rnd Solutions, Chennai, India, [2]Zifo RnD Solutions, Chennai, India

Alkaptonuria or 'black urine disease' is a rare autosomal recessive disorder caused by dysfunctional homogentisate 1,2-dioxygenase (HGD) gene (3q13.33) leading to accumulation of homogentisic acid in the body. This inborn error in metabolism of phenylalanine and tyrosine causes accumulation of homogentisic acid leading to ochronosis, pigmentation in the sclera, ear cartilage, mitral valve calcification and osteoarthropathy. Advances in sequencing technologies have helped us to map genetic variants associated with alkaptonuria in diverse populations and regions. Currently, no centralized resource of all the reported actionable variants with uniformity in annotation exists for the HGD gene. We have compiled HGD exonic variants from various data sources and systematically annotated their pathogenicity according to American College of Medical Genetics and the Association of Molecular Pathologists (ACMG-AMP) variant classification framework. A total of 1686 exonic variants were catalogued and manually curated, creating one of the most comprehensive Alkaptonuria variant databases (AKUHub) which is publicly available.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session II

PB4436 A database of genes and mutations associated with *chlamydia species* drug resistance.

Authors:

I. Benamri[1,2], A. Moussa[1], F. Radouani[2]; [1]Natl. Sch. of Applied Sci. of Tangier, Abdelmalek Essaâdi Univ., Tangier, Morocco, [2]Inst. Pasteur du Maroc, Casablanca, Morocco

Chlamydiae are intracellular bacteria responsible for redoubtable diseases in both human and animals. The commonly used drugs against *Chlamydiae* infections are antibiotics. However, the misuse and overuse of antimicrobial agents have resulted in an increase in the appearance of bacterial drug resistance, this phenomenon represents a growing public health problem worldwide. The main goal of this work is to implement a database for genes and mutations associated with *chlamydia species* resistance to drugs. This resource will be a user-friendly support to increase awareness of chlamydia species resistance to antibiotics, and provide support for potential inter investigator comparative studies. After a literature search targeting the collection of studies, we proceeded to a manual curation for all the papers gathered in order to extract the information needed to set up our database. Actually, our proposed database is presented as a web-based user-friendly interface to browse, search and visualise all the genes and mutations associated with *chlamydia species* resistance to drugs, in addition, it provides comprehensive information about *chlamydia species* and the genetic factors causing drug resistance. Hence, the users can easily retrieve and download all requested information.The developed portal is the first resource that gathers genes and mutations associated with chlamydia species drug resistance, it will be a relevant resource for drug resistance exploration based on bacterial genetic information. It will help investigators and physicians to improve the healthcare quality and infected patient’s therapeutic surveillance, to decrease the longer hospital stays, the higher medical costs and decrease the disease's complications. Finally, the database will be periodically updated to improve its services.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session III

PB4437 A deep population reference panel of tandem repeat variation

Authors:

H. Ziaei Jam[1], Y. Li[1], R. Devito[1], N. Mousavi[1], N. Ma[1], L. Ibra[2], Y. Adam[3], M. Maksimov[1], B. Huang[1], E. Dolzhenko[4], Y. Qiu[4], F. Kakembo[2], H. Joseph[2], B. Onyido[3], J. Adeyemi[3], M. Bakhtiari[1], J. Park[1], S. Javadzadeh[1], D. Jjingo[2], E. Adebiyi[3], V. Bafna[1], M. Gymrek[1]; [1]Univ. of California San Diego, La Jolla, CA, [2]Makerere Univ., Kampala, Uganda, [3]Covenant Univ., Ota, Nigeria, [4]Illumina Inc., San Diego, CA

Tandem repeats (TRs) are elements in the human genome that consist of repeated sequences of motifs, typically one or more nucleotides, and exhibit high variability. The human genome contains approximately 2 million distinct TR loci. Despite their abundance, TRs have proven challenging to study extensively due to the technical difficulties associated with sequencing and genotyping. Several tools, such as HipSTR, GangSTR, adVNTR, and ExpansionHunter, have been developed to genotype TRs using short-read sequencing data. These tools have been utilized to investigate TR variation in population-scale studies involving multiple cohorts. However, the existing methods predominantly capture different sets of TRs, and none encompass the entire range of TR variation. Moreover, the TR callsets generated from these methods have shown a significant bias toward individuals of European ancestry.

In this study, we introduce EnsembleTR, a graph-based approach that combines genotypes from various TR calling methods to produce consensus genotypes for each TR. We utilized EnsembleTR to analyze high-coverage whole genome sequencing data from 3202 samples from the 1000 Genomes Project and 348 samples from the H3Africa Project. As a result, we created a comprehensive reference panel of TR variation across diverse global populations, encompassing approximately 1.7 million TRs. To validate our findings, we conducted fragment analysis on 48 TRs across 31 samples and observed a remarkable concordance rate exceeding 98% when compared with the genotypes obtained using EnsembleTR.

Using the improved data from EnsembleTR, we studied patterns of TR variation across populations. Our catalog recapitulates known patterns, such as higher variation rates in African populations and the relationship between repeat length/motif and heterozygosity rates. We also uncover novel findings, including (1) context sequence features in the immediate flanking regions of repeats influencing TR stability, (2) population-specific repeat expansions, including a CAG repeat in intron 4 of CA10 and a TTC repeat upstream of NEXN that both are specifically expanded in African populations and (3) TRs associated with gene expression. Finally, we generated a phased reference panel that allows accurate imputation of TR variants from SNP data. Our results and analysis provide a genome-wide TR catalog as well as population-level TR characterizations that will open the path for further studies.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session I

PB4438 A familial, telomere-to-telomere (T2T) reference genome for human variation, recombination, and mutation

Authors:

D. Porubsky[1], G. A. Logsdon[1], N. Koundinya[1], W. T. Harvey[1], M. Noyes[1], Z. N. Kronenberg[2], K. M. Munson[1], K. Hoekzema[1], J. Knuth[1], G. Garcia[1], B. S. Pedersen[3], H. Dashnow[3], C. Nolan[2], E. Dolzhenko[2], C. Fanslow[2], C. Lambert[2], T. Mokveld[2], D. W. Neklason[4], W. S. Watkins[3], A. R. Quinlan[3], L. B. Jorde[3], M. A. Eberle[2], E. E. Eichler[1,5]; [1]Univ. of Washington, Dept. of Genome Sci., Seattle, WA, [2]PacBio, Menlo Park, CA, [3]Univ. of Utah, Dept. of Human Genetics, Salt Lake City, UT, [4]Univ of Utah, Dept. of Internal Med., Salt Lake City, UT, [5]Univ. of Washington, Howard Hughes Med. Inst., Seattle, WA

Telomere-to-telomere (T2T) chromosome assemblies generated from diploid samples provide access to complete genetic information, allowing us to revisit fundamental properties of mutation, transmission, and genetic recombination. We have generated deep Pacific Biosciences (PacBio) high-fidelity (HiFi), ultra-long Oxford Nanopore Technologies (ONT), Strand-seq, and Illumina whole-genome sequencing data to construct near-T2T, phased genome assemblies from primary material obtained from a 4-generation (G1-G4), 21-member CEPH reference family (1463). Using Verkko, a hybrid genome assembly pipeline, we created reference genome-grade assemblies that are highly accurate (QV>50) and contiguous (N50 >100 Mbp), resolving both the paternal and maternal haplotypes, including complete centromeres and complex regions of segmental duplications (SDs). We are constructing a comprehensive and validated catalog of >8 million single-nucleotide variants, indels, short tandem repeats, and structural variants for each haplotype from each member and assessing transmission characteristics, including a detailed assessment of inversion polymorphisms that associate with disease risk. In total, we have detected 121 inversions segregating in this family, of which 61 are larger than 50 kbp and flanked by SDs. We discover 72-110 de novo mutations per transmission and show that centromeres are remarkably stable when transmitted from generation to generation. We also sequence resolve the first de novo structural mutation of centromeres, consistent with higher-order repeat (HOR)-mediated changes of α-satellite DNA. We constructed a recombination map across the family and have identified 651 crossovers in G2 and G3 at high resolution (~17 kbp). Using Strand-seq, we are resolving the remaining gaps including complete resolution of the acrocentric arms and their mutation from generation to generation. The use of multiple orthogonal technologies, near-T2T phased-genome assemblies, and a multi-generation family to assess transmission as well epigenetic changes has the potential to create a “truth set” for all classes of human genetic variation upon which to test and benchmark new technologies and understand the most fundamental processes underlying human genetic variation.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session II

PB4439 † A Gene EXPRESSION RESOURCE TO EVALUATE HUMAN LIVER CELL LINES AS MODELS FOR HUMAN LIVER BIOLOGY

Authors:

B. Halligan[1], Y. Chen[2], A. Oliveri[1], M. Tohme[1], E. Speliotes[1]; [1]Univ. of Michigan, Ann Arbor, MI, [2]Univ. of Michigan-Ann Arbor, Saline, MI

Background: Liver disease may affect up to 54% of individuals in some populations. Liver disease is challenging to model in cell model. Primary hepatocytes can be difficult to obtain and dedifferentiate quickly in cell culture. Human hepatocellular carcinoma (HCC) lines can be easily propagated but are often aneuploid and important genes or biological pathways of interest may not be expressed. To aid in determining which cell lines may best human cell biology we used RNAseq to measure gene expression and compared these to normal human primary heptocytes. We also determined exonic variation in these cell lines. We created a new software application, CorrPaths, that reports gene expression and exonic variation across these cell lines to facilitate identifying cell lines suitable for modeling desired aspects of liver biology and diseases. **Methods:** RNAseq was performed on six primary hepatocyte datasets from GEO and compared with RNA seq from HepG2, HepG2-C3A, HuH-7, SNU-475 and THLE-1 cell lines. Gene expression levels for HCC lines were determined using TopHat and DESeq2 and compared to primary hepatocytes with CorrPaths. Select gene expression results were confirmed with Western blotting. Exome sequences from HCC lines were mapped to reference human genome sequences using GATK. **Results:** We found that many of the P450 cytochromes expressed by primary hepatocytes (CYP2E1, CYP2C8, CYP2C9, CYP3A4, CYP2A6) are nearly completely absent in all cell lines. Interestingly, we found all cell lines express the muscle form of glycogen synthase (GYS1) rather than the liver form of glycogen synthase (GYS2). This result was confirmed by Western blotting of GYS1 and GYS2 in HuH-7 cells compared with liver tissue extracts. Genotyping HuH-7 and HepG2 shows that both HuH-7 and HepG2 cell lines are homozygous for the PNPLA3 SNP (rs738409; I148M) known to increase NAFLD making them good models for this disease. The liver specific output from CorrPaths, LiverCellPath, can be used online or offline. It produces an Excel file that can be easily shared with other users, exported to HTML, and can be encrypted for use with sensitive data. **Conclusion:** Choosing the correct model biological system is critical for functional experimentation. Using LiverCellPath as an experimental design tool can guide users to chose a cell line suitable for their biology based on gene expression and exonic variation.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session III

PB4440 A genome-first approach to cancer phenotypes, penetrance and survival rates of individuals who harbor concurrent pathogenic/likely pathogenic variants in*CHEK2*and other cancer predisposition genes.

Authors:

S. Kim[1,2], M. Ramos[2], J. Kim[2], J. Haley[3], H. Rao[3], U. Mirshahi[3], Discover EHR collaboration, B. Graubard[2], H. Katki[2], D. Carey[3], D. Stewart[2]; [1]Johns Hopkins Med. Inst., Baltimore, MD, [2]Natl. Cancer Inst., Rockville, MD, [3]Geisinger, Danville, PA

Introduction: Using two population-scale exome-sequenced cohorts linked to the electronic health record, we investigated the prevalence and risk of cancer in individuals who harbor a germline pathogenic/likely pathogenic (P/LP) variation in *CHEK2* plus at least one of 14 other cancer-predisposition genes. **Methods:** From Geisinger and UK Biobank (UKBB) (age>18 yr, Geisinger: n=167,060, UKBB: n=469,681), we investigated cancer phenotypes in individuals with concurrent P/LP variants in *CHEK2* and 14 other cancer predisposition genes: *ATM, BARD1, BRCA1, BRCA2, CDH1, MSH6, NF1, PALB2, PIK3CA, PTEN, RAD51C, RAD51D, RECQL* and *TP53*. Electronic health records, cancer, and death registries documented with ICD9 and ICD-10/ICD10-CM codes were used to retrieve cancer phenotypes. The risk of cancers in individuals carrying concurrent P/LP variants in *CHEK2* and 14 other cancer predisposition genes (Group A) were compared with that of total *CHEK2* heterozygotes (Group B) as well as noncarriers. Demographic variables, including age, sex, smoking, alcohol, and body mass index (BMI) were examined as possible confounders using logistic regression. Kaplan-Meier survival curves for survival analyses and reverse Kaplan-Meier curves for age-dependent cancer penetrance were constructed, and hazard ratios were computed based on the Cox Proportional-Hazards model. **Results:** The numbers of individuals were as follows: Group A (Geisinger: n=61, 0.04%, UKBB: n=34, 0.007%) and Group B (Geisinger: n=3,086, 1.9%, UKBB: n=3,192, 0.68%). In UKBB, cancer prevalence was higher in Group A compared to that of Group B (OR:2.32, CI: 1.13-4.75, *P*=0.022), and age-dependent cancer penetrance was significantly increased in Group A versus Group B (*P*=0.011). These observations were not replicated in the Geisinger cohort. Overall survival rate and cancer-related mortality were not significantly different between groups A and B in UKBB and Geisinger. **Conclusions:** Our results of the genome-first approach revealed that individuals who harbor concurrent P/LP variants in *CHEK2* and at least one of 14 other cancer predisposition genes may have an increased risk of cancer and age-dependent penetrance compared to those of *CHEK2* heterozygotes. We anticipate that these results will inform cancer surveillance and genetic counseling.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session I

PB4441 A harmonized public resource of deeply sequenced diverse human genomes.

Authors:

M. Yohannes[1,2], Z. Koenig[1,2], L. Nkambule[1,2], J. Goodrich[1,2], H. Kim[1], X. Zhao[1,2,3], M. Wilson[1], G. Tiao[1], S. Hao[1,2,3], N. Sahakian[1], K. Chao[1], gnomAD Project Consortium, M. Talkowski[1,2,3], M. Daly[1,2,4], H. Brand[1,2,3], K. Karczewski[1,2], E. Atkinson[1,2,5], A. Martin[1,2]; [1]The Broad Inst. of MIT and Harvard, Cambridge, MA, [2]Massachusetts Gen. Hosp., Boston, MA, [3]Massachusetts Gen. Hosp. and Harvard Med. Sch., Boston, MA, [4]Inst. for Molecular Med. Finland, Helsinki, Finland, [5]Baylor Coll. of Med., Houston, TX

Underrepresented populations are often excluded from genomic studies due to limited resources supporting their analysis. The 1000 Genomes Project (1kGP) and Human Genome Diversity Project (HGDP), which have recently been sequenced to high coverage, are valuable genomic resources that capture global diversity and embrace open data-sharing policies. To harness the potential of these resources, we harmonized a set of 4,096 whole genomes from HGDP and 1kGP. We identified over 155 million high-quality single nucleotide variants (SNVs), insertions/deletions, and structural variants, revealing more variants compared to existing releases as a result of joint variant calling.

To explore ancestry compositions within geographical and genetic regions, we used global and subcontinental principal component analysis (PCA) and ADMIXTURE analysis. The SNV loadings were made freely available to enable the projection of user cohorts onto a unified reference PC space, allowing for ancestry modeling and inference. Additionally, we measured population genetic differentiation using common variants with Wright's fixation index, F_{ST}, while comparing rare variant sharing through pairwise doubleton counts using f_2 analysis. Comparisons of these divergence metrics provide insights into different historical time points, i.e. older population history with F_{ST} and more recent history with f_2. As expected, we identified more variants among individuals of African descent. Populations also clustered by geographical/genetic regions.

We also examined the relationship between F_{ST} and geographical distance based on great circle distances using the haversine formula, and pairwise geographic distances using five waypoints that reflect human migration patterns. The linear relationship between F_{ST} and distance was different between HGDP and 1kGP. The steeper slope observed in HGDP likely reflected the anthropological design intended to capture more divergent populations, whereas 1kGP represented some of the largest populations.

Importantly, we have publicly released this data without restriction. To facilitate community uptake of this invaluable genomics resource, we have also created two key resources for the genomics community: 1) a resource of haplotypes for phasing and imputation, and 2) detailed tutorials to facilitate common genomic analyses. This jointly called reference panel together with our detailed tutorials will empower genetic studies in diverse populations by equipping researchers with the necessary resources and tools that facilitate scalable and efficient analyses.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session II

PB4442 A large-scale genomic survey of coding sequence variation in rhesus macaques reveals substantial numbers of predicted damaging variants in disease-related genes.

Authors:

J. Rogers[1], M. Raveendran[1], R. A. Harris[1], H. Doddapaneni[1], D. Muzny[2]; [1]Baylor Coll. of Med., Houston, TX, [2]Baylor Coll. Med., Houston, TX

Rhesus macaques (*Macaca mulatta*) are the most widely used nonhuman primate models in biomedical research. This species is a well-documented and valuable model of infectious diseases such as HIV-AIDS, SARS-CoV-2, tuberculosis and others. In addition, rhesus macaques are widely studied as models of neurodevelopment, anxiety and affective disorders, diabetes, obesity, cognitive aging, reproductive biology, endometriosis and other human clinical disorders. We have generated deep whole genome or whole exome sequence data from 2,189 rhesus macaques bred in research colonies from more than 15 different US research centers. The dataset consists of 2,006 Indian-origin animals and 183 Chinese-origin animals. We mapped Illumina short read sequence data to the rhesus reference genome (Mmul_10) and called SNVs and small indels ($\leq$6 bp) using GATK and standard quality filters. The average read coverage for called variants was 42.7x. The Indian-origin animals exhibited 223,142 distinct missense, 3,618 stop gained and 778 splice donor or splice acceptor variants. The average number of SNVs analyzed per Indian-origin rhesus was 51,458. The total missense variants identified among Chinese-origin animals numbered 248,776, with 2,932 stop gained and 656 splice donor/acceptor variants. We mapped the macaque variants over to the hg38 reference genome and used CADD and ClinVar databases to assess the likely functional impact of each variant. There are 42 macaque variants observed more than once across the Indian-origin cohort and scored as "pathogenic" in ClinVar, including mutations that are known in humans to cause multiple myeloma, neurodevelopmental disorder, breast cancer, hypercholesterolemia and others. All the 42 variants have CADD scores of >22. The Chinese-origin rhesus macaques had 17 variants identified in ClinVar as pathogenic, with effects on kidney disease, adrenal disorder and intellectual development disorder among others. Overall, the Chinese-origin animals carry a similar array of predicted pathogenic variants to the Indian-origin animals but represent a smaller proportion of animals available for subsequent research. This dataset is the largest genomic survey reported to date for any laboratory nonhuman primate species. Our results illustrate the potential value of rhesus macaques as spontaneous models for the study of a wide diversity of human genetic diseases. Prior surveys by our lab and others of a smaller number of animals have led to the establishment of macaque models of disease that are currently being used to test novel therapies. These expanded data point towards further opportunities for similar pre-clinical studies.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session III

PB4443 A novel genomics analysis pipeline identifies mitochondrial somatic variants.

Authors:

K. Patel, J. Kwan, A. Jha; Yale, New Haven, CT

Background: Mutations in mitochondrial DNA (mtDNA) can impair mitochondrial function and can lead to a range of diseases, including those impacting cardiovascular function. The accumulation of somatic mitochondrial variants also contributes to cardiovascular disease but determining somatic mitochondrial variants remains a challenge.
Research question: We sought to evaluate which variant calling algorithms or combinations of them would be able to call somatic mitochondrial variants.
Methods: We used four different variant calling algorithms to extract mtDNA variants from input genomes from the 1000 Genomes Project. Realignment was performed on the patient genetic sample to align their input genome to the "Genome Reference Consortium Human Build 38" (HG38) reference genome. Four algorithms were used to detect mitochondrial variants: Mutect2, DeepVariant, FreeBayes with BCFtools, and Mutserve.
Results: The results show that Google's DeepVariant and GATK Mutect2 are the most effective algorithms for detecting mitochondrial variants. FreeBayes with BCFtools failed to detect any mitochondrial variants. DeepVariant identified a total of 17 variants, 12 of which had a quality score greater than 30. Mutect2 detected the greatest number of mitochondrial variants with 407 variants detected. Mutect2 found all but one variant that DeepVariant detected.
Conclusions: Although the variants we found came from a healthy patient population, we were able to identify mitochondrial somatic variants. We are currently using machine learning methods to create a customized machine learning method and apply this on larger patient cohorts like the UK Biobank and TOPMed to detect somatic mitochondrial variants that are associated with CVD and can predict outcomes.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session I

PB4444 A novel mouse model of uveal coloboma arises from genetic rearrangement on Chr13.

Authors:

Y. Wang[1], R. P. Alur[1], V. Kalaskar[1], A. George[1], E. Boobalan[1], A. Dutra[2], E. Pak[2], J. Thomas[2], L. Dong[1], M. Fruttiger[3], S. Pieke-Dahl[4], P. Hitchcock[4], B. Brooks[1]; [1]Natl. Eye Inst., Bethesda, MD, [2]NHGRI, Bethesda, MD, [3]Univ. Coll. London, London, United Kingdom, [4]Univ. of Michigan, Ann Arbor, MI

Purpose: Uveal coloboma is a congenital ocular malformation, caused by failure of the optic fissure to close at the 5th week of human gestation and at embryonic 11.5 days (E11.5) in the mouse. Currently, several genetic mutations that cause the disease have been identified. However, the genetic mechanisms by which the chromosomal rearrangement instigates optic fissure closure defects are poorly understood. Here we found that random insertion of a transgene (NSE-hVEGF) on chromosome 13 (Chr.13) of C57BL/6J mice generated an obviously retinal and iris coloboma (RICO). The region of insertion did not suggest any previously identified human coloboma genes. Therefore, it represents a novel dominant animal model of uveal coloboma. **Methods:** RICO mouse was generated by random insertion of transgene human VEGF driven by rat Neuron Specific Enolase (NSE) promotor on Chr.13. RICO mouse and embryos were analyzed via detailed histopathology. The insertion site was evaluated with the combination of Fluorescence In Situ Hybridization (FISH), Comparative Genomic Hybridization (CGH), BAC library construction, and whole-genome shotgun sequencing of RICO genomic DNA. The hVEGF RNA expression was detected by Basescope, and the inversional genes expression was evaluated by qRT-PCR at E11.5. **Results:** Transmission of RICO mice was autosomal dominant with nearly 100% penetrance. The uveal coloboma in mice are usually bilateral, affecting the retina, iris and optic nerve. FISH assay indicated the insertion site of the transgene at mouse Chr.13. Shotgun Sequencing analysis revealed that the chromosomal rearrangement occurred specifically between two distinct loci (4 Mb apart) in Chr.13, and about 20 copies of hVEGF transgene was inserted. The region of insertion did not suggest any previously identified human coloboma genes. CGH array shows no large duplications or deletions on Chr.13 after transgenic insertion. qRT-PCR results showed that a few development-related genes were altered in rearranged region. Basescope assay observed the specific hVEGF RNA expression in the margin of optical fissure at E11.5 of RICO mouse. Interestingly, insertion of hVEGF transgene into other chromosomes of independent lines of transgenic mice did not generate the RICO phenotype. **Conclusions:** Insertion of the hVEGF transgene into Chr.13 alter the genetic rearrangement and generate classic uveal coloboma in C57BL/6J mice. This transmission is autosomal dominant with nearly 100% penetrance. This result provides a new animal model to further understand the genetic and molecular mechanism of uveal coloboma.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session II

PB4445 Accelerated workflows for analyzing long-read sequenced genomes in the NIH All of Us Program

Authors:

E. Dawson[1], L. Slim[1], L. Paulin[2], F. J. Sedlazeck[2,3,4]; [1]NVIDIA, Santa Clara, CA, [2]Human Genome Sequencing Ctr., Baylor Coll. of Med., Houston, TX, [3]Human Genetics Ctr., Sch. of Publ. Hlth., Univ. of Texas Hlth.Sci. Ctr., Houston, TX, [4]Dept. of Computer Sci., Rice Univ., Houston, TX

The All of Us Project is a massive national sequencing initiative to gather health data from one million individuals in the US, with a focus on improving representation within genomic studies. This includes nearly 25,000 genomes to be sequenced using the Oxford Nanopore Promethion long read sequencing platform. To support the analysis of this data, we present a set of GPU-accelerated workflows that can be deployed locally or in cloud bioinformatics platforms. These workflows support basecalling, alignment of reads as well as variant calling of SNVs, indels and structural variants and phasing of small variants using a combination of best-practice tools. By incorporating a series of accelerated tools, we provide best-in-class performance and accuracy.

Leveraging these workflows in the cloud significantly increases the throughput of analyzing many genomes while also improving ease of use and the comprehensiveness across all variant classes. We demonstrate the accuracy of our workflows on the Genome-In-A-Bottle HG002 sample and provide runtime and cost numbers for a variety of depths of ONT data.

Standardized workflows provide an important community resource for reproducibility and a basis for further development. Cost-optimized workflows also make such analysis more accessible. These workflows can be adopted for Oxford Nanopore or PacBio data and scale from small to large labs. Our open-source workflows provide a best-practice starting point for long read studies and demonstrate the power of GPUs and workflows to democratize access to cutting edge bioinformatics analysis. They represent an approach for unifying and harmonizing variant calling across multiple larger consortia, generating a valuable resource for later studies relying on variant comparisons and allele frequency data. Furthermore, this harmonized workflow enables population merging of samples across ethnicities and as such represents a significant step forward for standardization for potential imputation servers to leverage the obtained long read data.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session III

PB4446 An automated quality control system for the biochemical genetics laboratory.

Authors:

P. Tanpaiboon[1], D. Salazar[2], L. T. Xu[1], R. Sharma[1]; [1]Quest Diagnostics, San Juan Capistrano, CA, [2]Quest Diagnostics Nichols Inst, San Juan Capistrano, CA

Background: Quality control (QC) in any clinical laboratory is critical to ensuring quality and accuracy of patient results. However, QC monitoring is complicated in the multi-analyte, multi-instrument assays common to biochemical genetics laboratories. An ineffective QC review can impact laboratory productivity, data accuracy, troubleshooting, audit preparedness, and client satisfaction. Comprehensive off-the-shelf QC management systems optimized for such highly complex assays and platforms are relatively scarce. A manual QC review process can impact laboratory productivity and increase risk of errors. In contrast, commercially available software may lack the ability to nimbly adapt to new assay requirements, which is crucial for large laboratories with comprehensive test menus and various assay platforms. Here we describe a novel software application that integrates, processes, and displays QC statistical parameters from multiple instruments in real time. Methods and Results: A customizable, cloud-based software application was developed to automate QC review and monitor assay performance. This QC program has modules for different assay platforms including complex mass spectrometry, chemistry analyzer, ELISA, and fluorescence-based plate reader assays. The program's functions include automated collection and analysis of assay data, Levey-Jennings charts with integrated data from multiple instruments-multiple HPLC channels, graphical data visualization, instrument data consolidation and centralization, assay monitoring, tracking of tests not performed, and a time-stamped QC audit trail. The program also generates automated email notifications for pending QC lot expiration, weekly/monthly review reminders, and alerts about abnormal patient results that enable prompt communication to providers. The software was developed and has been maintained by the Quest Diagnostics Bioinformatics team. The program was written in PHP and JavaScript using the Symfony 4 framework and Microsoft SQL Server, and hosted on internal servers via a Docker container instance. The servers are securely connected to Quest's network, and all users must have a security clearance from Quest before gaining access. There are several access levels to limit what a user can do.Conclusion: This program may improve assay quality and provide significant time savings (approximately 4 hours/day) by automating data entry and eliminating clerical errors. One benefit of this software is the ease of updating program capabilities and customizing them to meet specific and changing needs.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session I

PB4447 An imputation panel with disease-enriched variants for improving single-nucleotide and structure variants discovery on array data in Alzheimer's Disease

Authors:

P-l. Cheng[1], H. Wang[1], I. Horng[1], B. Dombroski[1], L-S. Wang[2], B. Vardarajan[3], G. Tosto[3], J. Farrell[4], B. Kunkle[5], G. Schellenberg[6], W-P. Lee[1]; [1]Univ. of Pennsylvania, Philadelphia, PA, [2]UPENN, Philadelphia, PA, [3]Columbia Univ., New York, NY, [4]Boston Univ Med, Laconia, NH, [5]Univ. of Miami, Miami, FL, [6]Univ Pennsylvania Sch Med, Philadelphia, PA

The discovery of structural variants (SVs) and short insertions/deletions (indels) is largely improved with the use of whole genome sequences (WGS) and has been demonstrated as a genetic factor associated with Alzheimer's disease (AD). Compared to WGS, genotype-array is cost-effective and has more data available than WGS for AD research. However, due to technology limitation, genotype-array is unlikely to support SV discovery. Although imputation could infer unobserved genotypes in a sample, current reference panels contain either none or a limited number of SVs, and thus disease-specific reference panels can outperform public reference panels when imputing disease causing variants. We first constructed an AD-specific panel including 51,459,037 single-nucleotide variants (SNVs) and 3,322,380 indels of 16,564 sequenced genomes from the Alzheimer's Disease Sequencing Project (ADSP). To assess the performance of our AD-specific panel, we performed imputation on 38,271 genotypes from Alzheimer's Disease Genetics Consortium (ADGC) and compared the imputation generated by our AD-specific panel with the one from TOPMed panel. Our AD-specific panel exhibited superior performance in rare variants. For example, rare exonic indels in ANO7, ZNF656, SCRT2, ABCA7, and PPDPF genes could only be imputed by our panel. Laboratory validation by PCR confirms the true imputation of these indels. To leverage the strengths of both the TOPMed and our AD-specific panels, we employed meta-minimac2 to combine imputations. The result shows that while the TOPMed panel demonstrated better performance in imputing common SNVs, our AD-specific panel was able to capture the rare disease-specific linkage disequilibrium (LD) structure that was overlooked by publicly available reference panels. Furthermore, we demonstrated extending the knowledge of indels gained from WGS to genotype-array data. To impute SVs, we incorporated 210,429 high-quality SVs, which were detected on ADSP WGS samples, into the panel. By using this SNP-SV panel, we successfully imputed SVs, ranging from 57bp to 39kb, located within AD risk loci, such as BIN1, PILRA, PLCG2, ADAM10, ABCA7, and APP, as well as genes found significant in association tests, such as NCK2, WDR12, TMEM106B, JAZF1, PLEKHA1, TPCN1, IGH, and WNT3/MAPT. The imputation quality of these AD related SVs on 38,271 genotypes was excellent in with R2=0.80±0.17. The project could effectively increase sample size for GWAS studies, enable SV discovery in numerous array data of AD, and facilitate finding novel genetic associations of AD.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session II

PB4448 Analysis of targeted and whole genome sequencing of PacBio HiFi reads for a comprehensive genotyping of genic and phenotype associated Variable Number Tandem Repeats.

Authors:

S. Javadzadeh[1], A. Adamson[2], J. Park[1], S-Y. Jo[3], Y-C. Ding[2], M. Bakhtiari[1], S. Neuhausen[4], V. Bafna[5]; [1]Univ. of California San Diego, La Jolla, CA, [2]Beckman Res. Inst. of City of Hope, Duarte, CA, [3]Yonsei Univ. Coll. of Med., Seoul, Korea, Republic of, [4]Beckman Res. Inst. of City of Hope, Duarte, CO, [5]Univ. of California, San Diego, La Jolla, CA

Variable Number Tandem Repeats (VNTRs) have repeat motif lengths greater than 5 bp. They have been associated with Mendelian and complex disorders but their complex structures are challenging to genotype. Of 10,529 gene-proximal VNTRs, nearly 40% exceeded Illumina short-read length (140 bp). Similarly, 60% of 53 phenotype-associated VNTRs (P-VNTRs) were long, requiring long-reads for precise genotyping. We explored the trade-off in scope (how many VNTRs) and coverage between targeted and whole genome HiFi sequencing for genotyping long VNTRs. We designed customized probes for 10,529 VNTRs and sequenced 7 samples for PacBio HiFi and Illumina sequencing followed by adVNTR genotyping. We compared these results with HiFi whole genome sequencing (WGS) data from 28 samples in the Human Pangenome Reference Consortium (HPRC).
Only 3,981 (38%) VNTRs and 4 of 53 P-VNTRs were spanned with > 15 reads using targeted probes. All 10,529 VNTRs showed high consistency of read coverage across the 7 samples (coefficient of variation CoV = 0.08), implying that the VNTR sequence, rather than individual samples, dictated the number of spanning reads. A smaller subset of 2,876 (1,422) VNTRs had very high median coverage of at least 50 (100) spanning reads. Among 5,638 VNTRs with low-coverage (<15), 67% were located within GC-rich regions (>60%). For another 16%, the closest probe was at least 1,000 bp from the VNTR possibly due to low-complexity intervening sequence. We also tested adVNTR genotype calls for overlapping VNTRs between Illumina and HiFi sequencing and obtained 97.4% consistency, with lack of spanning reads as the main reason for genotype discrepancy. Thus, good coverage is essential for accurate genotyping, but is possible only for a subset of targeted VNTRs.
In contrast, on the 40X WGS HiFi dataset, we captured 98% of all VNTRs and 49 (92%) P-VNTRs with > 15 spanning reads, albeit with a higher CoV (0.22). Among P-VNTRs,16% were highly polymorphic in repeat count (motif CoV up to 0.8), highlighting the importance of correct genotyping. Our findings demonstrate that targeted sequencing provides high and consistent coverage for a smaller subset of low-GC VNTRs, but WGS is more effective for broad sampling of a large number of VNTRs and P-VNTRs with sufficient coverage. The VNTR genotypes (repeat motif counts) for 10,529 VNTRs in 28 samples from the HPRC WGS cohort, will be provided as a resource for the community.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session III

PB4449 Analyzing bivariate cross-trait genetic architecture in GWAS summary statistics with the BIGA cloud computing platform

Authors:

Y. Li[1], F. Xue[1], B. Li[1], Z. Fan[2], J. Shu[3], J. Lin[4], X. Wang[1], C. Copana[1], B. Zhao[5]; [1]Purdue Univ., West Lafayette, IN, [2]Univ. of Pennsylvania, Philadelphia, PA, [3]Purdue Univ., Lafayette, IN, [4]Yale Univ., New Haven, CT, [5]Univ. of Pennsylvania, Dept. of Statistics and Data Sci., Philadelphia, PA

As large-scale biobanks provide increasing access to deep phenotyping and genomic data, genome-wide association studies (GWAS) are rapidly uncovering the genetic architecture behind various complex traits and diseases. GWAS publications typically make their summary-level data (GWAS summary statistics) publicly available, enabling further exploration of genetic overlaps between phenotypes gathered from different studies and cohorts. However, systematically analyzing high-dimensional GWAS summary statistics for thousands of phenotypes can be both logistically challenging and computationally demanding. In this paper, we introduce BIGA (http://bigagwas.org/), a website that offers unified data analysis pipelines and centralized data resources for cross-trait genetic architecture analyses using GWAS summary statistics. We have developed a framework to implement statistical genetics tools on a cloud computing platform, combined with extensive curated GWAS data resources. Through BIGA, users can upload data, submit jobs, and share results, providing the research community with a convenient tool for consolidating GWAS data and generating new insights.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session I

PB4450 Ancestry determination using REVEAL: VariantBank, an efficient storage, management, and computational analysis platform for VCF files

Authors:

M. Peterson, W. Jack, C. Garay, J. Zahiri, Z. Pitluk, S. Sarangi; Paradigm4, Inc., Waltham, MA

The growth in availability and accessibility of genomics data is transforming precision medicine and healthcare. Large genomic datasets provide numerous benefits such as targeted drug discovery, personalized treatment strategies based of pharmacogenomics, and expanding our understanding of disease, especially when ethnic cohorts can be created for complex diseases where multiple genetic factors are at play. However, conducting population genomics studies at scale present a variety of challenges such as - i) high-performance storage that is scalable, distributed, and supports both structured and unstructured data types; ii) efficient data management for querying, retrieval, and integration into existing pipelines; iii) versioning, security, and data provenance; iv) easy to use interfaces that employ elastic scaling for computational analyses; v) ability to query complementary multiomics data. In this work, we evaluated REVEAL, a bioinformatics platform developed by Paradigm4, for storage, management, and analysis of publicly available genomics data in the Variant Call Format (VCF). The VCF is a text file format generated during the variant calling process that contains genomic information, locations of variants in a group of sequenced samples, and associated metadata (e.g., version number, software, reference genome used). REVEAL facilitates easy loading of VCF files from cloud or local storage (e.g., AWS S3 bucket) into *flexFS*, a POSIX-compliant networked file system designed for high-throughput data transfer across thousands of nodes. Associated phenotypic data and sample metadata (e.g., biochemical measurements or prescription records) is also stored and organized into n-dimensional arrays, which are accessible via REST, R, and Python API's. REVEAL utilizes elastic scaling called Burst to run computational analyses such as VEP annotation or ancestry determination algorithms. The datasets that we loaded using REVEAL comprised of sample/ individual level gVCF files from the 1000 Genomes Project, Human Genome Diversity Project, and Simons Genome Diversity Project. Using the APIs we demonstrate querying of the VCF's for use cases such as identifying individuals with particular gene mutations, finding most frequent variants within gene regions, among others. We also show a scalable workflow for annotation with dbSNP using VEP, and finally we present results from implementing an machine learning algorithm, XGBoost, for ancestry determination in REVEAL.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session II

PB4451 Annotation and scoring of the deleteriousness of individual genetic variants in the 4th release of the Alzheimer's Disease Sequencing Project

Authors:

N. Kannan[1], N. Wheeler[1], Genome Center for Alzheimer's Disease, L-S. Wang[2], Y. Leung[2], W. S. Bush[1]; [1]Cleveland Inst. for Computational Biology, Dept. for Population and Quantitative Hlth.Sci., Case Western Reserve Univ., Cleveland, OH, [2]Dept. of Pathology and Lab. Med., Penn Neurodegeneration Genomics Ctr., Perelman Sch. of Med., Univ. of Pennsylvania, Philadelphia, PA

The Alzheimer's Disease Sequencing Project (ADSP) is a genetics initiative instituted by the National Institute of Aging to understand the genetic architecture of Alzheimer's Disease and related dementias. ADSP recently released its 4th (R4, latest) dataset consisting of 36,361 sequenced genomes, which contains genotype calls for 362 million genetic variants, the vast majority of which are rare.

Late Onset Alzheimer's Disease (LOAD) is highly polygenic, with much of its genetic architecture unknown. Predicting and scoring of the biological effects of individual variants provides valuable information when testing these variants for association with LOAD, and facilitates identification of novel associations.

We deployed CADD (Combined Annotation Dependent Depletion), SnpEff, and Ensembl VEP (Variant Effect Predictor) variant annotation and effect prediction tools locally to score potential impact and deleteriousness of ADSP R4 variants.

CADD provides both "raw" C-scores (that directly describe deleteriousness of a variant) and "scaled" C-scores (that describe the rank of deleteriousness of the variant relative to all other possible substitutions). A variant with a scaled C-score >= 10 is among the top 10% most deleterious substitutions possible for the human genome.

SnpEff was used to specifically predict potential LOF(Loss of Function) effects. We observed 1.9 million missense variants (0.76%), 76,672 frameshift variants (0.03%) and 754,927 splicing-related variants (0.3%) based on VEP consequence predictions. 443,333 variants (0.12%) were predicted to be LOF by snpEff.

Variants with predicted LOF had a average raw C-score of 1.98, while variants without predicted LOF had a significantly lower mean raw C-score of 0.40 ($t = 789.10$, $p < 0.001$), as expected. Conversely, among 443,333 variants with predicted LOF, 18,858 (4.25%) had C-scores >= 10, while 27,878,923 (7.68%) of variants without predicted LOF had C-scores >= 10 ($\chi^2 = 7377.20$, $df = 1$, $p < 0.001$).

Notably, the snpEff LOF designation only applies to coding variants, so the relative enrichment of higher C-scores for non-coding variants may indicate the presence of regulatory variants with equivalent deleterious effects to those observed in coding regions.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session III

PB4452 Applying ACMG/AMP Variant Curation Guidelines for SCID - ClinGen SCID VCEP Recommendations

Authors:

M. Zelnick[1,2], V. Jacovas[1,2], K. Namrata[1], X. Pan[1], S. Martin[3], J. Ross[4], B. McLean[4], S. McNulty[5], M. Cowan[6], J. Puck[7], M. Hershfield[8], J. Verbsky[9], J. Walter[10], E. Allenspach[11], A. Chan[6], N. S. Van Oers[12], R. Gosh[13], M. Piazza[14], B. Yuan[1,15], B. Johnson[3], I. Chinn[1,2]; [1]Baylor Coll. of Med., Houston, TX, [2]Texas Children's Hosp., Houston, TX, [3]Invitae Corp., Fort Lauderdale, FL, [4]Univ. of North Carolina Chapel Hill, Chapel Hill, NC, [5]Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, [6]Univ. of California San Francisco Benioff Children's Hosp., San Francisco, CA, [7]Univ California San Francisco, San Francisco, CA, [8]Duke Univ Med Ctr, Durham, NC, [9]Med. Coll. of Wisconsin, Milwaukee, WI, [10]Univ. of South Florida, Tampa, FL, [11]Seattle Children's Hosp., Seattle, WA, [12]UT Southwestern Med. Ctr., Dallas, TX, [13]NIH, Bethesda, MD, [14]Prevention Genetics, Marshfield, WI, [15]Human Genome Sequencing Ctr., Houston, TX

Introduction: The Clinical Genome Resource (ClinGen) has established Clinical Domain Working Groups (CDWG) across various medical subspecialties. The Immunology CDWG was created in 2019 to uncover the relationships between genes and variants connected with Inborn Errors of Immunity (IEI). Severe Combined Immunodeficiency Disease (SCID) is a rare IEI resulting in severe T- and B-cell immunodeficiency and heightened infection susceptibility. Without rapid diagnosis and definitive treatment, SCID typically results in mortality in the first year of life. Currently, variations in 19 genes result in SCID. The role of the Variant Curation Expert Panels is to provide evidence-based interpretations of genomic variants to aid in rapid SCID diagnosis and promote precision clinical care. The ClinGen SCID VCEP aimed to establish gene-specific guidelines for variant classification since there was a lack of gene-specificity in the 2015 guidelines provided by the American College of Medical Genetics and Genomics (ACMG)/Association for Molecular Pathology (AMP).
Methods and Findings: The SCID VCEP developed 20 rules from the 28 ACMG/AMP criteria to accurately interpret variants with seven SCID-associated genes. The genes of interest include those most commonly identified as defective in North American SCID cases: *IL2RG*, *JAK3*, *ADA*, *DCLRE1C*, *IL7R*, *RAG1*, and *RAG2*, with slight differences in rules applied between each gene. The new criteria were tested on both common and rare variants to ensure the effectiveness of the specifications. To confirm validity, all variants were curated in pairs, and reviewed by both Immunology and genetics experts. The feedback led to iterative refinements of the specifications during pilot testing of the specifications. This process resulted in comprehensive and finalized guidelines ready for implementation in the sustained phase of curation.
Conclusions: This standardized and evidence-based interpretation of genetic variants associated with SCID is essential to ensure accurate diagnosis, effective clinical management, and informed genetic counseling for affected families.The efforts of the SCID VCEP represent a significant step forward in our understanding of SCID and its associated genes. It has the potential to elucidate pathogenicity, enhancing clinical decision-making and improving overall patient outcomes significantly. The rigorous process of curation and expert evaluation has yielded specifications that are finely tuned and designed to meet the highest clinical standards of accuracy and precision.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session I

PB4453 Assessing and addressing the burden of VUSs within adult medical genetics practices

Authors:

A. Stergachis, A. Folta, A. E. Sedeno Cortes, M. Jose, M. Horike-Pyne, A. E. McEwen, S. Fayer, L. Starita, M. Dubard-Gault, F. M. Hisama; Univ. of Washington, Seattle, WA

Genomic medicine aims to use patient-level genetic data to improve clinical care. However, the majority of clinically encountered variants are classified as variants of uncertain significance (VUS), which cannot be used for making clinical decisions given their unknown relationship to disease. To better understand the burden of VUSs within clinical medicine, we created the Brotman Baty Institute Clinical Variant Database (BBI-CVD), which is an electronic health record linked (EHR-linked) database of phenotype and clinical germline genetic variant information from ~4,000 patients with rare genetic disorders cared for at two adult tertiary medical centers. We demonstrate that the VUS burden is not equally shouldered across different patients at these centers. Specifically, certain testing indications result in >3-fold higher rate of VUS identification relative to pathogenic (P) or likely pathogenic (LP) variants. For example, 2.52 VUSs are returned for every P/LP variant identified in someone undergoing testing for cardiomyopathy, whereas the rate of VUSs to P/LP variants was 0.74 in individuals undergoing testing for sensorineural hearing loss. In addition, we found that individuals with different self reported ancestry had markedly different rates of VUS identification. For example, 2.49 VUSs are returned for every P/LP variant identified in individuals identifying as Black or African American, whereas the rate of VUSs to P/LP variants was 0.98 in individuals identifying as European ancestry. Finally, we demonstrate that we can leverage the BBI-CVD in conjunction with emerging variant-level functional data to enable the reclassification of VUSs for individual patients cared for at these medical centers. Specifically, integration of BRCA1 functional data with the BBI-CVD enabled us to reclassify one patient's VUS as LP and two patients' VUSs as likely benign, enabling updated care recommendations by their providers. Overall, the BBI-CVD helps illuminate the real-world burden of VUSs within adult tertiary medical genetics practices, and offers a resource to improve the diagnosis and care of patients with rare disorders.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session II

PB4454 Assessing the Usage of Human Phenotype Ontology Terms in Facilitating Discovery of Disease Causal Genes

Authors:

J. Chan[1], K. Wang[2,3]; [1]The Children's Hosp. of Philadelphia, Philadelphia, PA, [2]Children's Hosp. of Philadelphia, Philadelphia, PA, [3]Perelman Sch. of Med., Univ. of Pennsylvania, Philadelphia, PA

Phenotypic abnormalities are often associated with genetic disorders and analyzing these abnormalities can assist physicians in identifying potential disease-causing genes in patients with rare diseases. In this study, our objective is to evaluate the effectiveness of computationally generated Human Phenotype Ontology (HPO) terms in identifying disease causal genes in patients with genetic diseases. The CHOP Arcus database contains over 71,000 patients diagnosed with genetic diseases, each of whom has associated HPO terms generated by natural language processing algorithms on clinical narratives. By employing gene prioritization tools and leveraging the HPO terms, we conducted an analysis to determine the ability of using HPO terms to facilitate the discovery of causal genes or diseases. For a subset of patients, we also compared the results from computationally generated HPO terms with those generated by human experts during patient encounter. Our results suggest that phenotypic information can facilitate physicians to effectively narrow down their focus when assessing potential avenues for diagnosis or treatment.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session III

PB4455 Barriers to Uploading Genetic Testing Report to Online Patient Data Sharing Registry

Authors:

B. VanHoute[1], J. M. Savatt[2], C. Cooney[3], A. M. Rossel[1], C. Ineke[1], A. Morgan[4], D. Azzariti[5], H. L. Rehm[6], C. L. Martin[2], E. R. Riggs[7]; [1]Geisinger, Lewisburg, PA, [2]Geisinger, Danville, PA, [3]Geisinger, Jersey City, NJ, [4]Geisinger, Kingston, PA, [5]ClinGen Resource, Cambridge, MA, [6]Broad Inst., Cambridge, MA, [7]Geisinger, Huntsville, AL

Purpose: GenomeConnect, the Clinical Genome Resource (ClinGen) patient registry, enables participants to share their genetic and health data. Structured genetic data collected from participant-uploaded reports, along with participant-provided phenotypes, are shared with databases, such as ClinVar, improving overall genomic knowledge. Participants are also able to learn about research opportunities, connect with others with similar variants and receive updates about their genetic results. Although many benefits of the registry rely on individuals sharing their genetic reports, to date only 1,656 of 4,778 participants (34.7%) have shared a report. This ongoing study aims to investigate barriers to report sharing.
Methods: A brief online survey was distributed to all consented and eligible participants that had not uploaded a genetic report as of May 2023 (n=2,351). The survey included skip logic, inquiries about the results of the participant's testing, recollection of reminder emails, demographic questions, an option to participate in a semi-structured interview, and instructions for how to upload a report.
Results: As of June 2, 2023, 102 participants completed the survey (4.3% response rate). This survey is ongoing with an expected end date of June 17, 2023. Of respondents, the majority recalled GenomeConnect participation (n=78/102, 76.5%) and were interested in sharing a genetic report (n=56/70, 80.0%). Many respondents indicated they did not have a copy of the report (n=17/70, 22.9%), did not know uploading was part of participation (n=18/70, 25.7%), forgot (n=11/70, 15.7%), experienced difficulty (n=25/70, 35.7%), or did not have time (n=7/70, 10%). Only 4 respondents (5.7%) reported privacy concerns, and other barriers such as a lack of benefit were selected by single participants. Since survey distribution, 69 recipients shared a report or indicated they would like GenomeConnect to directly request their results.
Conclusions: Our study demonstrates that the majority of respondents are willing to share their genetic report as part of GenomeConnect participation. Factors impacting sharing are related to report availability and the upload process. Future work, including semi-structured interviews, will further characterize barriers to report sharing and inform strategies to support participants. Survey data alone suggest a need to revisit current resources, such as email reminders, videos, and email instructions, and to explore the creation of additional report sharing supports. Ultimately, engaging as many patients in data sharing as possible will improve understanding of genes and genomic variants.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session I

PB4456 Batch ClinVar Submission Support in ClinGen's Variant Curation Interface (VCI)

Authors:

M. Wright[1], C. Preston[1], G. Cheung[1], M. Mandell[1], B. Wulf[1], L. Babb[2], M. DiStefano[2], S. Harrison[3], C. Klein[1], R. Shapira[1], I. Keseler[1], D. Ritter[4], N. Shah[4], K. Riehle[4], A. Milosavljevic[4], S. Plon[5], T. Klein[1], Clinical Genome Resource (ClinGen); [1]Stanford Univ. Sch. of Med., Palo Alto, CA, [2]The Broad Inst., Boston, MA, [3]Ambry Genetics, Salt Lake City, UT, [4]Baylor Coll. of Med., Houston, TX, [5]Baylor Coll. Med., Houston, TX

The NIH-funded Clinical Genome Resource Consortium (ClinGen) Variant Curation Interface (VCI) is a global, open-source variant classification platform for supporting the application of evidence criteria and classification of variants based on the ACMG/AMP sequence variant classification guidelines. To facilitate evidence-based improvements in human variant classification, the VCI is publicly available to the worldwide genomics community. The VCI is among a suite of tools developed by ClinGen, and supports an FDA-recognized human variant curation process of ClinGen Variant Curation Expert Panels (VCEPs). ClinGen is expanding to involve more curators and teams of curators working today (affiliations) as part of an increasing scale of activities that will increase genetic variants across a greater number of genes. The variant curation workflow is intended to support dissemination of variant curations into two repositories: the ClinGen Evidence Repository (for approved ClinGen VCEPs) and ClinVar. Support for dissemination of curations into the Evidence Repository from the VCI is an API-driven process via the ClinGen Data Exchange; however, support for ClinVar submission from the VCI has been largely manual to date. Here we present the first in a series of planned software features to provide a better ClinVar submission experience for VCI users. This batch submission feature allows users to create an active batch of curated variants and then download a pre-formatted file of the curation data, which can easily be submitted to ClinVar. This provides a faster process for users submitting variant interpretations to ClinVar. It also allows VCI users to annotate which of their curations have been submitted to ClinVar. Future feature plans for ClinVar submission from the VCI include providing VCI users with real-time feedback on the status of their ClinVar submissions, as well as direct API-based submissions from the VCI. These features together will further streamline workflows for both ClinGen VCEP and non-ClinGen VCI users, and support the ClinGen goal of creating scalable curation workflows to support the clinical genomics community.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session II

PB4457 Black and African American Connections to Parkinson's Disease Study

Authors:

P. Wild Crea[1], M. Makarious[2], K. Levine[3,4], D. Vitale[1,4], M. Koretsky[3], M. Nalls[4,3], Z-H. Fang[5], D. Hall[6], T. Xie[7], M. Padmanaban[7], E. Shamin[8,9,10], D. Standaert[11], M. Dean[11], E. Disbrow[12], A. Rawls[13], J. Solle[14], N. Louie[14], K. Billingsley[15,3], P. Alvarez Jerez[16,3], D. Hernandez[15], S. Arepalli[1], L. Malik[3], A. Miano-Burkhardt[15,3], H. Leonard[3,4], H. Iwaki[3,4], C. Blauwendraat[15,3], A. Singleton[15,3], S. Bandres Ciga[3], Black and African American Connections to Parkinson's Disease Study, Global Parkinson's Genetics Program; [1]Lab. of Neurogenetics, NIH, Bethesda, MD, [2]NIH, Bethesda, MD, [3]Ctr. for Alzheimer's Disease and Related Dementias, NIH, Bethesda, MD, [4]Data Tecnica, Washington, DC, [5]German Ctr. for Neurodegenerative Diseases (DZNE), Tübingen, Baden-Württemberg, Germany, [6]Dept. of Neurological Sci., Rush Univ. Med. Ctr., Chicago, IL, USA, Chicago, IL, [7]Dept. of Neurology, Univ. of Chicago Med., Chicago, Illinois, USA, Chicago, IL, [8]Human Motor Control Section, Natl. Inst. of Neurological Disorders and Stroke, NIH, Bethesda, MD, [9]Kaiser Permanente Mid-Atlantic States, Largo, MD, [10]MidAtlantic Permanente Res. Inst., Rockville, MD, [11]Dept. of Neurology, Univ. of Alabama at Birmingham, Birmingham, AL, [12]LSU Hlth.Shreveport Ctr. for Brain Hlth., Shreveport, Louisiana, United States of America, Dept. of Neurology, LSU Hlth.Shreveport, Shreveport, LA, [13]Univ. of Florida Norman Fixel Inst. for Neurological Diseases, Gainesville, FL, [14]Dept. of Clinical Res., Michael J. Fox Fndn. for Parkinson's Res., New York, NY, [15]Lab. of Neurogenetics, Natl. Inst. on Aging, NIH, Bethesda, MD, [16]UCL Movement Disorders Ctr., Univ. Coll. London, London, United Kingdom

Title: Black and African American Connections to Parkinson's Disease Study

Aim: BLAAC PD is a multi-center study recruiting Black and African American individuals with Parkinson's Disease (PD) and healthy controls. BLAAC PD aims to provide a platform for replication studies to explore the relevance of genetic findings reported in other populations and investigate genotype-phenotype correlations. The ultimate goal is to create a foundational cohort to assess diverse aspects of PD in this historically excluded population and serve as a model for diversity and equity in research.

Background: Understanding genetic mechanisms across diverse populations can provide unique insights into complex traits like PD. Our current insights of the genetics underlying PD etiology has been disproportionately based on European ancestry populations. This has led to a significant gap in our knowledge about the disease's genetics and clinical characteristics in underrepresented populations, particularly individuals of African and African admixed ancestries.

Materials & Methods: BLAAC PD collects samples from six sites across the United States. A total of 147 cases and 174 controls have been collected. Following DNA extraction, samples are genotyped and imputed, followed by ancestry assessment through a pre-trained machine learning model based on reference sample series. A comprehensive assessment was conducted to investigate known and novel genetic contributors.

Results: Our analyses showed consistent differences in variant frequencies, magnitude of effects and risk alleles for disease-causing mutations in PD known genes including SNCA, VPS35, LRRK2, PRKN, PINK1, DJ1 and GBA. Additionally, we screened genetic risk loci known to be associated with PD in multiple ancestral populations including over 100 loci. We also assessed structural variants in early-onset and familial PD cases.

Discussion: These findings highlight the need for diverse representation in genetic research. Insufficiently diverse genetic data may exacerbate health disparities once translated to the clinic. BLAAC PD will help resolve the cross-ancestry applicability of drug targets, treatments and preventative measures by elucidating the genetic architecture of disease in African and African admixed ancestries and also providing a platform for replication studies of previous work in other populations. Looking ahead, BLAAC PD plans to continue recruiting and genotyping participants to serve as a foundational cohort for diverse genetic studies.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session III

PB4458 Build the coordinates-compatible, consensus human reference genomes.

Authors:

L. Wang[1], T. Ma[1], J. Qi[2], E. Z. Prestano[1], Y. Asmann[3], J. Nie[4]; [1]Mayo Clinic, ROCHESTE, MN, [2]Hefei Inst.s of Physical Sci., Hefei, China, [3]Mayo Clinic, Jacksonville, FL, [4]Hefei Cancer Hosp., Hefei, China

Motivation:The current human reference genome (GRCh38) suffers from "reference bias" as approximately 70% of the sequences are derived from a single individual. This bias significantly affects the results of most next-generation sequencing-based genomic assays.
Methods:We have developed a tool called "genometa" for constructing consensus reference genomes (CRGs) using the standard reference genome as a template. Genometa achieves this by replacing minor alleles on the standard reference genome with non-reference major alleles (NRMAs). NRMAs are genomic variants that are dominant (allele frequency > 0.5) in specific populations but are not represented in the standard reference genome. Substitution variants were directly incorporated into the template without altering the original coordinates. For insertions and deletions (INDELs), genometa first identifies the Balanced Insertion and Deletion (BID) regions and then integrates these regions into the reference template using a dynamic programming algorithm that maximizes the utilization of INDEL variants.
Results:On average, we obtained approximately 2.35 +/- 0.072 million NRMA variants for each population from gnomAD, consisting of 84% SNVs and 16% Indels. We incorporated around 1.97 +/- 0.057 million substitution variants and 5133 +/- 584 BID regions into GRCh38 to construct population-specific CRGs. As a result, we generated CRGs for 10 populations (AFR, AMI, AMR, ASJ, EAS, FIN, MID, NFE, SAS, OTH) included in the gnomAD project.
To compare the CRGs with the standard GRCh38, we downloaded 42 whole-genome sequencing data of 6 AFR, 27 EAS, 6 NFE, and 3 MID individuals from the Human Genome Diversity Project (HGDP). Using the BWA-MEM, we aligned the WGS data of these individuals to both GRCh38 and the corresponding CRGs. Consistently across all tested samples, the alignment rates showed significant improvement when aligned to the respective CRGs. Compared to the alignment with GRCh38, for every 1 million read pairs, we observed an average increase of 2409 +/- 463 properly-paired reads. Additionally, there was an average increase of 30542 +/- 2939 mapped bases and a reduction of 63784 +/- 4591 mismatches.
Summary:Population-specific CRGs offer a substantial improvement in alignment rates, thereby significantly impacting genomic and epigenomic analyses. This enhancement enables the accurate identification of DNA variants and cis-regulatory elements, contributing to a more comprehensive understanding of genetic and human diseases. The compatibility of CRGs' coordinates also facilitates seamless cross-sample comparisons, data integration, and visualization.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session I

PB4459 Building a phenotype for *ALPK3* loss-of-function mutation heterozygotes: Leveraging PheWAS to enhance reverse phenotyping studies

Authors:

N. Singh-Miller[1], A. Awan[1], C. Wilczewski[1], L. Hellwig[2], M. Haigney[2], L. Biesecker[1], H. Mo[1], C. Turner[1]; [1]Ctr. for Precision Hlth.Res., Natl. Human Genome Res. Inst., NIH, Bethesda, MD, [2]Ctr. for Military Precision Hlth.(CMPH), Uniformed Services Univ. of the Hlth.Sci., Bethesda, MD

Large patient datasets linking genomic data with clinical data, such as the *All of Us* (*AoU*) program, offer an invaluable opportunity to explore genotype-phenotype relationships and advance our understanding of heritable conditions. However, the limited depth of ICD-10 encoded electronic medical record (EMR) data coupled with restrictions on patient follow-up poses challenges to comprehensive phenotyping. Smaller cohorts of recallable patients available for targeted deep-phenotyping overcome these challenges. However, these cohorts are limited by their size and the cost of resource-intensive recall to phenotype traits with potentially small expected effect sizes. In this study we implemented an innovative phenotyping approach to address these challenges by using phenome-wide association studies (PheWAS) to improve the design of a targeted reverse phenotyping study aimed at investigating novel gene-disease relationships for *ALPK3*. Putative loss of function (pLoF) variants in *ALPK3* are known to cause a severe autosomal recessive cardiomyopathy phenotype. Currently there is an emerging phenotype associated with *ALPK3* pLoF heterozygosity, however the overall phenotype and penetrance in phenotypically-unselected genotype-positive individuals remains unknown.

We assembled a genotype-positive cohort of individuals from the *AoU* database based on their heterozygosity for a pLoF variant in *ALPK3*; clustering pLoF heterozygotes to increase the likelihood of identifying statistically significant features. We have initially identified fifteen such features for stop-gain mutations with a p-values less than 2×10^{-3}; including cardiac phenotypes such as cardiogenic shock (OR=16, p-value= 1.9×10^{-4}), as well as numerous inflammatory phenotypes such as mastoiditis (OR=23, p-value = 2.4×10^{-5}), optic neuritis (OR=8, p-value = 8.7×10^{-5}), postphlebitic syndrome (OR=32, p-value = 9.4×10^{-4}), and chronic pericarditis (OR=25, p-value = 1.6×10^{-3}). With this insight, we designed and initiated a phenotype-specific reverse phenotyping study with 10 genotype-positive patients in the NIH Reverse Phenotyping Core (RPC) cohort.

Through this methodology, we demonstrate that reverse phenotyping studies with re-contactable cohorts can be guided further with clinical informatics using large cohorts that, for now, are unavailable for follow-up phenotyping.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session II

PB4460 Building the ClinGen Pathogenicity Calculator Version 2.0 by leveraging ClinGen API Microservices.

Authors:

K. Riehle[1], A. R. Jackson[1], N. Shah[1], A. A. Zuniga[1], T. Farris[1], A. Kodochygov[1], K. Yu[1], A. Mihajlović[2], B. Jevtić[2], M. Milinkov[2], N. Vučinić[2], D. Jevtić[2], N. Martinović[2], M. Wright[3], C. Preston[4], M. Mandell[5], G. Cheung[5], D. Ritter[1], J. Arce[1], V. Dang[1], G. Noble[1], S. Harrison[6], A. Byrne[7], D. Azzariti[8], E. Riggs[9], M. DiStefano[10], L. Babb[11], K. Ferriter[11], E. McCormick[12], X. Gai[13], L. Shen[14], M. Falk[15], A. Wagner[16], J. Arbesfeld[17], M. Cline[18], A. F. Rubin[19], O. Griffith[20], M. Griffith[21], T. Klein[22], S. Plon[23], A. Milosavljevic[1], Clinical Genome Resource; [1]Baylor Coll. of Med., Houston, TX, [2]Persida Inc., Brooklyn, NY, [3]Stanford Univ. Sch. of Med., Palo Alto, CA, [4]Stanford Sch. of Med., Menlo Park, CA, [5]Stanford Univ. Sch. of Med., Stanford, CA, [6]Ambry Genetics, Salt Lake City, UT, [7]Broad Inst. of MIT and Harvard, Cambridge, MA, [8]ClinGen Resource, Cambridge, MA, [9]Geisinger, Huntsville, AL, [10]The Broad Inst., Macungie, PA, [11]The Broad Inst. of MIT and Harvard, Cambridge, MA, [12]Children's Hosp. of Philadelphia, Philadelphia, PA, [13]CHILDREN S Hosp. OF LOS ANGELES, Los Angeles, CA, [14]Children's Hosp. Los Angeles, Los Angelese, CA, [15]CHOP, Philadelphia, PA, [16]Nationwide Children's Hosp., Columbus, OH, [17]The Ohio State Univ. Coll. of Med., Columbus, OH, [18]UC Santa Cruz, Santa Cruz, CA, [19]WEHI, Melbourne, Australia, [20]Washington Univ. Sch. of Med., St Louis, MN, [21]Washington Univ, St. Louis, MO, [22]Stanford Univ, Stanford, CA, [23]Baylor Coll. Med., Houston, TX

The Clinical Genome Resource (ClinGen) suite of microservices provides a platform to build applications within and outside of ClinGen by leveraging Application Programming Interfaces (APIs). The ClinGen APIs and messaging queues have been employed within ClinGen to support variant curation, which includes data aggregation, curation, and dissemination of published pathogenicity assertions. Version 1.0 of the Pathogenicity Calculator was aimed at non-ClinGen users, supported ACMG/AMP v3.0 2015 guidelines, and was utilized as a classic siloed application. Version 2.0 is upgraded to support an early draft of the ACMG/AMP/CAP/ClinGen SVC v4.0 2023 guidelines and leverages ClinGen API microservices, including the ClinGen Allele Registry, ClinGen Linked Data Hub, and ClinGen Criteria Specification Registry. In addition to the utility to the end-users, Version 2.0 thus provides a new model for how interfaces and workflows within and outside of ClinGen may leverage the ClinGen Resource.

Calculator 2.0 users first identify a variant of interest, uniquely identified by a stable canonical allele identifier (CA ID) obtained from free, open, on-demand registration within the Allele Registry. Supporting evidence is obtained and aggregated in the form of excerpts linked to a given CA ID in the Linked Data Hub, which facilitates open sharing of linked data from external sources and automated calculation of some evidence codes as Calculator input. ClinGen variant curation expert panel (VCEP) specifications are created, edited, and approved following the ClinGen expert panel process within the Criteria Specification Editor and are publicly-accessible via the Criteria Specification Registry. Criteria specifications are gene and disease specific structured data developed from a combination of empirical analysis and expert recommendations. Calculator 2.0 can use the tailored VCEP-specific gene-condition criteria specifications from the Criteria Specification Registry with a flexible user interface to solicit the required user input and calculate variant pathogenicity scores.

In conclusion, Version 2.0 of the ClinGen Pathogenicity Calculator exemplifies the power of ClinGen API microservices to serve as a platform for rapid development of new applications within and beyond ClinGen. This microservice-oriented model for interface development is particularly relevant for established variant curation and interpretation workflows outside of ClinGen, as they may tap into ClinGen API microservices to accelerate development, leverage ClinGen knowledge within their workflows, and adopt the latest professional guidelines.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session III

PB4461 Calibrating CADD on a gene level for enhanced causal variant filtration

Authors:

B. Pusey[1], J. Fu[2], E. Macnamara[3], W. Gahl[4], D. Adams[5]; [1]Natl Human Genome Res. Inst., Homer, AK, [2]NHGRI, Rockville, MD, [3]NIH, Bethesda, MD, [4]NHGRI (NIH), Bethesda, MD, [5]NIH, Bethesda, MD

The NIH Undiagnosed Diseases Program (UDP), a member of the Undiagnosed Diseases Network (UDN), was established with the goal of conducting research for patients with significant illnesses that remain undiagnosed despite extensive medical evaluation. Genome sequencing approaches became available around the same time that the UDP started and are important tools for making diagnoses in some cases. Given the increasing prevalence of exome and genome sequencing in standard medical care, many current UDP applicants have been evaluated with this technology before they apply. As a result, a major focus of UDP genomic analysis is assessment of variants classified as “variants of unknown significance”. In this forest of variants it can be difficult to know which has a truly significant CADD (Combined Annotation Dependent Depletion) score, and the current practice has been to use an arbitrary or semi-arbitrary cut off based on a general view of the genome in its entirety. What can be determined to be a significant CADD score varies across genes dependent on a multitude of factors such as the gene size, and inherent genetic mutation burden etc.. Using gnomAD (Genome Aggregation Database) a standard distribution curve was calculated on a per gene basis. Using these distributions z-scores were calculated creating a calibrated score (cadd_z) that reflects the genes inherent biases given a large population in which individuals with severe disease have been culled from the cohort. Using these recalibrated cadd_z scores we then reanalyzed out cohort of more than 600 families and compared the score to those obtained using our original filters with a hard CADD score cut off. Diagnosed cases were used as controls to confirm that no causal variants were lost. The calibrated cadd_z score resulted in a reduction in noise from filtration results as well as providing candidate variants of interest that had previously been excluded. CADD scores can then be re-quantified to a per gene z-score allowing for filtering relative to the scores of the gene as a whole rather than an subjective value. Furthermore, this approach can be abstracted out to other genomic scoring methodologies. Using our new CADD threshold methodology, we identified new potential candidate variants of interest.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session I

PB4462 Calypso: a comprehensive platform supporting team-based, longitudinal genomic diagnostic care

Authors:

A. Ward[1,2], I. Cooperstein[1], T. Di Sera[1], S. Georges[1], A. Pitman[1], M. Tristani-Firouzi[1], G. Marth[1]; [1]Univ. of Utah, Salt Lake City, UT, [2]Frameshift Labs, Cambridge, MA

Patients presenting with complex phenotypes represent some of the most challenging diagnostic cases. To achieve a genetic diagnosis, large interdisciplinary teams are required to work collaboratively, sharing their own unique expertise and skills. Treating physicians provide detailed understanding of the clinical presentation and family history; medical geneticists brings a wide and deep knowledge of genetic diseases; bioinformaticians provide computational analysis skills and an ability to adjudicate the role of individual variants; and diagnostic pathologists synthesize all information to draw conclusions on the clinical significance of candidate variants. Currently available software tools cater primarily to those with computational expertise rather than clinical expertise and focus on diagnostic analysis at a single point in time, despite the fact that patient phenotypes can evolve over time. Consequently, treating physicians, or genetic counselors can be left out of the diagnostic process. Further, failure to follow a patient over a potentially lengthy diagnostic process where phenotypes and/or variant interpretations can change can potentially lead to missed diagnoses.
Calypso is a web-based software system being developed to address these concerns and comprises:

1. A computational backend to perform comprehensive annotation and filtering of incoming genomes and enable re-annotation and re-analysis of all variants on a periodic or user triggered basis,
2. User-friendly web tools including the iobio suite to ensure each case is summarized with quality statistics, clinical information and other analytics that ensures all team members can work as part of an effective collaborative team,
3. Case and cohort level dashboards that employ visual elements to summarize important case information (e.g. candidate variants and case history), as well as interactive analytics across cohorts of cases,
4. A robust communication infrastructure to support close collaboration between all diagnostic team members

Initial versions of Calypso have been deployed at the University of Utah, where it is used to aid team-based diagnostic analysis as part of the rapid NICU sequencing program and at the UDN.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session II

PB4463 CARTaGENE PheWeb: presenting GWAS and PheWAS results in a Quebec cohort

Authors:

V. Chapdelaine[1,2,3], G. Lettre[4,5], S. Gagliano Taliun[4,5], D. Taliun[1,2,3]; [1]McGill Univ., Montreal, QC, Canada, [2]Canadian Excellence Res. Chair in Genomic Med., McGill Univ., Montréal, QC, Canada, [3]Victor Phillip Dahdaleh Inst. of Genomic Med. at McGill Univ., Montréal, QC, Canada, [4]Université de Montréal, Montreal, QC, Canada, [5]Montreal Heart Inst., Montréal, QC, Canada

CARTaGENE is a large longitudinal population-based cohort and a prospective health study with genetic data and a wealth of phenotypic information available to empower discoveries to better understand human health and disease. It currently includes 29,330 array-genotyped individuals recruited in the province of Quebec, Canada, of which 25,681 are of European genetic ancestry. Many of these individuals are French Canadians, a population that originated from immigration in the 17th century of 8,500 mostly French settlers to Quebec followed by serial founder effects. CARTaGENE thus provides a unique opportunity to better understand genetic risk in a founder population.

We performed a PheWAS on ~100M genetic variants imputed with the state-of-the-art TOPMed reference panel (R2 > 0.3) across 701 traits (326 binary, 375 continuous), including diseases, disease-related, behavioral, and various biochemical and anthropometric measurements. We used the regenie generalized mixed model association test to account for case-control imbalance (for binary traits) and relatedness among participants. Our analyses included sex-specific traits and chromosome X. We identified 30,678 significant associations (p-value < 5×10-8) within 1,327 independent loci. All summary statistics are publicly available to browse, visualize and download via the CARTaGENE PheWeb web-based interactive tool.

Our resource provides the most comprehensive view of genetic association results for the currently largest genotyped population cohort in Quebec. We have made these data publicly available to the worldwide research community to support a variety of downstream analyses, including replication, fine-mapping, construction of polygenic risk scores, and Mendelian randomization to better understand the genetic contributors to disease risk.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session III

PB4464 Cavatica-based Variant WorkBench

Authors:

Y. Guo[1], Q. Li[1], J. Rozowsky[2], J. Costanza[3], M. Mattioni[2], J-P. Thibert[3], C. Botek[3], M. Brown[1], D. Higgins[1], Y. Zhu[1], B. Farrow[1], A. Heath[1], A. Resnick[1], V. Ferretti[3], Kids First Data Resource Center; [1]Children's Hosp. of Philadelphia, Philadelphia, PA, [2]Velsera, Cambridge, MA, [3]Ctr. Hosp.ier Univ.ire Sainte-Justine, Montréal, QC, Canada

The Gabriella Miller Kids First Pediatric Research Program (Kids First) aims at facilitating researchers to uncover new insights into the biology of childhood cancer and structural birth defects, including the discovery of shared genetic pathways between these disorders. Kids First Data Resource Center developed the Kids First Data Resource Portal (KFDRP), which is a centralized data platform for both Kids First and collaborative cohorts. Currently, KFDRP facilitates data access for 37 studies with over 33,700 participants. To assist researchers and clinicians in identifying potential disease-causing germline variants in the human genome, KFDRP released the Variant WorkBench (VWB), which facilitates variant querying, manipulation, analysis, and visualization. Cavatica is a widely used bioinformatics analysis platform developed by Velsera (formally Seven Bridges). KFDRP-based VWB is now transitioned to a Cavatica-based version to have better support for data import/export, more convenient sharing of notebooks and data, and easier billing details, with the same level of scalable, cloud-based computing. Additionally, Cavatica-based VWB has incorporated multiple public reference databases, including Cancer Hotspots, ClinVar, COSMIC, dbNSFP, gnomAD, TOPMed, as well as gene-phenotype links provided by OMIM, HPO, Orphanet, and the Deciphering Developmental Disorders Project. These databases have been converted to parquet files, reducing process time and memory usage compared to regular text files, thus speeding up data retrieval and querying. Furthermore, Cavatica empowers users to create and share custom analysis workflows tailored to specific research questions and datasets, enhancing data utility and analysis accuracy. Here we present a gene-based variant filtering workflow that aggregates information from dbNSFP, HGMD, TOPMed, gnomAD, and ClinVar. We identified 1202 pathogenic/likely pathogenic variants in *NF1*, a gene that helps regulate cell growth, across 582 pediatric brain tumor samples within 6 mins. Among the variants, there are 21 protein-altering variants. Overall, the upgraded VWB is more powerful, user-friendly, and adaptable.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session I

PB4465 Characterization of HLA variation in Quebec from whole-genome samples in the population-based CARTaGENE cohort.

Authors:

P. McClelland[1], K. Lo[2], G. Lettre[3], C. Bhérer[1], D. Taliun[1]; [1]McGill Univ., Montreal, QC, Canada, [2]Montreal Heart Inst., Montreal, QC, Canada, [3]Université de Montréal, Montreal, QC, Canada

The human leukocyte antigen (HLA) system contains genes involved in many immunological processes, and genetic variation in this region is associated with many immune- and disease-related phenotypes, including type I diabetes, multiple sclerosis, and rheumatoid arthritis. Frequencies of genetic variants in the HLA region differ significantly across populations, making population-specific HLA characterization necessary to understand the prevalences and functional effects of HLA alleles within a given population. Here we characterize HLA alleles in individuals from the province of Quebec, Canada, using high-depth whole-genome sequencing (WGS) data from 2,180 participants from CARTaGENE, a population-based cohort that has collected high-quality biosamples and extensive health and phenotypic information from individuals across six different regions in Quebec. We performed HLA typing of class I and II classical alleles (HLA-B, -C, -DPA1, -DPB1, -DQA1, -DQB1, -DRB1, -E, -F, and -G) using HLA*LA. After quality checks, 43,189 alleles were retrieved from 2,173 participants. 286 individual HLA alleles were observed at least once in the CARTaGENE WGS dataset: 84 HLA-B, 42 C, 17 DPA1, 34 DPB1, 11 DQA1, 17 DQB1, 55 DRB1, 5 E, 2 F, and 19 G. Of these, 222 HLA alleles were also present in the most extensive publicly available multi-ancestry high-resolution HLA reference panel, HLA-TAPAS (n=21,546). 35 out of 222 HLA alleles were at a frequency significantly different in the CARTaGENE cohort (p < 2.25 x 10-4). Frequencies of 71 low-resolution (one-field) HLA allele types were also previously reported in the Hema-Quebec donor biobank in 17 administrative regions of Quebec. We compared allele frequencies by sampling site in CARTaGENE to the closest correlating administrative region in the Hema-Quebec cohort. In the CARTaGENE WGS dataset, frequencies of 38 shared HLA alleles correlated well with all regions combined in Hema-Quebec (Pearson's r = 0.9841); 2 alleles were at significantly different frequencies in CARTaGENE (p < 1.32x 10-3). For individual CARTaGENE sampling strata, HLA allele frequencies were well correlated with corresponding Hema-Quebec regions, with Pearson's r ranging from 0.88-0.93. This sequencing-based HLA allele data can be used in future to construct an HLA-specific imputation reference panel for imputing uncaptured HLA alleles into ~30,000 genotype-array samples also available within CARTaGENE, allowing for cohort-scale analysis into HLA-disease associations in the Quebec population.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session II

PB4466 Characterization of human neuroblastoma cultured cells lines used for interpretation of coding and noncoding variation in neurodevelopmental diseases

Authors:

H. B. Heins, J. G. Manuel, J. K. Ng, T. N. Turner; Washington Univ. Sch. of Med., St. Louis, MO

Neurodevelopmental disorders (NDDs) affect >1% of the population. While there is an established contribution of rare protein-coding variants and large copy number variants to NDDs, the role of rare noncoding variation is an active area of investigation. From whole-genome sequencing (WGS) data in autism, we have identified a contribution of de novo variants in promoter and enhancer regions of the genome. We are now beginning to functionally characterize these variants in in vitro assays. To optimize these experiments, we are deeply characterizing available human cultured neuroblastoma cell lines including IMR-32, SH-SY5Y, and SK-N-SH as well as the commonly used HEK293 cell line. These cell lines have been utilized in several studies to assess coding and noncoding variation, but in-depth characterization of their genomic, epigenomic, and functional genomic characteristics are not available. This information is critical for determining the best cells for functional testing. We are utilizing several technologies including Hi-C sequencing, ATAC-Seq, short-read and log-read WGS, short-read and long-read RNA sequencing, and karyotype analysis to characterize these cell lines. By karyotype analysis, we found that each cell line exhibited aneuploidy with an IMR-32 karyotype of 48~49,XY,add(1)(p36.1),+add(1)(,add(2)(p21),+6,+12,der(16)t(15;16)(q11.2;q11.2)[cp11], a SH-SY5Y karyotype of 47,X,add(X)(p22.1),add(1)(q21),+7,add(8)(q24.1),add(9)(q34),add(14)(q22),add(22)(q13)[20], a SK-N-SH p36 karyotype of 47,X,add(X)(p22.1),+del(1)(p22),+7,add(8)(q24.1),add(9)(q34),add(14)(q22),add(22)(q13)[cp21], and a HEK293 karyotype of 67~74<3N>,XXX,+X,+X,add(X)(p22.1),add(X)(q26),add(1)(q42)x2,+2,-3,4,+5,+6,+6x2,add(6)(q25),-8,-8x2,+9,-10,+10,add(12)(p13),-13,-14,-15,-15x2,-16,+17,-18,+18,+21,+22,+22x2,+1-5mar,1~3dmin[cp16]. Another interesting aspect of our work beyond building the profiles for functional testing is our application of >40× long-read WGS. This work will provide new insights into de novo assembly generation of these aneuploid genomes and will be informative in other contexts (e.g., cancer). Our hope is to provide this data for our highly characterized HEK293 and human neuroblastoma cell lines to the research community to further the understanding of genomic and regulatory elements and noncoding variation in them.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session III

PB4467 Characterization of long non-coding RNA and circular RNA in the aging human hippocampus across three library types.

Authors:

A. Knyshov[1], J. Hu[1,2], M. Lin[1], B. Zhou[1], J. A. Empawi[1], S. Puri[3], T. D. Stein[4], L. A. Farrer[1,5], B. Wolozin[3], X. Zhang[1,5]; [1]Dept. of Med. (BioMed. Genetics), Boston Univ. Chobanian & Avedisian Sch. of Med., Boston, MA, [2]Bioinformatics Program, Boston Univ., Boston, MA, [3]Dept. of Anatomy & Neurobiology, Boston Univ. Chobanian & Avedisian Sch. of Med., Boston, MA, [4]Dept. of Pathology and Lab. Med., Boston Univ. Chobanian & Avedisian Sch. of Med., Boston, MA, [5]Dept. of Biostatistics, Boston Univ. Sch. of Publ. Hlth., Boston, MA

Ribosomal RNA depletion (ribo-minus) is commonly used for RNA sequencing to capture poly(A)+ mRNA, non-coding RNAs (ncRNAs), or protein-coding mRNAs that are not polyadenylated. The poly(A)-selected (poly(A)+) protocol enriches poly(A)+ transcripts including mRNAs and ncRNAs. However, many functional long transcripts (>200 nt) are known to lack poly(A) tails including circular RNAs (circRNAs), which are much less abundant than protein-coding mRNAs in human tissues.

In order to enrich these poly(A)- transcripts and estimate their function in aging brains, in this study, we first applied a polyA-depleted (poly(A)-) protocol to generate RNAseq data for 128 human hippocampus samples. For the same samples, ribo-minus and poly(A)+ RNAseq data were also generated in a similar manner (~40M 100bp paired reads per sample) using Illumina NovaSeq at the Genome Center at Yale University. We then systematically assessed the expression of linear long ncRNA (lncRNA), detected circRNA, and compared their expression levels across the three library types.

Our results show that both ribo-minus and poly(A)- sequencing perform well for linear lncRNA profiling in hippocampus, with poly(A)- libraries generally having a slightly higher number of detected expressing genes (per sample median n=2,661 at TPM > 0.1) compared to ribo-minus libraries (per sample median n=2,591 at TPM > 0.1). Similarly, the number of detected circRNA is only slightly higher in poly(A)- libraries (per sample median n=2,936 at FPB > 0.1) compared to the ribo-minus libraries (per sample median n=2,886 at FPB > 0.1). These results suggest that both strategies are effective for studying ncRNA in human tissue.

Additionally, we compared poly(A)- and poly(A)+ data generated from the same samples to predict the polyA status of protein-coding and lncRNA transcripts in human hippocampus. A ROC-based method was applied using a set of previously known status genes (n=6,104) to determine the classification cut-offs. Our results show that hippocampus samples have a higher proportion of non-polyadenylated transcripts, including more non-polyadenylated protein-coding mRNAs compared to what was reported before for other human tissues and cell lines. Our study offers a rich source of data for exploring the poly(A)- landscape of transcriptome in the human aging brain, but also suggests the importance of understanding the potential pitfalls of each sequencing strategy and making informed methodological choices, especially when studying non-coding RNAs.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session I

PB4468 Characterizing the Natural History of *SYNGAP1*, *PCDH19*, and *CHD2* Disorders: Insights from the Ciitizen Database

Authors:

S. Sengupta[1], N. Chourasia[2,3], E. Bonkowski[4], H. Mefford[5]; [1]St. Jude Children's Res. Hosp., Memphis, TN, [2]Univ. of Tennessee Hlth.Sci. Ctr., Memphis, TN, [3]Le Bonheur Children's Hosp., Memphis, TN, [4]Univ. of Washington / St. Jude Children's Res. Hosp., Memphis, TN, [5]St. Jude Children`s Res. Hosp., Memphis, TN

Background:*SYNGAP1*-related intellectual disability, *PCDH19*-related epilepsy, and *CHD2*-related neurodevelopmental disorders are genetic disorders characterized by intellectual disability, epilepsy, and developmental delays. This study aims to conduct a comprehensive natural history analysis using electronic medical record data for *SYNGAP1*, *PCDH19*, and *CHD2*-related disorders gain insights into the clinical, demographic, and genetic characteristics of patients and understand the disease progression and outcomes.Methods:We use data generated through the Ciitizen®, a wholly owned subsidiary of Invitae Corporation, to acquire clinical information, genetic variants, demographic details, and longitudinal follow-up data of the participants. Descriptive analysis of the data summarizes the clinical and demographic characteristics, including age, sex, age of onset, seizure types, and comorbidities. Longitudinal analysis will investigate disease trajectories over time, assessing factors such as age at onset, seizure frequency, and cognitive development. Our ultimate goal is to further define the natural history to enable clinical trial readiness, to inform efficient treatment approaches, and to enhance clinical care for patients with these mutations.Results:We have obtained data for 205 patients (n=150 *SYNGAP1*, n=13 *PCDH19*, n=42 *CHD2*). Our preliminary analysis of 92 patients (median age 4 years) with known histories of epilepsy and primarily *de novo* *SYNGAP1* mutations show that more than 50% of pathogenic variants are SNVs, followed by deletions in 37%. For atonic seizures, we discovered a trend toward decreased mean seizure frequency over a period of 20 months following diagnosis. The downward trend is consistent with documented prescription of anti-seizure drugs. We are currently evaluating the association between age of seizure onset and developmental outcomes, longitudinal analysis of developmental milestone achievement, cognitive outcomes, and treatment response in to better understand the disease trajectory of *SYNGAP1*-related intellectual disability. We will perform parallel analyses for *PCDH19*- and *CHD2*-related disorders and subsequently perform comparative analyses across the disorders.Discussion:Comprehensive analysis of longitudinal medical record data will provide insights into the natural history and disease progression of *SYNGAP1*-, *PCDH19*-, and *CHD2*-related disorders. The implications for understanding disease mechanisms, identifying potential therapeutic targets, enabling clinical trial readiness, and informing clinical interventions will be discussed.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session II

PB4469 ClinGen Glaucoma Expert Panel curation guidelines improve MYOC variants classification.

Authors:

E. Souzeau[1], P. Graham[2,3], J. Hadler[1,4], S. W. J. Tompson[5], K. Whisenhunt[5], J. E. Craig[1], A. W. Hewitt[3], O. M. Siggs[1,6], S. Chakrabarti[7], J. D. Hulleman[8], F. Pasutto[9], T. L. Young[5], D. A. Mackey[2,3], K. P. Burdon[3], A. Dubowsky[4]; [1]Flinders Univ., Adelaide, Australia, [2]Univ. of Western Australia, Perth, Australia, [3]Univ. of Tasmania, Hobart, Australia, [4]SA Pathology, Adelaide, Australia, [5]Univ. of Wisconsin-Madison, Madison, WI, [6]Garvan Inst. of Med. Res., Sydney, Australia, [7]L V Prasad Eye Inst., Hyderabad, India, [8]Univ. of Texas Southwestern Med. Ctr., Dallas, TX, [9]Univ. Erlangen-Nuremberg, Erlangen, Germany

Aim and Methods: Pathogenic variants in Myocilin (MYOC) are the most common cause of Mendelian open-angle glaucoma. The Clinical Genome Resource (ClinGen) Glaucoma Variant Curation Expert Panel (VCEP) recently published rule specifications for MYOC variant classification. Here we present the results of the curation of MYOC variants using the newly specified rules by the Glaucoma VCEP. **Results:** The Glaucoma VCEP curated 268 MYOC variants: 9 were classified as benign (B), 36 as likely benign (LB), 193 as variants of uncertain significance (VUS), 22 as likely pathogenic (LP) and 8 as pathogenic (P). All LP/P variants were located in the conserved olfactomedin domain in exon 3. A previous ClinVar classification was reported for 31.0% of variants (83/268). Fifty-eight (70.0%) reached an overall classification concordance with ClinVar (B/LB vs VUS vs LP/P), while 7 (8.4%) had discordance with clinical implications (LP/P downgraded to VUS). The VCEP rules led to the reclassification of 10 variants from VUS to B/LB and the resolution of 8 variants with conflicting interpretations in ClinVar. The majority of the VUS did not have functional evidence available (93.8%, 181/193). **Conclusions:** The refined variant curation guidelines for MYOC from the Glaucoma VCEP have improved variant classification and increased the number of variants listed in ClinVar. Additional functional evidence may further improve the classification by decreasing the proportion of VUS.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session III

PB4470 Clinical picture of galactosemia type IV: Nationwide survey in Japan.

Authors:

Y. Saito[1], Y. Wada[2], N. Ichinoi[3], Y. Nakajima[4], S. Ajihara[5], K. Murayama[6], R. Kagawa[7], T. Tanaka[8], C. Numakura[9], T. Hamazaki[10], N. Igarashi[11], H. Esaki[12], T. Kono[13], T. Sawada[14], T. Sawada[15], H. Nyuzuki[16], H. Hirai[17], S. Fumoto[18], J. Matsuda[19], A. Matsunaga[20], S. Maruyama[21], K. Yamaguchi[22], M. Yoshino[23], A. Kikuchi[1], S. Kure[1]; [1]Dept. of Pediatrics, Tohoku Univ. Graduate Sch. of Med., Sendai, Japan, [2]Dept. of Pediatrics, Tohoku Univ. Sch. of Med., Sendai, Japan, [3]Dept. of Pediatrics and Child Hlth., Nihon Univ. Sch. of Med., Tokyo, Japan, [4]Dept. of Pediatrics Fujita Hlth.Univ. Sch. of Med., Toyoake, Japan, [5]Dept. of Pediatrics & Clinical Genomics, Saitama Med. Univ., Saitama, Japan, [6]Dept. of Metabolism, Chiba Children's Hosp., Chiba, Japan, [7]Dept. of Pediatrics, Hiroshima Univ., Graduate Sch. of BioMed. and Hlth.Sci., Hiroshima, Japan, [8]Dept. of Pediatrics, Natl. Hosp. Organization Hokkaido Med. Ctr., Sapporo, Japan, [9]Dept. of Pediatrics, Yamagata Univ. Sch. of Med., Yamagata, Japan, [10]Osaka Metropolitan Univ., Osaka, Japan, [11]Dept. of Pediatrics, Toyama Prefectural Central Hosp., Toyama, Japan, [12]Dept. of Pediatrics, Sasebo City Gen. Hosp., Sasebo, Japan, [13]Div. of Endocrinology and Metabolism, Saitama Children's Med. Ctr., Saitama, Japan, [14]Dept. of Pediatrics, Kumamoto Univ. Hosp., Kumamoto, Japan, [15]Dept. of Pediatrics, Izumi City Gen. Hosp., Osaka, Japan, [16]Div. of Pediatrics, Dept. of Homeostatic Regulation and Dev., Niigata Univ. Graduate Sch. of Med. and Dental Sci., Niigata, Japan, [17]Dept. of Pediatrics, Shikoku Central Hosp. of the Mutual Aid Association of Publ. Sch. Teachers, Shikokuchuo, Japan, [18]Dept. of Pediatrics, Kyorin Univ. Sch. of Med., Mitaka, Japan, [19]Dept. of Pathophysiology and Metabolism, Kawasaki Med. Sch., Kurashiki, Japan, [20]Dept. of Pediatrics, St. Marianna Univ. Sch. of Med., Kawasaki, Japan, [21]Dept. of Pediatrics, Kagoshima Univ. Graduate Sch. of Med. and Dental Sci., Kagoshima, Japan, [22]Dept. of Pediatrics, NHO Kokura Med. Ctr., Kitakyushu, Japan, [23]Dept. of Pediatrics, Kyushu Hosp., Japan Community Hlth.care Organization, Kitakyushu, Japan

Introduction: Galactosemia type IV (MIM# 618881, galactose mutarotase [GALM] deficiency) is caused by a lack of GALM, an enzyme responsible for catalyzing the interconversion between β-D-galactose and α-D-galactose in the Leloir pathway. Given the recent identification, there is limited knowledge regarding GALM deficiency. This study aims to investigate the prevalence, genotypic characteristics, symptoms, and prognosis of GALM deficiency.Methods: A nationwide survey on clinical profiles of GALM deficiency was conducted between February 2022 and March 2023, targeting 529 core hospitals to cover positive cases through newborn screening.Results: A total of 41 patients with GALM deficiency were identified. The estimated prevalence of GALM deficiency in Japan was 1 in 164,333 individuals. Two pathogenic variants, c.294delC and c.424G>A, accounted for 73% of all identified variants. Four patients (10.3%) presented with infantile cataracts, which were successfully resolved through lactose restriction without surgical intervention. There was no significant association between cataract formation and mean galactose levels before the onset of cataracts. During the neonatal period, liver dysfunction, portosystemic shunt, cholestasis, and hepatomegaly were observed in 22.5%, 7.9%, 5.0%, and 2.6% of patients, respectively. The liver dysfunction manifested as an elevation in transaminase levels, which occurred within the first three months of life and spontaneously improved by 1-year-old. The median values of AST and ALT were 135.5 U/L and 163.5 U/L, respectively. In this population, only one patient (2.4%) developed attention-deficit/hyperactivity disorder. No patient was diagnosed with developmental delay, unlike GALK1 deficiency. No patients experienced hypoglycemia, encephalopathy, ovarian dysfunction, or fractures. The median of maximum galactose levels recorded was 27.9 mg/dL, which was lower than that observed in GALK1 deficiency. All patients with GALM deficiency were started on lactose restriction, and 24.4% of them completed the restriction.Conclusion: Our study indicates that GALM deficiency exhibits a milder phenotype compared to GALK1 deficiency. Early detection of GALM deficiency would hold significance meaningful because lactose restriction has proven to be effective in treating cataracts possibly associated with GALM deficiency.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session I

PB4471 Cloud-scale training and education in the NHGRI Analysis, Visualization, and Informatics Lab-space (AnVIL)

Authors:

N. Kucher[1], M. Schatz[1], A. Philippakis[2], AnVIL Team; [1]Johns Hopkins Univ., Baltimore, MD, [2]Broad Inst. of MIT and Harvard, Cambridge, MA

As stakeholders in biomedical research increasingly adopt cloud computing, educational platforms and materials are needed to support the transition of genomic data science researchers to the cloud, for expert bioinformaticians and the next generation alike. Smoothly managing informatics training with software, data, and participant access as well as new elements relevant to the cloud, such as computing environment and billing are necessary for successful adoption.

The NHGRI Genomic Data Science Analysis, Visualization, and Informatics Lab-space or AnVIL, is a secure, cloud-based platform for storage, management, and analysis of genomic and related datasets that offers many opportunities for supporting genomics research training in the cloud. A major component of AnVIL is Galaxy, an interactive analysis platform that empowers users with and without programming experience to perform complex bioinformatics analysis from a user-friendly web interface. Galaxy in AnVIL offers participant and billing management, dedicated computing resources, and a consistent but customizable software environment all within a security perimeter necessary for use of protected and private datasets. These advantages result in faster jobs, reduced data transfer, tracked class activity and spending, and flexibility for in person and virtual training events.

Building on the rich resource of Galaxy training materials, we have developed scalable training and outreach content for AnVIL. Galaxy's extensive training network offers dozens of tutorials for the life sciences, including for genome assembly and variant analysis, functional genomics, proteomics, and epigenetics as well as other data intensive sciences such as climate research or machine learning. The AnVIL materials guide users in onboarding and explore the necessary building blocks for independent use of the AnVIL platform. In this presentation we will showcase the available AnVIL training resources, how they fit in the larger Galaxy Training Network, and summarize our experiences about lessons learned from onboarding users for using cloud resources.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session II

PB4472 Combining national health register data with biobank's genomic resources enables selection of healthy controls for any disease.

Authors:

K. Silander, N. Eklund, K. Kantojärvi, P. Laiho, M. Sauramo, A. Kyttälä, T. Wahlfors; Finnish Inst. for Hlth.and Welfare, Helsinki, Finland

Background: The Finnish Institute for Health and Welfare (THL) is an independent state-owned expert and research institute promoting the welfare and health of the population. THL has maintained the national Care Register for Health Care since 1967. This register contains diagnosis data derived from hospitals and primary health care of all people living in Finland. Since 1996, the diagnoses have been recorded with ICD-10 codes. THL Biobank hosts samples and data from more than 230000 sample donors. The research collections include population-based health examination surveys, disease-specific cohorts, and other research cohorts collected nationwide. Imputed genome-wide SNP data is available for more than 130000 donors. The population-based cohorts (N=45000) include data on demographics, lifestyle and health status, clinical data and samples (DNA, serum, and plasma) collected during a baseline clinical visit, NMR metabolomics data and GWAS data. These cohorts also include a harmonized set of variables created for the study of healthy aging. **Material and methods:** To build better research resources, we have created an internal database that contains ICD-10 diagnosis codes and dates for all biobank sample donors, derived from the Care Register for Health Care. The register-based diagnosis data can be combined with omics and survey data stored in the biobank to identify different subsets of healthy sample donors. **Results:** We tested our resource for three common disease categories: cancer, diabetes, and dementia. Among 22000 donors from population-based cohorts who have reached the age of 70 years, we identified 18000 sample donors without a diagnosis of cancer (ICD-10 codes C00-C99) by age 70 and 14000 donors without any diagnosis of cancer so far. There are 19000 donors without diabetes diagnosis (E08-E11, and E13) by age of 70, and more than 20000 donors without any diagnosis of dementia (F00, F01, F03) so far. **Discussion:** Using this unique resource, different criteria can be used to identify sets of confirmed healthy controls for any given disease. It is also possible to select healthy controls for a combination of several different disease categories. The identification of confirmed healthy controls can be expanded to include all biobank sample donors with genomic data (130000 donors) using various selection criteria.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session III

PB4473 Comprehensive molecular epidemiology of spinocerebellar degeneration in Japan based on J-CAT, a nation-wide patient registry

Authors:

Y. Hama[1], H. Date[1], H. Mizusawa[2], Y. Takahashi[1], J-CAT (Japan Consortium of Ataxias); [1]Natl. Ctr. Hosp., Natl. Ctr. of Neurology and Psychiatry, Kodaira, Japan, [2]Natl. Ctr. of Neurology and Psychiatry, Kodaira, Japan

Objective: Spinocerebellar degeneration (SCD) is a heterogeneous group of disorders presenting with ataxia. Approximately one-third of SCD patients are diagnosed as hereditary SCD in Japan. Many causative genes for SCD have been identified, particularly in whole genome/exome sequencing (WGS/WES) era. The aim of the study was to elucidate the comprehensive molecular epidemiology of SCD in Japan. **Methods:** From 2016 through May 2023, 2679 patients have been registered in J-CAT nation-wide with informed consent and 2197 subjected to mutational analysis. The number of patients with positive or negative family history was 1154 or 1043, respectively. Initial mutational analysis included PCR fragment analysis and repeat-primed PCR of SCA1, SCA2, MJD/SCA3, SCA6, SCA7, SCA8, SCA12, SCA17, SCA31, SCA36, DRPLA, and HD, followed by repeat-primed PCR for the pathogenic AAGGG or ACAGG repeat expansions in *RFC1* for 264 mutation-negative patients whose ages of onset were above 30 excluding those with autosomal dominant mode of inheritance (those who had an affected parent or children). WGS was conducted for 1058 mutation-negative patients and variant detection in 612 genes associated with ataxia has been completed in 838 patients. Candidates for pathogenic mutations identified in WGS were confirmed by Sanger sequencing method. **Results:** Mutational analysis established the diagnosis in 1048 patients (47.7%). The diagnostic rate of SCD patients with positive or negative familial history was 70% or 23%, respectively. In total, 41 genes were identified as causative in this cohort. The number of patients with major disease types were as follows: 311 with SCA31, 275 with SCA6, 184 with MJD/SCA3, 73 with DRPLA, 34 with SCA2, 31 with SCA1, 20 with SCA 36, 15 with SCA8, 14 with EA2, 13 with HD, 9 with SCAR8, 9 with CANVAS, 7 with SCA42, 6 with Spastic ataxia, Charlevoix-Saguenay type, 6 with SCA13, and 5 with SCA17. **Conclusions:** J-CAT has elucidated the latest and the most comprehensive evidence for molecular epidemiology of SCD in Japan. Next generation sequencing is a time and cost-effective method for genetic diagnoses, and an accurate diagnosis can result in better management. Even after WGS, a definitive diagnosis could not be provided for 52% of the patients. Further research is required for better diagnostic efficiency.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session I

PB4474 Custom genome assembly of KOLF2.1J and its benefit across neuronal cell types.

Authors:

P. Alvarez[1,2], B. Walenz[1], A. Rhie[1], E. Lara[1], A. Beilina[1], D. Patel[1], F. Grenn[1], K. J. Billingsley[1], R. Genner[1,3], E. Gustavsson[2], M. Ryten[2], M. R. Cookson[1], B. Paten[4], M. Jain[5], A. M. Phillippy[6], C. Blauwendraat[1]; [1]NIH, Bethesda, MD, [2]Univ. Coll. of London, London, United Kingdom, [3]Johns Hopkins Univ., Baltimore, MD, [4]UC Santa Cruz, Santa Cruz, CA, [5]Northeastern Univ., Boston, MA, [6]Natl. Institues of Hlth., Bethesda, MD

Objectives To generate and characterize a custom genome assembly from the KOLF2.1J cell line and explore its impact on the analysis of omics data generated from derived neuronal and glial cell types.
Background While induced pluripotent stem cells have gained popularity to study neurodegenerative diseases (NDDs), the heterogeneity of stem cells results in difficulty when comparing results. The iPSC Neurodegenerative Disease Initiative (iNDI) identified the KOLF2.1J cell line and prioritized its use as a reference line for studying NDDs due to its high performance and neutral genetically encoded risk for Alzheimer's Disease (AD). This line, and its derivatives expressing >100 variants related to AD and other dementias, is publicly available. Current genetic analyses are limited by use of the human reference genome for sequence alignments. We present here a telomere-to-telomere draft of the custom genome assembly for KOLF2.1J to enable more accurate investigation of NDD genomics.
Methods KOLF2.1J IPSCs were sequenced using PacBio HiFi, Oxford Nanopore Technologies (ONT) Ultra Long sequencing and Hi-C. Verkko/1.2 assembler drafted a custom genome with this data. In parallel, KOLF2.1J was differentiated into neurons and microglia, and sequenced using ONT whole genome sequencing and bulk RNA sequencing, Illumina bulk RNA sequencing, and CAGE-seq. DNA and RNA were basecalled using Guppy and mapped to hg38, CHM13, and the new KOLF2.1J assembly using Minimap2. Variant calling was performed with Sniffles2 and transcript discovery and quantification was performed with Stringtie and Talon. Methylation from the ONT long-read genome sequencing data was plotted using Modbamtools.
Results We generated a complete genome of KOLF2.1J, and identified many structural variants including some impacting the coding structure of genes. We identified that RNA mapping increased unique mapping with the KOLF2.1J assembly compared to hg38 and CHM13v2.0 for both short-and long-read data, indicating that using a custom assembly generates more accurate genomics results compared to general reference genomes. Methylation data showed clear differences across neuronal cell types including in cell specific marker genes (i.e.TREM2 for microglia). Furthermore, we identified several positions in the genome with significant haplotype-specific methylation differences.
Conclusion Here we describe the path towards precision iPSC technology by generating a custom genome assembly of a widely used cell line. We show that a custom genome assembly greatly improves the accuracy of several types of omic data and recommend this approach for commonly used cell lines.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session II

PB4475 *De novo* genome assemblies from two Indigenous Americans from Arizona identifies new polymorphisms in non-reference sequences

Authors:

C. Koroglu Altok[1], P. Chen[2], M. Traurig[1], C. Bogardus[1], L. Baier[1]; [1]NIH/NIDDK, Phoenix, AZ, [2]Jilin Univ., Changchun, China

The goal of improving disease prevention and treatment in all people generated a collective effort to increase diversity in human genetic studies, particularly among underrepresented populations. However, analysis of DNA sequence reads involves the initial step of alignment to the human GRCh38/hg38 reference genome, which does not adequately represent genomes of non-European ancestry. In this study we used *de novo* sequence assemblies to identify non-reference segments (NRSs) in genomes of an Indigenous American cohort from Arizona. These novel segments were then included in our whole-genome sequencing (WGS) variant calling pipeline. Two DNA samples (isolated from one female and one male) were sequenced using the HiFi SMRT method and the long sequence reads (mean length = 12 kb) were assembled using HiFiasm assembler. Each *de novo* assembly was compared to the hg38 reference genome using MUMmer software to identify NRSs > 100 bp. Each assembly included ~17 Mb of NRS which consists mostly of repeat elements. Forty NRSs totaling 240 kb were uniquely anchored to the primary assembly of hg38. These NRSs were enriched in gene regions with 17 segments positioned within introns. Among them, a 187 bp NRS that anchored 79 bp downstream of the *HCN2* third exon was the only instance where the reads in both *de novo* assemblies were identical. We then assessed the existence of the 187 bp NRS in 50 additional samples including those from other ancestries by Sanger sequencing; none of the five Caucasian genomes contained the *HCN2* NRS. In contrast, more than half of the Indigenous genomes (20 out of 34) and a few from Mexican and African American ancestries had the *HCN2* NRS. The *HCN2* NRS spans several putative transcriptional regulatory elements, and genotyping of this NRS in ~8,000 DNA samples to determine the frequency of the segment and any potential associations with metabolic disease in our Indigenous cohort is currently underway. In addition to characterizing the *HCN2* NRS, we are also determining the effect of all inserted NRSs on variant calling. We edited hg38 by inserting NRSs into their anchor sites. Alignment and variant calling using the GATK pipeline were performed on 387 WGS samples from our Indigenous cohort using either hg38 or the modified reference (hg38-NRS). Comparison of the annotated variants yielded >50,000 SNVs that were detected in at least 5% of the samples using hg38-NRS as reference which were not captured using hg38. In conclusion, this study showed that inclusion of population-specific NRSs dramatically changed the variant profile in an under-represented ethnic group, revealing common variation not detected by our previous population-level WGS and genotyping studies.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session III

PB4476 De-identification of clinical notes using open source software and commercial cloud-based NLP service together

Authors:

H. Qiu[1], F. Mentch[1], J. Glessner[1,2], H. Hakonarson[1,2]; [1]Children's Hosp. of Philadelphia, Philadelphia, PA, [2]Dept. of Pediatrics, The Perelman Sch. of Med., Univ. of Pennsylvania, Philadelphia, PA

Unstructured free text in electronic health records captures rich phenotypic information. In our research setting, studies and analyses are typically conducted on anonymized datasets. For genomic researchers that require review of free text clinical notes by data analysts, de-identification needs to be performed to remove protected health information (PHI).

To address this, we utilized the open source MITRE Identification Scrubber Toolkit (MIST) for identifying and redacting HIPAA identifiers in free-text clinical notes. Since MIST is a type of supervised machine learning based on conditional random fields (CRF) model, a gold standard reference corpus composed of custom local data is required for optimal model building. The reference corpus is typically curated by human annotators. Construction of such a gold standard corpus is laborious and time consuming. Instead of human annotators, we tried Amazon's cloud-based natural language processing (NLP) service - Comprehend Medical - to extract potected health information (PHI) from the selected reference dataset. Comprehend Medical's output json files were transformed into MIST annotation files (in mat-json format) using custom scripts. The result mat-json files were subsequently treated as human-gold annotations for the model building and evaluation.

By combining commercial NLP service for initial reference dataset annotations and open source MIST tool for additional clinical notes de-identification, we established a scalable, time-saving and cost-effective automated de-identification workflow that demonstrated excellent performance. Here we present in detail how we implemented this method and share our experience and lessons learned. We successfully applied this method in multiple internal research projects with good user feedback.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session I

PB4477 Developing a shareable HLA imputation reference panel for the Arab population

Authors:

H. Mbarek[1], C. Demay[1], G. Butler-Laporte[2], C. Saad[1], D. Velayutham[3], Z. Jan[3], P. Jithesh[3], Z. Calderone[4], S. Ismail[5], Qatar Genome Program Research Consortium; [1]Qatar Genome, Doha, Qatar, [2]Univ. of Oxford, Oxford, United Kingdom, [3]Coll. of Hlth.& Life Sci., Hamad Bin Khalifa Univ., Doha, Qatar, [4]American Red Cross, Holland Lab. for BioMed. Sci., Baltimore, MD, [5]Qatar Genome, Qatar Fndn., Doha, Qatar

The Human Leukocyte Antigen (HLA) complex is one of the most polymorphic regions of the human genome, showing significant diversity between different populations. Identifying HLA specific alleles is crucial for a wide range of research and clinical purposes, as the HLA system's genes are associated with susceptibility to over a hundred diseases, including infectious, inflammatory, and autoimmune conditions.
Although the genome-wide association studies (GWAS) and HLA allele imputation framework have revealed numerous associations between the HLA and human diseases, most of the shared imputation bases and tools have been generated from subjects of European origin.This method is known to be imprecise for individuals from ancestral groups poorly or not at all represented in imputation panels, running the risk of omitting potentially relevant HLA associations.
To solve it, we developed an HLA reference panel dedicated to the Arab population using 14,669 whole genome sequenced individuals from the Qatar Genome Program. After raw sequence quality control accounting for HG38 alternative HLA haplotypes, we designed a meta-algorithm combining HLA haplotype calls (IMGT/HLA v3.49) from three state-of-the art algorithms (HISAT-genotype, HLA-HD and T1K) to obtain high-quality HLA allele calls for 26 genes at a 3-field resolution (i.e. all protein-coding and synonymous variants). This includes all clinically relevant classical HLA genes, as well as non-classical HLA class I and II genes, including HLA-DRB3/4/5 and MICA/MICB. Then, these data trained a deep learning model for HLA genotype imputation using the cutting-edge HLA allelic imputation method DEEP*HLA. We obtained a shareable model that respects privacy requirements of both the reference and imputed samples.
Our reference panel and deep learning model are the first of their kind for the Arab population. They will help describe the genetic and phenotypic landscape of the HLA locus, improve understanding of HLA-associated diseases in Arab populations worldwide, and discover new alleles. It will also contribute to improving transplantation practices, mitigating or preventing transfusion-associated graft-versus-host disease, and discovering new host-pathogen interactions that may serve as potential therapeutic targets.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session II

PB4478 † Developing and sharing polygenic risk scores for 4,206 brain imaging-derived phenotypes for 400,000 UK Biobank subjects not participating in the imaging study.

Authors:

X. Yang[1], B. Li[1], Z. Fan[2], D. Ding[2], J. Shu[1], M. Ritchie[2], G. Nave[2], P. L. Michael[2], T. Li[3], H. Zhu[4], B. Zhao[2]; [1]Purdue Univ., West Lafayette, IN, [2]Univ. of Pennsylvania, Philadelphia, PA, [3]Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, [4]M D Anderson Cancer Ctr, Houston, TX

The UK Biobank's brain imaging data is an essential resource for clinical research, but its cost and difficulty in obtaining limit the imaging study to only 100,000 participants, leaving the majority of UKB subjects without imaging data. However, because imaging-derived phenotypes (IDPs) are heritable, and most UKB subjects have genetic information available, it's possible to predict IDPs for UKB subjects outside the imaging study using genetic data. To this end, this study systematically developed and evaluated biobank-scale genetic polygenic risk scores (PRS) for 4,206 IDPs from multiple brain imaging modalities and processing pipelines. The results indicate that the majority of IDPs (64.76%, 2,774/4,206) were significantly predicted by PRS developed by subjects with both genetic and imaging data. Moreover, genetically predicted IDPs showed associations with a wide range of complex traits and diseases, with the patterns being consistent across different imaging pipelines. These findings suggest that genetic prediction through PRS is a cost-effective and practical way to make the UKB imaging study more beneficial to a broader population. The PRS data resources developed in this study have been made publicly available through Zenodo and will be returned to the UK Biobank.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session III

PB4479 Development of a Korean imputation reference panel for imputing variants within the major histocompatibility complex locus

Authors:

C. Yu, K. Kim; Kyung Hee Univ., Seoul, Korea, Republic of

Genetic variants in the major histocompatibility complex (MHC) region are strongly associated with various human diseases. However, it is challenging to narrow down the causal variants mainly due to ancestry-specific long-range linkage disequilibrium in an extremely complex genetic architecture of the MHC region and the technical difficulty of typing HLA variants (= amino acids and classical alleles) and C4 variants (= C4 isotypes and structural variations) in large-scale populations. Here, our study aimed to construct a Korean MHC imputation reference panel for imputing HLA and C4 variants simultaneously, based on the neighboring SNP haplotypes. To this end, we obtained whole-genome sequencing data from 844 unrelated individuals and typed eight classical HLA genes (HLA-A, -B, -C, -DQA1, -DQB1, -DPA1, -DPB1, and -DRB1) at the G-group resolution by HLA-LA, which utilized a population reference graph. Additionally, we estimated the copy number of *C4*, its two alleles *C4A* and *C4B*, and viral sequences HERV within *C4* using a Gaussian Mixture Model by GenomeSTRiP. To determine the haploid-level copy number of these *C4*-related variants, haplotype clustering was conducted using a Gaussian Mixture Variational Autoencoder model with 733 pruned SNPs around *C4*. The haplotype clusters detected were assigned to the most likely haploid-level copy numbers through an iterative process to maximize the concordance rate between typed and estimated diploid-level copy numbers. Haploid-level HLA and *C4* variants were encoded in a binary format, along with the SNP genotypes using Beagle5. The Korean MHC imputation reference panel comprises 42,141 MHC SNPs, 570 HLA classical alleles at various resolutions, 1,399 HLA amino acid positions, and 17 binary-coded *C4*-related copy numbers. The reference performance was evaluated by measuring imputation accuracy through a leave-one-out cross-validation method. Haploid-level concordance rates were >99% for all eight HLA genes at the G-group resolution and >90% for all four *C4*-related copy number variations. In summary, we characterized the complex structure of *C4* and HLA genes in a Korean population and developed an MHC imputation reference panel that has great potential in future MHC association mapping studies in the Korean population. **Funding:** This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (NRF-2022R1A2C2092164).

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session I

PB4480 Development of Biobank Network for Research and Development of Precision Medicine in Japan

Authors:

S. Ogishima[1], Y. Murakami[2], T. Morisaki[3], Y. Goto[4], E. Noiri[4], Y. Asano[5], M. Morita[6], M. Yoshida[7], M. Muto[8], T. Tanaka[7], H. Nishiyama[9], H. Matsuoka[10], T. Kosho[11], K. Matsuda[3], T. Tomita[5], H. Nishihara[12], H. Nakae[13], J. Ikeda[13], H. Yokota[1], K. Ohneda[1], K. Noguchi[1], S. Nagaie[1], F. Nagami[1], T. Takagi[14]; [1]Tohoku Univ., Sendai, Japan, [2]Inst. of Med. Sci., Univ. of Tokyo, Tokyo, Japan, [3]Tokyo Univ., Tokyo, Japan, [4]Natl. Ctr. for Global Hlth.and Med., Tokyo, Japan, [5]Natl. Cerebral and Cardiovascular Ctr., Osaka, Japan, [6]Okayama Univ., Okayama, Japan, [7]Tokyo Med. and Dental Univ., Tokyo, Japan, [8]Kyoto Univ., Kyoto, Japan, [9]Tsukuba Univ., Tsukuba, Japan, [10]Kobe Univ., Kobe, Japan, [11]Shinshu Univ. Sch. of Med., Matsumoto, Japan, [12]Keio Univ., Tokyo, Japan, [13]CIBER, Tokyo, Japan, [14]Toyama Intl. Univ., Toyama, Japan

Biospeimen and data from quality-controlled biobanks are key for research and development in genomic medicine. In Japan, there exists over 50 biobanks, which stores over 1,210,000 biospecimen provided by over 730,000 doners. Biobanks hold quality-controlled biospeimen and data, but research institutions have had difficulty finding where to find the biospeimen and data they need for their own genomic medicine research and development. To address this issue, we established a Japanese biobank network connecting the three major biobanks (Tohoku Medical Megabank project, Biobank Japan, National Center Biobank Network and the university hospital's biobanks (Kyoto University, Tokyo Medical and Dental University, Tsukuba University, Okayama University, Kobe University, Shinshu University) in Japan. Our biobank network aims at research and development of operational support for promotion of utilization of biospecimen and data stored in biobank toward realization of genomic medicine. For this aim, we have developed a biobank crosssearch system on biospecimen and data stored in our biobank network. Research institutions can find biospeimen and data they need for their own genomic medicine research and development. To implement a cross-search service for biospecimens and data, we standardized the metadata of biobank holdings of biospecimen and data. The metadata was standardized by Minimum Information About BIobank data Sharing (MIABIS) in accordance with the biobank data model of the European biobank network BBMRI-ERIC. Now our biobank network stores 1,043,880 biospecimen and 321,112 molecular data provided by 583,773 donors. We also setup the web-based coordination system of fast access to biospecimen and data to meet the requests by academic/commercial users using the biobank cross-search system. Utilization across various biobanks will be promoted by our biobank crosssearch system and support of matching between academic/commercial users, and researches and developments toward realization of genomic medicine will be expected to be accelerated.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session II

PB4481 Development of the ClinGen Rheumatologic Autoimmune Disease Clinical Domain Working Group (RAD CDWG) as an exemplar of complex disease curation

Authors:

R. Shapira[1], C. J. Klein[1], I. M. Keseler[1], G. Cheung[1], O. A. Garcia[1], M. A. Fernandez Vina[2], T. R. Merriman[3], A. Tin[4], S. E. Plon[5], M. W. Wright[1], I. Aksentijevich[6], S. Mack[7], S. L. Bridges, Jr.[8], T. E. Klein[1,9], on behalf of ClinGen Rheumatologic Autoimmune Disease Clinical Domain Working Group; [1]Dept. of BioMed. Data Sci., Stanford Univ., Stanford, CA, [2]Dept. of Pathology, Stanford Univ., Stanford, CA, [3]Univ. of Alabama at Birmingham, Birmingham, AL, [4]Univ. of Mississippi Med. Ctr., Jackson, MS, [5]Baylor Coll. Med., Houston, TX, [6]NIH - NHGRI, Bethesda, MD, [7]Univ. of California San Francisco, San Francisco, CA, [8]Hosp. for Special Surgery, New York, NY, [9]Dept. of Med., Stanford Univ., Stanford, CA

The Clinical Genome Resource (ClinGen) is an NIH-funded consortium dedicated to building a central resource that defines the clinical relevance of genes and variants for use in precision medicine and research. Until recently, gene curation efforts have focused exclusively on monogenic disease. The ClinGen Rheumatologic Autoimmune Disease Clinical Domain Working Group (RAD CDWG) was established to be an exemplar of curating complex genetic disorders.

The RAD CDWG has adopted a structure unique among ClinGen CDWGs, comprised of two framework development subgroups in addition to curation expert panels. Within this integrated and iterative model, disease area experts inform the development of and pilot new frameworks which can then be generalized by ClinGen for other domains.

The RAD CDWG is also addressing monogenic gene-disease relationships with two newly established ClinGen Gene Curation Expert Panels (GCEPs): the Monogenic Autoinflammatory Diseases GCEP and the Systemic and Incomplete Lupus Erythematosus GCEP. Future plans include the development of an Autoinflammatory Diseases Variant Curation Expert Panel (VCEP) in 2024.

The Human Leukocyte Antigen Working Group (HLA WG) was established to create an HLA-disease curation framework and develop a tool that can take in and process the complex HLA nomenclature for curation of HLA-disease associations and HLA-drug interactions. This nomenclature presents a unique challenge to biocurators, particularly given its evolution over time and inconsistent allele naming in the literature. To address this challenge, an extensive list of "common" HLA alleles (persisting at frequencies >0.0001 across multiple populations) is being built into the new curation interface. In conjunction with detailed documentation for mapping identifiers in the literature to the built-in list, this functionality will minimize the need for biocurators with HLA-specific expertise and help ensure data quality.

To fully address the breadth of complex disease inheritance within the RAD clinical domain, the Multigenic Disease Taskforce was established in 2023 with the goals of developing 1) a framework for determining what loci are important to curate, 2) a framework for curation of causal gene(s) at a locus, and 3) guidelines for determining the potential clinical utilities of causal genes of multigenic diseases. Gout has been selected as the test phenotype, as an exemplar complex disease that is not significantly attributable to monogenic or HLA associations. In subsequent phases, the framework will be extended to and tested on other diseases, first within the RAD clinical domain, and ultimately other clinical domains.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session III

PB4482 Electronic phenotyping in the UK Biobank using the OMOP Common Data Model.

Authors:

D. Yang, W. S. Bush; Case Western Reserve Univ., Cleveland, OH

The UK Biobank (UKBB) is a large-scale registry containing medical, genetic, and lifestyle data from over 500,000 individuals from the UK, but its data is not in an easily usable format for research. Previously, an open-source tool called Delphyne was developed by the Hyve to perform extract, transform, and load (ETL) processing and to deploy the UKBB data using the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM). Each time investigators access UKBB data, distinct participant identifiers are assigned to assure participant anonymity, therefore existing OMOP instances of the UK Biobank are not usable across projects or institutions. The ETL process, a crucial part of the data conversion to OMOP, can be a lengthy and complex task if not facilitated by effective tools like Delphyne. Here, we investigate the feasibility and efficiency of applying Delphyne to our local version of the UKBB. We first adapted and applied the published ETL to our UKBB dataset to create an OMOP version, documenting specific UKBB files and code modifications needed. Using electronic phenotyping of aging conditions including Alzheimer's disease (AD) and chronic kidney disease (CKD), we compare data distributions from mapped clinical data to summary diagnoses released by the UKBB and other published analyses of these phenotypes from the UKBB. To further validate the accuracy of our ETL process, we incorporated additional rounds of iterative testing and refinements, ensuring our electronic phenotyping aligns well with the original UKBB data context. OMOP helped to facilitate easier organization and manipulation of large and dense UKBB phenotype information and allowed increased use and harmonization of the data for our research projects. Existing OMOP tools such as the Data Quality Dashboard and the ATLAS web interface are connected to the UKBB OMOP instance to provide easier quality control and data browsing capabilities. We developed a comprehensive tutorial that serves as a guide for other investigators intending to undertake a similar adaptation of their UKBB datasets. This tutorial outlines the steps, challenges encountered, and potential solutions, thus reducing the learning curve and promoting a wider application of the OMOP CDM in future UKBB research. This work streamlines the use of UKBB for more efficient and reproducible research, potentially accelerating the discovery of better insights in the realm of aging and chronic diseases. In addition, our work also highlights the potential of effective data transformation tools like Delphyne, enabling large-scale studies to glean meaningful interpretations from vast datasets.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session I

PB4483 † Empowering Researchers: Web-Based Repository for Tertiary Analysis of Public Bulk and Single-Cell RNA Sequencing Data.

Authors:

X. Zhang, D. Cheng, Z. Ouyang, Y. Zhu; BioInfoRx, Inc., Madison, WI

High-throughput RNA sequencing technologies (bulk and scRNA-Seq) have revolutionized the field of transcriptomics, providing insights into complex biological processes. The raw sequencing data files from primary data analysis are traditionally deposited to GEO or SRA databases. Recently, uniformly processed secondary analysis results from published RNA-Seq data (e.g., gene counts from ARCHS4 or recount2) have made reproducible analysis and massive data mining possible. However, tertiary data analysis, or interpreting the results in the context of the study, remains challenging for typical scientists without computational skills.
To overcome this challenge, we have developed web-based databases and data mining tools to enable scientists to conduct sophisticated tertiary analyses without requiring computational expertise. For bulk RNA-Seq data, >30,000 human and mouse projects encompassing >1 million samples from ARCHS4 are loaded into a web database. Researchers can easily access the projects and samples, utilize a streamlined workflow to perform differential gene expression analysis, and load the results in QuickOmics, a Shiny-based visualization system. QuickOmics includes a range of tertiary analysis tools such as PCA, gene expression plots, heatmaps, co-expression networks, functional enrichment, and pathway analysis.
For single-cell datasets, we collected published data reported by the authors and processed the results into h5ad format. In cases where important information, such as cell type annotation, is missing in public repositories, we have reached out to authors to obtain the required details. A standardized scRNA-Seq analysis pipeline has also been utilized to reprocess raw data for a subset of datasets. Currently, our database contains over 400 single-cell RNA-Seq datasets, comprising more than 17 million cells, with ongoing expansion. Researchers can explore these datasets visually using the interactive and scalable cellxgene VIP tool. This tool empowers researchers to gain insights into cell composition, gene expression profiles, and differentially expressed genes across cell types, utilizing over 20 commonly used plotting functions and high-level analysis techniques in single-cell research.
Comprehensive documentation is provided to assist users to fully utilize these tools. We will present several case studies where the databases and tools can quickly reproduce published results. The easy access to RNA-Seq tertiary analysis will foster collaboration, promote reproducibility and facilitate new discoveries.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session II

PB4484 Enhancing dbSNP and ALFA for Genomic Research: Commemorating dbSNP's 25th Anniversary

Authors:

L. Phan, W. Qiang, H. Zhang, G. Shekhtman, E. Moyer, L. Ziyabari, R. Villamarin, T. Hefferon; NIH, NLM/NCBI, Bethesda, MD

Genomic research has witnessed remarkable progress over the past 25 years, and this year marks the 25th anniversary of the Single Nucleotide Polymorphism Database (dbSNP). In light of this significant milestone, this presentation outlines crucial improvements to dbSNP and the Allele Frequency Aggregator (ALFA) databases to meet the evolving demands of researchers and clinicians in precision medicine. The proposed enhancements encompass data integration, quality control, advanced annotation tools, expanded population-specific allele frequencies, improved web search tools, and an enhanced user interface. By incorporating these updates, we aim to provide researchers with reliable and comprehensive genomic information, enabling robust genotype-phenotype associations and accelerating discoveries in disease genetics and personalized medicine. As we celebrate the 25th anniversary of dbSNP, this presentation not only highlights its legacy but also showcases how these improvements will revolutionize healthcare practices, empowering the scientific community to unlock the full potential of genomics.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session III

PB4485 Enhancing genomic understanding in the Taiwanese population: Individual-level Structural Variant profiling and HLA genotyping by WGS.

Authors:

J. Hsu[1], Y. Tseng[1], P. Chen[1], P-M. Chien[2], S-K. Lai[3], C-Y. Chen[2], P. Chen[1]; [1]Natl. Taiwan Univ., Taipei City, Taiwan, [2]Natl. Taiwan Univ., Taipei, Taiwan, [3]Natl. Taiwan Univ. Hosp., Taipei City, Taiwan

Accurately detecting genetic variations in Biobank-scale whole-genome sequencing (WGS) data is essential for precisely constructing a comprehensive variation profile and characterizing allele frequencies. This study focuses on improving genomic understanding in the Taiwanese population by addressing challenges related to individual-level structural variant (SV) profiling and human leukocyte antigen (HLA) genotyping. Previously published databases, such as gnomAD and HGSVC, have not adequately benchmarked SV calling using materials from the Genome in a Bottle (GIAB) consortium for their SV callsets. Moreover, the representation of Asians, including Taiwanese individuals, is limited in these databases. To overcome these challenges, we utilized the GIAB released workflow and callsets to evaluate nine SV callers, achieving an overall F1 score, precision, and recall of 0.85, 0.89, and 0.82, respectively, for large germline deletions. In parallel, this study compares WGS-based HLA genotyping algorithms, including HLA-VBSeq, Kourami, and HISAT-genotype. By analyzing miscall patterns using 800 samples with confirmed HLA genotypes, the study establishes the best available WGS-based HLA genotyping workflow and suggests complementary approaches for different HLA genes. We recommend using HISAT-genotype as the primary algorithm combined in-house checkpoints with HLA-VBSeq to reach at least 96% sensitivity for HLA class I genes at the two field resolutions. With an established WGS-based SV profiling and HLA genotyping workflow, individual-level SV and HLA genotypes can be resolved for samples from the Taiwan Biobank (TWB), facilitating improved clinical applications and personalized medicine. This research significantly enhances genomic studies in the Taiwanese population by integrating comprehensive SV analysis and improved HLA genotyping. The findings provide valuable insights into molecular diagnosis and the understanding of genetic variations, thereby offering essential information for healthcare applications in Taiwanese and Asian populations.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session I

PB4486 † Enhancing Molecular Diagnosis of Mendelian Disorders: An Automated Phenotype-Genotype Database Using Large Language Models

Authors:

Z. Chen[1], J. Cai[1], Y. Yang[2], Z. Zhao[1], G. Li[1], K. Xu[1], Q. Li[1], N. Wu[1]; [1]Peking Union Med. Coll. Hosp., Beijing, China, [2]Queen Mary Univ. of London, London, United Kingdom

Mendelian disorders affect millions of live births each year, underscoring the pressing need for molecular diagnoses provided by exome sequencing (ES). A typical ES analysis necessitates the interpretation of hundreds of variants, a task heavily reliant on phenotype-genotype databases such as The Human Gene Mutation Database (HGMD), ClinVar, and Online Mendelian Inheritance in Man (OMIM). However, these existing databases encounter two major challenges: 1) a lack of adequate phenotype-genotype associations; 2) an inability to accurately represent the probability of phenotype occurrences. To address these challenges, this study aims to establish an automated phenotype-genotype database of Mendelian disorders, capitalizing on recent advancements in large language models (LLMs). First, we developed a novel search strategy. Leveraging the similarity computation capabilities of LLMs, we employed semantic similarity searches instead of traditional keyword searches to identify all relevant literature on Mendelian disorders on Pubmed (a total of 34,957,128 articles). Using recall and precision as the evaluation metrics and HGMD as the "gold standard" (a total of 78,023 articles), semantic similarity search strategies notably outperformed keyword searches, yielding a relative increase ranging from 61.9% to 190.5% for recall, and -25.8% to 96.8% for precision. Second, leveraging the in-context learning and instruction-following abilities of LLMs, we developed a pipeline to efficiently extract phenotypes from the retrieved articles. When using recall and precision as the evaluation metrics, the LLMs-based pipeline successfully retrieved 79.7% of phenotypes with a precision rate of 90.3% in a manually curated test set of 208 articles. Here, we presented a prototype for a Marfan Syndrome phenotype-genotype database, developed through the LLMs-based strategies (Table 1, https://www.dropbox.com/s/eb4mzbxcsmb66fk/Table_1.csv?dl=0). Our approach has demonstrated enhanced precision and recall, as well as more accurate information extraction. This revealed considerable promise for the automatic construction of a phenotype-genotype database, and thus facilitating the molecular diagnosis of Mendelian disorders.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session II

PB4487 Expanding the GENCODE geneset - addition of "non-canonical" ORFs to the Human reference annotation

Authors:

T. Hunt, J. M. Mudge, J. Gonzalez, J. Loveland, A. Frankish, The GENCODE consortium; EMBL-EBI, Hinxton, United Kingdom

The GENCODE consortium produces detailed reference annotation of all human and mouse protein-coding genes, pseudogenes, long non-coding RNAs and small RNAs. Accurate gene annotation is of fundamental importance for genome biology and clinical genomics; annotation that is incorrect or incomplete impacts downstream analysis and introduces potentially significant errors.

Recent work on enhancing the GENCODE geneset has focused on the identification of novel ORFs, including potential translations within lncRNAs and the UTRs of protein-coding genes. Extensive translation occurs outside of 'canonical' human protein-coding genes which can be readily identified by Ribosome Profiling. The first stage of this community-driven work involved making a consensus set of Ribo-seq ORFs identified by seven recent experimental publications mapped to GENCODE annotations. This resulted in the identification of 7,264 Ribo-seq ORFs which were recently published (PMID: 35831657).

However whether these ORFs are truly functional or simply molecular "noise" is open to discussion. Ribosome profiling does not demonstrate protein existence, and we have thus far classified only ~50 Ribo-seq ORFs as protein-coding by using evolutionary methods to distinguish protein-level constraint. Furthermore, our high stringency conventional mass spectrometry analysis finds little proteomics support for Ribo-seq ORF proteins. To investigate further, we have performed an extensive bespoke analysis of immunopeptidomics datasets. We observe that large numbers of Ribo-seq ORFs generate proteins that are presented by the HLA system within cell-lines or cancer samples, raising their potential utility as biomarkers or therapeutic targets. Finally, we recognize that translation can instead have a purely regulatory function. Around half of our Ribo-seq ORFs are uORFs within protein-coding genes, and these are also of emerging clinical interest.

Thus, we see great value in producing a reference GENCODE annotation of 'non-normal' translation. We will discuss our efforts to expand our current "non-canonical" ORF catalog, with a community-focused approached, including the expansion of our 5' UTR annotation with state-of-the-art transcriptomics data to ensure that it encompasses putative uORFs; the development of a new evolutionary approach to infer their function; and the development of gold standard approaches to analyse Ribo-seq data.

Combined, this expansion and improvement of the GENCODE catalog will provide a gold standard resource for both basic genome biology research and clinical genomics applications.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session III

PB4488 Exploring Genetic Polymorphisms of Amino Acid Concentrations in a Community-Based Japanese Cohort of 10,000.

Authors:

R. Toki[1], S. Harada[1], S. Kojima[1,2], S. Fushiki[1,2], A. Hirata[1], M. Iida[1], N. Miyagawa[1], S. Edagawa[1], A. Miyake[1], Y. Sutoh[3,4], H. Ohmomo[3,4], Y. Otsuka-Yamasaki[3,4], A. Shimizu[3,4], T. Okamura[1], T. Takebayashi[1]; [1]Dept. of Preventive Med. and Publ. Hlth., Keio Univ. Sch. of Med., Tokyo, Japan, [2]Konica Minolta, Inc., Tokyo, Japan, [3]Div. of Bioinformation Analysis, Iwate Tohoku Med. Megabank Organization, Iwate Med. Univ., Yahaba, Japan, [4]Div. of Bioinformation Analysis, Inst. for BioMed. Sci., Iwate Med. Univ., Yahaba, Japan

Introduction: Plasma amino acid concentrations have been recognized as vital markers for various diseases and these are influenced by a complex interplay of genetic and environmental factors. Although genome-wide association studies (GWAS) have successfully identified genetic variants affecting plasma amino acid concentrations, most predominantly focus on European populations. Therefore, there is an unaddressed need for further exploration of these genetic associations in the Japanese population. **Objective:** Our study aims to identify genetic polymorphisms associated with plasma amino acid concentrations in a Japanese population, comprising 10,000 individuals. **Methods:** We genotyped 10,933 participants from the Tsuruoka Metabolomics Cohort Study (TMCS) baseline surveillance using the Axiom™ Japonica Array™ NEO (Thermo Fisher Scientific). Following quality control, we conducted cross-imputation with impute4 using the 3.5KJPNv2/1000 genome phase3 panel in line with our prescribed pipeline. Subsequently, we conducted a GWAS using BOLT-LMM (v2.3.6) with 20 common amino acid concentrations in plasma measured by capillary electrophoresis-mass spectrometry as traits. We then fine-mapped the obtained summary statistics and cross-referenced our lead single nucleotide polymorphisms (SNPs) with previous reports in the GWAS Catalog. **Results:** Association analysis of 11,584,837 SNPs identified polymorphisms significantly associated ($p < 5 \times 10^{-8}$) with 18 of the 20 plasma amino acid concentrations, excluding glutamic acid and leucine, and detected 85 significant locus-to-metabolite associations at 43 genes. Of these associations, 43 loci had no reported one-to-one association with the trait, and 14 loci had no reported association with any plasma metabolite in the previous metabolome GWAS. For instance, SNPs in *AGXT* and *MTHFR* were shown to be significantly associated with plasma serine concentrations for the first time in this study, and both were identified as genes encoding necessary metabolic enzymes involved in serine metabolism. In addition, polymorphisms commonly associated with multiple amino acid concentrations were detected at 15 loci where associations were noted, further suggesting the possibility of pleiotropy. **Conclusion:** Our study focused on genetic factors on plasma amino acid concentrations by identifying multiple genetic polymorphisms in the Japanese population. The unique findings, including potential novel mutations and pleiotropy, call for further validation and underline the need for more comprehensive studies in different ethnicities to fully understand the global genetic architecture of these traits.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session I

PB4489 ExtCodonIslands: A novel algorithm for extracting outlier regions in translational efficiency.

Authors:

M. Cloward, E. Nysetvold, P. Ridge; Brigham Young Univ., Provo, UT

Codon bias is the unequal distribution of synonymous codons across a gene or genome. Codon choice is subject to selection and is non-random. Codon choice is highly correlated with the most abundant tRNA. This is energetically desirable because the anticodons bind more efficiently to the codon during translation, unlike codons that do not bind in the wobble position. Codons may be labeled as efficient (i.e., their cognate tRNA is highly expressed) and translated quickly, or inefficient and translated more slowly. Groups of consecutive, inefficient codons appear in many genes. For example, a ramp sequence is consecutive inefficient codons at the 5' end of the coding region. Here we name outlier regions of efficient or inefficient codons "codon islands." Like ramp sequences, the presence or absence of codon islands is conserved. Efficiency-altering mutations in these codon islands may have phenotypic consequences even when the mutation is synonymous (e.g., inefficient regions may slow the ribosome to allow for protein folding that must occur before downstream translation). We present ExtCodonIslands, an algorithm to identify codon islands, and we describe the frequency, locations, and conservation of codon islands.
ExtCodonIslands uses codon efficiencies provided by the user or calculates efficiencies from a CDS file. Each codon in each sequence in the input file is assigned an efficiency. ExtCodonIslands uses a sliding window mean to remove noise and simulate ribosome movement along the transcript. Outliers are identified if window mean clusters exceed the interquartile range multiplied by plus or minus 1.5.
We ran ExtCodonIslands on 144,412,469 sequences from 69,224 archaea, bacteria, fungi, invertebrates, plants, protozoa, mammalian vertebrates, other vertebrates, and viruses. We identified 100,113,205 codon islands or 0.73 islands per transcript. Islands per transcript ranged from 0.29 in mammals to 1.04 in fungi. Islands were more abundant at the 5' and 3' ends of transcripts that contained them and less abundant in the center. Excluding viruses, the tool ran at about 1.7 seconds per 1000 transcripts on a standard laptop. Our study provides important insights into codon efficiency clustering patterns in all domains of life.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session II

PB4490 Extending Pan-UK Biobank to rare variants through systematic gene-based association testing across continental ancestry groups in the UK Biobank.

Authors:

J. Sealock[1,2], W. Zhou[1,2], K. Tsuo[2], M. Kanai[2], W. Lu[3], Y. Wang[1], Y. He[4], R. Gupta[5], R. Grant[2], P. Darnowsky[2], M. Solomonson[2], R. Walters[1], E. Atkinson[6], B. Neale[1], A. Martin[1], K. Karczewski[7]; [1]Massachusetts Gen. Hosp., Boston, MA, [2]Broad Inst. of MIT and Harvard, Cambridge, MA, [3]Broad Inst., Cambridge, MA, [4]Harvard Med. Sch., Boston, MA, [5]Harvard Med. Sch. / Broad Inst., Cambridge, MA, [6]Baylor Coll. of Med., Houston, TX, [7]Broad Inst., Medford, MA

Rare-variant association studies (RVAS) provide new opportunities to directly assess phenotypic effects of gene disruption by aggregating rare damaging variants, particularly predicted loss-of-function (pLoF) variants, within a gene. We previously conducted RVAS on thousands of phenotypes in the UK Biobank (UKB); however, analyses were performed using only individuals with European ancestry (EUR), thus limiting the detection of rare variants to a single ancestry. Here, we make use of the full UKB dataset to conduct RVAS on all genetic ancestry groups.
We performed gene burden testing in the two largest non-EUR ancestries in UKB, CSA (Central/South Asian; N = 8,471) and AFR (African; N = 6,281), using SAIGE-GENE+ on pLoF, missense, and synonymous variants. These analyses included some of the largest gene burden tests for a subset of phenotypes in these ancestries. In CSA, we discovered 284 significant associations among 468 phenotypes. Notable associations included changes in blood biomarker levels with pLoF variants in *HBB*, such as red blood cell count. We found 12 pLoF associations not seen in EUR, including an association between *HBB* and direct bilirubin. Analyses in AFR identified 575 significant associations across 353 tested phenotypes. We replicated a known association between decreased LDL levels and damaging variants in *PCSK9*. There were 20 pLoF associations not seen in EUR, including between *UGT1A3* and total bilirubin. Among significant hits, the correlation of effect estimates with EUR was 0.50 ($p = 2 \times 10^{-7}$) for CSA and 0.40 ($p = 2 \times 10^{-4}$) for AFR, suggesting that these effects are directionally consistent, but in some cases providing additional information on top of EUR effects.
Inclusion of non-EUR samples in genetic studies is imperative to maximizing scientific discoveries. While some of the analyses conducted here remain underpowered, the results represent some of the largest RVAS of non-EUR populations to date. Effect estimates were not perfectly correlated between EUR and AFR/CSA, and several associations not seen in EUR emerged, indicating the potential for uncovering new genes by using the full dataset. Our results illustrate the utility of incorporating diverse populations for RVAS and provide a step forward for creating equitable and diverse population genetics resources.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session III

PB4491 FinnGen PGS Browser: a platform for polygenic risk score analyses and interpretation from a cohort of 400,000 Finns.

Authors:

N. Kolosov[1], FinnGen, V. Llorens[2], T. P. Sipilä[2], M. Aavikko[2], M. P. Reeve[2], S. Ripatti[2], A. Palotie[2,3,4], M. J. Daly[2,3,4], M. Artomov[1,5]; [1]Nationwide Children's Hosp., Columbus, OH, [2]Inst. for Molecular Med. Finland (FIMM), Helsinki Inst. of Life Sci. (HiLIFE), Univ. of Helsinki, Helsinki, Finland, [3]Broad Inst. of Harvard and MIT, Cambridge, MA, [4]Massachusetts Gen. Hosp., Boston, MA, [5]The Ohio State Univ. Coll. of Med., Columbus, OH

Polygenic scores (PGSs) have become a vital resource for biomedical research due to the increasing availability of genome-wide association study results, genetic data, and electronic health records from large biobanks such as FinnGen. PGSs have a broad range of applications, such as predicting a patient's disease status, estimating the age of disease onset, identifying phenome-wide associations, or assessing shared etiology between phenotypes. However, for researchers and medical professionals to easily comprehend and utilize this information, it is crucial to present it in a clear and concise manner.

To address this need, we've created a "FinnGen PGS browser" - a web-based application that allows researchers to access a variety of PGS-based experimental results conducted in 392,649 participants and offers advanced filtering, visualization, and risk prediction capabilities.

We analyzed 129 polygenic score models (PGSs) derived from external genome-wide association studies (GWAS) and 4,995 traits from FinnGen (n=392,649). For each PGS model, we conducted a phenome-wide association study to investigate its association with 4,995 phenotypes for case/control status using logistic regression, and age of disease onset using Cox proportional hazards. On average, we observed 349 significant (p-value < $2*10^{-6}$) associations per PGS model for case/control status and 127 for age of onset. For each phenotype, we presented the corresponding odds ratios, hazard ratios, and area under receiver operating characteristic curve (ROC AUC) for all PGSs.

Additionally, we identified 15 phenotypes with ROC AUC values for PGS models exceeding 0.6 and trained corresponding Cox proportional hazards deep neural networks using polygenic scores and clinical data to provide the predictions for the external samples. The performance of these models was evaluated on test samples using the time-dependent concordance index (C^{td}-index), ranging from 0.63 to 0.89.

We envision that the PGS browser presented in this study will inspire future hypothesis generation for researchers and help facilitate further integration of PGSs into clinical practice.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session I

PB4492 † Gene Expression and Splicing QTL Analysis of Blood Cells in African American Participants from the Jackson Heart Study

Authors:

J. Wen[1], Q. Sun[2], L. Huang[3], L. Zhou[1], M. F. Doyle[4], L. Ekunwe[5], N. C. Olson[6], A. P. Reiner[6], Y. Li[7], L. M. Raffield[8]; [1]Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, [2]UNC-Chapel Hill, Chapel Hill, NC, [3]Univerisity of North Carolina at Chapel Hill, Chapel Hill, NC, [4]Univ. of Vermont, Burlington, VT, [5]Jackson Heart Study, Jackson, MS, [6]Univ. of Washington, Seattle, WA, [7]Univ North Carolina, Chapel Hill, NC, [8]UNC - Chapel Hill, CHAPEL HILL, NC

Most gene expression and alternative splicing quantitative trait loci (eQTL/sQTL) studies have been biased toward European ancestry individuals. Here, we performed eQTL and sQTL analysis using TOPMed whole genome sequencing-derived genotype data and RNA sequencing data from stored peripheral blood mononuclear cells in 1,012 African American participants from the Jackson Heart Study (JHS). At a false discovery rate (FDR) of 5%, we identified 4,798,604 significant eQTL-gene pairs, covering 16,538 unique genes; and 5,921,368 sQTL-gene-cluster pairs, covering 9,605 unique genes. About 31% of detected eQTL and sQTL variants with a minor allele frequency (MAF) $gt 1% in JHS were rare (MAF $lt 0.1%), and therefore unlikely to be detected, in European ancestry individuals. We also generated 17,630 eQTL credible sets and 24,525 sQTL credible sets for genes (gene-clusters) with lead QTL p $lt 5e-8. Finally, we created an open database, which is freely available online (http://jhsqtl.genetics.unc.edu/), allowing fast query and bulk download of our QTL results.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session II

PB4493 GeneMatcher and VariantMatcher, tools designed for phenotypic and genomic data sharing

Authors:

R. Paulo Martin[1], S. Griffith[2], C. Antonescu[3], L. Vail[3], J. Posey[4], Z. Coban-Akdemir[4], S. Jhangiani[5], K. Doheny[6], J. Lupski[4], D. Valle[2], A. Hamosh[7], N. Sobreira[8]; [1]Johns Hopkins Med., Baltimore, MD, [2]Johns Hopkins Univ, Baltimore, MD, [3]Johns Hopkins Sch. of Med., Baltimore, MD, [4]Baylor Coll. of Med., Houston, TX, [5]Baylor Coll. Med., Houston, TX, [6]Ctr. for Inherited Disease Res. (CIDR), Baltimore, MD, [7]Johns Hopkins Univ., Baltimore, MD, [8]Johns Hopkins Univ, Crownsville, MD

GeneMatcher (genematcher.org), is a freely available, web-based tool designed to facilitate connection of individuals around the world with interest in a specific gene with options to tailor matches using phenotype criteria. It has been used primarily to establish collaborations aimed to identify novel disease gene candidates. Over the last ~10 years it has been acknowledged in more than 747 publications describing more than 560 novel disease gene discoveries. GeneMatcher is also a founding member of Matchmaker Exchange (MME; www.matchmakerexchange.org), that allows GeneMatcher submitters to match genes across MME. As of June 1st, 2023, there are 17,132 GeneMatcher submitters from 105 countries who have made 74,420 submissions involving 15,392 unique genes, with 10,775 genes having > one match.VariantMatcher (variantmatcher.org) is also a freely available web-based tool developed to connect individuals around the world with interest in a specific variant. It enables sharing of de-identified variant-level and phenotypic data from research participants for discovery of disease-causing variants and genes. The database contains rare (MAF < 1% in gnomAD), nonsynonymous SNVs identified in 6,827 VCF files (957,249 unique variants) of affected and unaffected individuals sequenced as part of multiple projects along with their phenotypic information. Users are able to query the database upon registration and administrator approval. Queries are made on "chr:coordinate refAllele > altAllele" format using hg19 assembly (hg18 and hg38 will be lifted over to hg19 prior to matching). Matches are based on genomic coordinates, although phenotypic features can be added and shared in the email notifying the match. VariantMatcher is also connected to the Beacon Network through the Beacon protocol. As of June 1st, 2023, VariantMatcher had 901 submitters from 50 countries; 11,265 variants in 252 genes (224 of which are known disease genes). 281 variants have matched to 1 of more of 2,199 individuals. Among these 281 variants, 24 were classified as pathogenic or likely pathogenic in ClinVar, 155 were classified as variant of uncertain significance (VUS) and 42 were classified as benign or likely benign. To date, most of the matches helped to rule out the queried variant as causative for the phenotype being investigated.GeneMatcher and VariantMatcher have been widely used around the globe and have enabled researchers, health care providers, and patients to share their phenotypic and genomic data and build new connections to increase discovery rate of novel etiologies for diseases of unknown cause providing insights into both human biology and disease.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session III

PB4494 † Generation of a multi-ancestry Leukocyte Telomere Length resource in the Million Veteran Program

Authors:

K. Iyer[1,2], P. Vembu[2], R. Guarischi-Sousa[1,2], P. Billing-Ross[3,2], P. Tsao[1,2], S. Clarke[1,2], T. Assimes[1,2]; [1]Dept. of Med., Stanford Univ., Palo Alto, CA, [2]VA Palo Alto Hlth.Care System, Palo Alto, CA, [3]Dept. of Genetics, Stanford Univ., Palo Alto, CA

Telomeres, repetitive DNA sequences that protect chromosomal ends, shorten with age, and have been associated with several age-related chronic illnesses. Both genetics and environmental factors impact telomere length. However, telomere length has been predominantly studied in European populations, due to the scarcity of multi-ancestry biobanks that could facilitate such comprehensive investigations. This gap poses a great risk of exacerbating disparities in aging research.

The Million Veteran Program (MVP) presents a unique and unparalleled chance to bridge the existing gaps in telomere research concerning diversity and aging demographics. We leveraged the pre-existing sequenced data of 78,385 diverse MVP participants to generate the aging phenotype of leukocyte telomere length (LTL) using the widely implemented bioinformatics tool in the field - TelSeq. These participants with a mean age of 65 years, exhibited a demographic distribution of approximately 69% European ancestry, 24% African ancestry, 6% Hispanic ancestry, and 1% Asian ancestry. LTL estimates for this analytic sample set showed a median of around 2kb (IQR_{all} = 1.7kb - 2.3kb; $Median_{EUR}$ = 1.97kb, IQR_{EUR} = 1.67kb - 2.28kb; $Median_{AFR}$ = 2.03kb, IQR_{AFR} = 1.73kb - 2.36kb; $Median_{HISP}$ = 1.97kb, IQR_{HISP} = 1.64kb - 2.35kb; $Median_{ASN}$ = 1.86, IQR_{ASN} = 1.55kb - 2.3kb). Among these participants, we observed a statistically significant inverse correlation between age and estimated LTL (r= -0.26, p<2E-16), which was consistent across all ancestry groups (r_{EUR} = -0.27, p_{EUR}<2E-16; r_{AFR} = -0.23, p_{AFR}<2E-16; r_{HISP} = -0.21, p_{HISP}<2E-16; r_{ASN} = -0.21, p_{ASN}<2E-16). To further investigate this relationship, within each ancestry group we stratified the participants into 20-year age intervals and examined the correlation within each bin. We observed a marginally declining correlation with increasing two decades intervals, corroborating the fact that rate of telomere attrition decreases in older individuals.

Next steps include making the resource available to active MVP investigators for the (1) identification of genetic determinants of LTL, and (2) association of LTL with common chronic diseases encountered by veterans. Considering that majority of MVP samples have undergone sequencing using the Illumina's NovaSeq platform and processed exclusively at a single center, we expect minimal batch effects that could compromise the accuracy of our LTL estimates. Therefore, we plan to expand the scale of LTL data generation by including additional 75,000 samples. This expanded dataset will likely contribute to a more comprehensive understanding of LTL and its associations with diverse populations and aging.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session I

PB4495 Genome in a Bottle benchmarks in the era of complete human genomes

Authors:

J. Zook; Natl. Inst. of Standards and Technology, Gaithersburg, MD

Recent advances in genome sequencing, assembly, and polishing have made high-quality, nearly-complete diploid genome assemblies possible. The Genome in a Bottle Consortium (GIAB) is using these diploid assemblies to generate benchmark sets for evaluating accuracy of challenging small variants and structural variants. Previous versions of the GIAB benchmark sets relied primarily on variant calls from mapped reads. These mapping-based benchmark sets excluded regions with large variants and complex variants, excluding ~8% of the GRCh38 reference genome plus 7% of sequence missing from the reference. Initial exploratory work in assembly-based variant calling provided promising results, first with the development of a benchmark for the difficult, highly variable but medically relevant MHC region, then with a benchmark set targeting 273 medically relevant genes not sufficiently covered by our existing mapping-based benchmark sets. Building off this initial work we developed a framework for generating and evaluating assembly-based benchmarks sets. We form these benchmarks using the assembly-based variant caller dipcall, followed by excluding regions with potential assembly errors, alignment errors, and variant types problematic for current benchmarking tools. This framework enables the automated generation of small and structural variant draft benchmark sets for high-quality diploid assemblies aligned to multiple references (GRCh37, GRCh38, and T2T-CHM13). We are currently working with the GIAB community to evaluate draft assembly-based benchmarks for tandem repeats and for X and Y chromosomes, with the latter derived from complete chromosome assemblies. For the X and Y benchmark, we found it reliably identifies errors across a variety of high-quality variant callsets from different technologies, though we needed to exclude assembly errors in very long homopolymers as well as a complex gene conversion-like event containing the gene TSPY2 where assembly-assembly alignment is not standardized. As we include increasingly complex variants in the benchmark, we are also developing new tools to compare these variants, e.g., enabling robust benchmarking in tandem repeats. The GIAB and T2T Consortia are also working towards a near-perfect assembly with homopolymer errors corrected, as well as methods to form benchmarks for regions with complex variants. These new curated benchmarks will be valuable as the community moves towards variant calling in the most challenging regions, spurring development of increasingly accurate sequencing technologies and bioinformatics methods.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session II

PB4496 GenomeConnect: Patient completed health surveys can increase phenotypic data available in the public knowledge base

Authors:

C. Cooney[1], J. M. Savatt[2], D. R. Azzariti[3], H. L. Rehm[4], C. L. Martin[2], E. R. Riggs[5]; [1]Geisinger, Danville, NJ, [2]Geisinger, Danville, PA, [3]ClinGen Resource, Cambridge, MA, [4]Massachusetts Gen. Hosp., Boston, MA, [5]Geisinger, Huntsville, AL

ClinGen's patient registry, GenomeConnect, has recognized that patients can serve as an important source of genomic and health data. GenomeConnect collects data from genetic reports and health surveys from participants who consent to de-identified data sharing and recontact. All participants are encouraged to complete a general health survey; some are offered additional sub-surveys based on their responses. Responses from participant surveys are mapped to Human Phenotype Ontology (HPO) terms to enable structured data sharing. Detailed phenotypic information can be valuable when evaluating gene-disease validity and variant pathogenicity. GenomeConnect, in consultation with ClinGen's curation expert panels, designed 8 detailed sub-surveys to capture additional phenotypic data: developmental milestones, developmental services and family history, cardiomyopathy, seizures, arrhythmia, RASopathy, cancer, and heart defects. To assess the value of participant completed health surveys, we reviewed the number of completed surveys and HPO terms collected for all GenomeConnect participants. Of 4,588 participants, 3,459 (75%) completed the general health survey resulting in an average of 2 HPO terms per participant (0-258). Based on their responses, 2,236 (49%) participants were eligible to complete at least one sub-survey. Nine hundred and twenty-seven participants (41%) completed at least one sub-survey with a complete rate of: 35% developmental milestones and developmental services and family, 39% cardiomyopathy, 28% seizures, 45% arrhythmia, 7% RASopathy, 44% additional cancer, and 40% heart defect. Preliminary data suggest that the sub-surveys provide an additional HPO term per participant (range 0.8-8.5) though the number of potential HPO terms varies per sub-survey. GenomeConnect surveys also provide additional data that cannot be described using HPO terms, such as treatment, family history, age of symptom onset and diagnosis. The development of additional sub-surveys is advancing patient-provided phenotypic data collection, which can be used to inform gene and variant classification efforts including those of ClinGen expert curation panels. Further, GenomeConnect surveys are publicly available, and modified versions are being used by other advocacy groups thus increasing detailed phenotypic data collection. GenomeConnect participants continue to provide valuable genetic and health data to the public knowledge base. In future work, we plan to compare GenomeConnect's ClinVar submissions to those from other submitters and determine the number of variants where GenomeConnect is the only source of structured phenotype data.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session III

PB4497 † Genome-wide maps of regulatory enhancer-gene interactions across human cell types and tissues.

Authors:

K. Mualim[1,2]; [1]Stanford Univ., Stanford, CA, [2]Carnegie Inst. for Sci. - Dept. of Plant Biology, Stanford, CA

Identifying transcriptional enhancers and their target genes is essential for understanding gene regulation and the impact of human genetic variation on disease. Here we create and evaluate a resource of >13 million regulatory enhancer-gene interactions across 352 cell types and tissues, by integrating predictive models, measurements of chromatin state and 3D contacts, and large-scale genetic perturbations generated by the ENCODE4 Consortium. We first create a systematic benchmarking pipeline to compare predictive models, assembling a dataset of 10,411 element-gene pairs measured in CRISPR perturbation experiments, >30,000 fine-mapped eQTLs, and 569 fine-mapped GWAS variants linked to a likely causal gene. Using this framework, we develop a new predictive model, ENCODE-E2G, that achieves state-of-the-art performance across multiple prediction tasks, demonstrating a strategy involving iterative perturbations and supervised machine learning to build increasingly accurate predictive models of enhancer regulation. The ENCODE-E2G atlas of regulatory enhancer-gene interactions across the genome reveals global properties of enhancer networks, identifies differences in the functions of genes that have more or less complex regulatory landscapes, and improves analyses to link noncoding variants to target genes and cell types for common, complex diseases. By interpreting the model, we find evidence that, beyond enhancer activity and 3D enhancer-promoter contacts, additional features guide enhancer-promoter communication including promoter class and enhancer-enhancer synergy. Altogether, these genome-wide maps of regulatory enhancer-gene interactions, benchmarking software, predictive models, and insights about enhancer function provide a valuable resource for future studies of gene regulation and human genetics.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session I

PB4498 Genome-wide prediction of dominant and recessive neurodevelopmental disorder risk genes

Authors:

A. Zoghbi[1], B. Weido[1], J. Dhindsa[1], A. Shetty[1], C. Sands[1], S. Petrovski[2], D. Vitsios[3], R. Dhindsa[1]; [1]Baylor Coll. of Med., Houston, TX, [2]AstraZeneca, Fulbourn, United Kingdom, [3]AstraZeneca, CAMBRIDGE, United Kingdom

Despite great progress in the identification of neurodevelopmental disorder (NDD) risk genes, there are thousands that remain to be discovered. Computational tools that provide accurate gene-level predictions of NDD risk can significantly reduce the costs and time needed to prioritize and discover novel NDD risk genes. Here, we first demonstrate that machine learning models trained solely on single-cell RNA-sequencing data from the developing human cortex can robustly predict genes implicated in autism spectrum disorder (ASD), developmental and epileptic encephalopathy (DEE), and developmental delay (DD). Strikingly, we find differences in gene expression patterns of genes with monoallelic and biallelic inheritance patterns. We then integrate these expression data with 300 orthogonal features in a semi-supervised machine learning framework (mantis-ml) to train inheritance-specific models for ASD, DEE, and DD. The models have high predictive power (AUCs: 0.84 to 0.95) and top-ranked genes were up to two-fold (monoallelic models) and six-fold (biallelic models) more enriched for high-confidence NDD risk genes than genic intolerance metrics. Across all models, genes in the top decile of predicted risk genes were 60 to 130 times more likely to have publications strongly linking them to the phenotype of interest in PubMed compared to the bottom decile. Collectively, this work provides highly robust novel NDD risk gene predictions that can complement large-scale gene discovery efforts and underscores the importance of incorporating inheritance into gene risk prediction tools (https://nddgenes.com).

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session II

PB4499 Genomic and Clinical Correlates of Plasma Urate in the "All of Us" Cohort

Authors:

O. Shchelochkov[1], A. Awan[2], T. Ferrara[2], S. Goleva[2], D. Schlueter[3], J. Keaton[4], J. Denny[5], H. Mo[2]; [1]Purine and Pyrimidine Metabolism Unit, NHGRI, Bethesda, MD, [2]Cohort Analytics Core, NHGRI, Bethesda, MD, [3]Cohort Analytics Core, NHGRI, BETHESDA, MD, [4]Ctr. for Precision Hlth.Res., NHGRI, Bethesda, MD, [5]All of Us, NIH; Cohort Analytics Core, NHGRI, Bethesda, MD

Variations in blood urate are associated with clinical, genomic, and environmental factors, but their causal interactions require further clarification. Using a population-scale dataset ("All of Us"), we explored causal relationships between urate levels, genomic correlates, and co-morbidities. The All of Us cohort currently includes 245,400 participants with genome sequencing, demographics surveys, and electronic health records, including lab measurements. Blood urate level was extracted from electronic health records, formatted and standardized for analysis using the QualityLab pipeline (PMID: 33441150). Participants with end-stage renal disease, blood cancers, or urate levels exceeding 3 standard deviations from the mean were excluded. 25,261 participants met the inclusion/exclusion criteria. Using a trans-ancestry GWAS analysis to identify polymorphisms associated with urate levels, we replicated several known GWAS associations. Blood urate was associated with SNPs in *SLC2A9* (rs13115469-C, beta = 0.39 mg/dL, p = 1.24e-119); *ABCG2* (rs1481012-G, beta = 0.22 mg/dL, p = 1.18e-19); *SLC22A12* (rs524023-T, beta = 0.12 mg/dL, p = 6.95e-15); and *SLC17A1* (rs1165154-C, beta = 0.11 mg/dL, p = 1.17e-12). We then used PheWAS to explore the association between clinical phenotypes and urate level (direct-PheWAS). After Bonferroni correction, 387 phecodes with significant associations were identified, of which 30 were related to the cardiovascular system. We then used the genetic risk scores (GRS) calculated from the four identified SNPs to perform GRS-PheWAS as a comparison. Only gout and its sub-phecodes were significant in the GRS-PheWAS, of which none were in the cardiovascular system. Urate was associated with gout with an OR = 1.70 (1.64 - 1.76) per 1 mg/dL increment in the direct-PheWAS, while the uric acid level estimated by GRS had OR = 3.01 (2.58 - 3.51) per 1 mg/dL increment. We conclude that genetic determinants of urate had a larger effect size on the risk of gout than the actual urate measurements; and that other, likely non-genetic factors linked to urate have a larger effect size on the risk of cardiovascular diseases. Replication of these findings are required in an independent cohort.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session III

PB4500 † Genomic, Epigenomic, and Functional Genomic Characterization of Mouse Neuronal Cell Lines.

Authors:

J. G. Manuel, H. B. Heins, J. K. Ng, T. N. Turner; Washington Univ. Sch. of Med., St. Louis, MO

Neuronal cell lines are used to study the function of genes and variation. In-depth molecular characterization of these cell lines is crucial to research. In our study, we focus on two commonly used mouse neuronal cell lines called HT-22 and Neuro-2a. These lines have not been studied in-depth for genomic, epigenomic, and functional genomic features. Characterization of the cell lines will help determine genomic variation, gene expression, methylation, open chromatin regions, and chromatin interactions. These can be beneficial in prioritizing genes and noncoding regions that can be studied in a specific cell line. We characterized the HT-22 and Neuro-2a cell lines by performing karyotype, Illumina whole-genome sequencing (WGS), PacBio WGS, Hi-C, ATAC-seq, Illumina RNA-seq and PacBio Iso-Seq. Karyotyping, Illumina WGS, and PacBio WGS were used to identify the genomic variation within these cell lines with PacBio WGS also providing information on 5mC methylation. PacBio Iso-Seq was used to generate full length RNA isoforms, and Illumina RNA-seq was used for quantification of expression levels. ATAC-seq was used to identify regions of open chromatin, and Hi-C identified chromatin interactions. By using the methods previously stated, we were able to generate an in-depth characterization of each cell line with long-read and short-read sequencing data. Our unpublished data is available through NCBI BioProject (PRJNA938057) with a UCSC genome browser trackhub and other information available at https://github.com/TNTurnerLab/HT22_genome_epigenome_functional_genomics_project. This type of characterization had not been done before for these cell lines even though they have been used extensively in studies of genes and therapeutics. One particularly innovative aspect of our work is our long-read WGS (98X coverage) of the Neuro-2a cell line using the newly released PacBio Revio sequencer. In our sequencing of Neuro-2a on three Revio SMRT cells, we identified “expected” coverages of at least 30X with 30X, 32X, and 36X genome coverage per SMRT cell, respectively. Variant calling was highly consistent across all three SMRT cells with >98% concordance comparing SMRT cells for small variants and >88% concordance for structural variants. The correlation of methylation features was $r > 0.94$ in comparisons of the three SMRT cells. This is just one interesting aspect of our work and in our presentation, we will discuss several others. In-depth characterization of the HT-22 and Neuro-2a cell lines will be a resource to other researchers to identify whether their gene or noncoding regions of interest can be studied using these cell lines as a model.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session I

PB4501 Genomics 2 Proteins portal: A discovery tool to link genetic screening outputs to protein sequence and structure

Authors:

S. Iqbal[1], J. Safer[1], D. Nguyen[1], S. Kwon[1], P. May[2], D. Hoksza[3], A. Campbell[1], A. Burgin[1]; [1]Broad Inst. of MIT and Harvard, Cambridge, MA, [2]LCSB, Esch-belval, Luxembourg, [3]Charles Univ., Prague, Czech Republic

We live in the era of big biological data where on the one hand, deep learning methods and experimental techniques have made millions of high-quality protein structures accessible to the biomedical community, and on the other hand, an unprecedented number of genetic variations in the general population and those associated with monogenic and complex diseases have been identified and accumulated in multiple databases. Efficient computational tools are needed to connect these multi-omics data of disparate types (e.g., genomic vs. transcripts vs. protein structural data), for a holistic interpretation of the molecular effect of inherited clinical and population genetic variations, as well as those generated by functional genomics approaches (e.g., base editing, multiplexed assay). Toward this goal, we developed an integrated bioinformatic method established by programmatic access and linking of APIs to dynamically query, retrieve, and connect genetic variations and transcripts to protein sequence annotations and structures, wrapped within an interactive web interface, called Genomics 2 Proteins portal (G2P; g2p.broadinstitute.org). Using the method, we mapped over 18,000,000 human genetic variations from all protein-coding genes (20,271 genes; 49,675 Ensembl and 53,389 RefSeq transcripts) onto protein sequences and structures, with comprehensive annotations such as physicochemical properties of amino acids, structural information, functional features from UniProt, post-translational modifications, and readouts from multiplexed assays of variant effect. By exploiting experimentally solved (Protein Data Bank) and predicted (AlphaFold database) protein structures, G2P covers 99% of human proteins corresponding to 75,573 structures. Additionally, the portal includes an “Interactive Module” that allows users to securely upload genetic variation annotations at the protein residue level along with other protein sequence annotations, e.g., domains, drug-binding pockets, and conservation scores, and map them over to the target protein’s structure, essentially generalizing this capability of linking genomics to proteins beyond the genetic variations and protein structures available in the databases. The portal will serve as an easy-to-use, discovery tool for researchers and scientists from different subfields of biology and medicine to understand the relationship between genetic variations and their molecular phenotype, i.e., how variations map to protein structures, modulate structure-function relationship and cause diseases, and thereby design better therapeutics.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session II

PB4502 GWAS SVatalog: a visualization tool to aid fine mapping of GWAS loci with structural variations.

Authors:

S. Chirmade, Z. Wang, S. Mastromatteo, B. Thiruvahindrapuram, D. Roshandel, T. Nalpathamkalam, F. Lin, K. Keenan, J. Whitney, S. Allana, L. Strug; The Hosp. for Sick Children, Toronto, ON, Canada

Structural variants (SVs) are large polymorphisms that range from 50 bp to several megabases. These genomic modifications have been shown to impact different diseases through a variety of mechanisms by affecting gene expression, regulation and function. SVs have been notoriously difficult to genotype but advances in sequencing methods, particularly the emergence of long-read technologies, have progressed the understanding of genome-wide characteristics and frequencies of these alterations. Genome-wide association studies (GWAS) conventionally use single nucleotide polymorphisms (SNPs) and have significantly enhanced our understanding of contributing loci and the disease model for common traits. However, SNPs explain only a small percentage of phenotypic variation and may be tagging other types of polymorphisms that could fill this gap in variation.

Motivated by a GWAS of cystic fibrosis (CF) disease severity, we identified a suggestive locus tagging a SNP-SV pair, rs10273639 and 20-kb deletion, and discovered its contribution to the risk of intestinal obstruction at birth (meconium ileus) in CF. We hypothesized that other disease-associated loci reported in the literature are tagging unknown or unexplored SVs and could be the cause of GWAS signals.

We developed GWAS SVatalog, an open-source web tool that uses GWAS Catalog's vast amount of trait-associated SNP data to visualize SNP-SV pairs in high linkage disequilibrium (LD). Using 101 samples sequenced by Pacific Biosciences long-read technologies, we merged SVs from two different calling pipelines, pbsv and Sniffles2, and combined them with SNP data. This workflow was benchmarked using Genome in a Bottle's HG002 public datasets. LD between SVs (MAF > 0.1) and GWAS Catalog SNPs was then calculated using SHAPEIT4, generating R2 and D' statistics.

We found 35,855 SVs to be in high LD (D' >= 0.9) with 64,955 GWAS SNPs, accounting for over 26% of the total SNPs found in the GWAS Catalog. This suggests the impact of SVs at GWAS loci requires further bioinformatic and functional investigation to determine whether the SVs explain the GWAS signal.

GWAS SVatalog allows the user to visualize the SNP-SV relationships at a specific locus in the form of an interactive plot by filtering for their specified criteria (genomic region or phenotype). The corresponding GWAS Catalog, SV and LD data can also be downloaded by the user for subsequent analysis.

Understanding the existence of common SVs at GWAS loci will provide a source of putative causal polymorphisms that have yet to be explored. GWAS SVatalog will facilitate this understanding, guiding future functional investigation and disease mechanism studies.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session III

PB4503 High resolution characterization of known pathogenic repeat expansion loci in over 1,000 long-read genomes

Authors:

I. Xu[1], M. Danzi[1], S. Fazal[1], E. Dolzhenko[2], M. Eberle[2], AllOfUS Long Reads Working Group, S. Zuchner[3]; [1]Univ. of Miami, Miami, FL, [2]Pacific BioSci.s, Oceanside, CA, [3]Univ. of Miami Miller Sch. of Med., Miami, FL

Tandem repeats (TRs) are highly polymorphic, repetitive stretches of DNA motifs. Expansion of TRs cause over 50 diseases, many with neurological symptoms. Due to the limits of short-read genome sequencing in resolving large tandem repeats and a lack of normative genome-wide high-quality data in control populations, TRs are largely understudied. TRs are likely implicated in many more disease phenotypes. Working as part of the Long Reads Working Group within the All Of Us Research Project program, we have genotyped over 1.7 million TR loci in 1,027 phased control genomes using long-read PacBio Hifi sequencing. This resulted in an unprecedented three billion TR allele sequences, complete with length, motif composition, motif purity, and flanking sequence variation. Here, we characterize the ~50 known pathogenic TR loci in this control cohort. We show the distributions of repeat lengths, motif compositions, repeat purity, and flanking sequence variation, as well as estimates of premutation allele frequencies and carrier frequency for autosomal recessive disorders. We utilize the RFC1 locus as an example to showcase the extreme length and motif polymorphism present at some pathogenic TR loci which can be studied with this dataset. This data resource will aid in the identification of new pathogenic TR expansions.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session I

PB4504 Highly faceted scientific management model catalyzes large scale team science to help build the future of genomic medicine.

Authors:

L. Antonacci-Fulton, S. Cody, T. Lindsay, H. Lawson, T. Wang; Washington Univ., St. Louis, MO

Washington University School of Medicine has a long history in the field of genomics contributing to the generation of the original human genome project and it is now home to organizational centers that drive large-scale, team-based genomic research.

• Human Pangenome Reference Consortium (HPRC) - Building human pangenome resources will lead the scientific and clinical communities from a linear human reference to a reference comprised of many genomes which better represent human variation, leading to improved outcomes in the clinic.

• Impact of Genomic Variation on Function (IGVF) - Study of genomic variation on function leading to the creation of a catalog of results and resources utilized to understand phenotypes

• Somatic Mosaicism Across Human Tissue (SMaHT) - a systemic study of genetic differences within the body to create a catalog of variation to understand how they impact biology and disease.

HPRC will allow for the development of a multi-allelic reference comprised of many genomes that better define diversity. This new reference landscape along with information about somatic variation will allow for a broader investigation of genomic variation on function. The resulting information will be a full collection of functionalized germline and somatic variations.

Funded components comprise the consortia, including an organizational unit, data producers, analysis core, and groups funded for tool and technology development. Resources are regularly cataloged and made available to the scientific community. Outreach activities include publications, presentations, public-facing websites, and social media. The organizational framework for each consortium includes a steering committee and working groups focused on specialized work streams.

Our organization center employs Wiki software, allowing for the collaborative organization and management of information and resources generated by the consortium. Google and Box store and retain files. Communication tools enhance daily interactions and involve SLACK and email. Many consortia also have mid-year virtual and annual, in-person meetings that allow the consortia researchers to come together to take stock of their current work and make plans for the future.

Centralizing activities within a multi-consortium support center yields the quick start of consortiums, an operational economy of scale, quick adoption of new organizational methods and policy changes, systematic in-reach and outreach methods, and the straightforward cross-pollination of systems, ideas and interactions between the consortia.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session II

PB4505 hipFG: Harmonization and Integration Pipeline for Functional Genomics

Authors:

J. Cifello, P. P. Kuksa, N. Saravanan, O. Valladares, Y. Leung, L-S. Wang; Univ. of Pennsylvania, Philadelphia, PA

Functional genomics (FG) data such as genomic interactions, quantitative trait loci (QTLs), and epigenetic markers provide insightful biological context to genome-wide association study (GWAS) hits. However, the inclusion of such FG data within genomic workflows is often hindered by inconsistent standards, file formats, and metadata vocabularies. For example, effect sizes for expression QTLs will correspond to alternate alleles in some studies but tested alleles in others. We developed the Harmonization and Integration Pipeline for Functional Genomics (hipFG) to normalize heterogeneous FG datasets via standardization of both data and metadata, and organization of the standardized outputs for rapid datatype-specific indexing and access. Applicable FG datasets may include chromatin interactions, QTLs, and other annotated genomic intervals. This is achieved via automatic generation of customized datatype-specific pipelines. For example, QTLs are corrected to have effect sizes always with respect to non-reference alleles, and chromatin interactions are reformatted such that both anchors of each interaction are searchable.

Harmonized FG outputs have uniform (BED) file standards and are organized into output folders by assay, datatype, and more. These FG outputs are compatible with rapid query of genomic regions of interest via tools such as Bedtools, Giggle, and BEDOPS. hipFG pipelines are scalable and executable sequentially or in parallel. Lastly, hipFG generates standardized metadata describing FG outputs, expanding upon a user-provided minimal data descriptor table. To test hipFG, we processed three disparate data sources -- all of eQTL catalogue, EpiMap, 3DGenome 988 datasets, >18B records -- to functionally annotate and prioritize 10,823 variants of genome-wide or suggestive significance from the Bellenguez, et. al. '22 Alzheimer's Disease GWAS. This yielded a 1 MB locus at chr17 (q21.31) with the most variants-tissue-assay combinations of all tested loci.

In conclusion, hipFG allows users to quickly integrate genome analysis workflows with diverse FG datasets spanning many cell types, tissues, and assays. hipFG is freely accessible at https://bitbucket.org/wanglab-upenn/hipFG/.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session III

PB4506 How Open Targets resources use Genetics data to facilitate target identification and prioritisation

Authors:

A. Buniello[1,2], D. Suveges[1,2], C. Cruz[1,2], M. Bernal[1,2], P. Uniyal[1,2], I. Lopez[1,2], J. Xiangyu[1,2], Y. Tsepilov[1,2], D. Considine[1,2], H. Nicholls[1,2], K. Tsukanov[1,2], C. Mehta[1,2], L. Fumis[1,2], R. Martinez Osorio[1,2], E. Barkhordari[1,2], H. Cornu[1,2], S. Horswell[1,2], S. Young[1,2], M. Ghoussaini[3,2], D. Ochoa[1,2], D. Hulcoop[4,2], I. Dunham[1,2], E. McDonagh[1,2]; [1]EMBL-EBI, Hinxton, Cambridge, United Kingdom, [2]Open Targets, Cambridge, United Kingdom, [3]Wellcome Trust Sanger Inst, Cambridge, United Kingdom, [4]GSK, Stevenage, United Kingdom

Open Targets is an innovative industry-academia partnership that uses human genetics and genomics data to facilitate systematic drug target identification and prioritisation. Its two flagship open source bioinformatic resources are Open Targets Genetics (https://genetics.opentargets.org/), identifying targets based on GWAS and functional genomics and the Open Targets Platform (https://platform.opentargets.org/), integrating public domain data to enable target identification and prioritisation. Genetics evidence is of key importance for identifying and prioritising targets for novel drug development, as drugs developed to modulate target-disease associations with supporting genetic evidence are more likely to succeed; however, the challenge arises with connecting evidence from all genome wide association studies (GWAS) loci to the likely disease-causing genes in a systematic way. Open Targets Genetics was established as a resource to address this challenge and provide post-GWAS analysis for the scientific community. It brings together data from 36,224 human GWAS from the GWAS Catalog, FinnGen and UK Biobank with functional genomics datasets (eQTLs, pQTLs and sQTLs). 3.7 million colocalisation tests are performed and made available for GWAS-GWAS and GWAS-quantitative trait loci. These large scale genetics and functional genomics datasets feed into a machine learning model, developed to provide a predicted locus-to-gene (L2G) score to help prioritise the most likely disease-causing genes from GWAS. Open Targets Genetics is a key source of genetic evidence that is integrated with >20 different data sources feeding into the Open Targets Platform to provide target-disease associations for the context of drug discovery. As well as L2G predictions for common traits from Open targets Genetics, rare disease and somatic variation evidence from clinical resources such as ClinVar, PanelApp and ClinGen are integrated, scored and combined into the Platform to provide extensive genetic evidence for target-disease associations. Gene Burden test results that aggregate rare and ultra-rare variants at the gene level to provide gene to phenotype associations. The Platform integrates a comprehensive set of burden test results that aggregate rare and ultra-rare variants at the gene level to provide gene to phenotype associations. We have also shown that two thirds of drug approvals have underlying genetic evidence for the target-disease hypothesis. We believe these systematic, open source approaches address a fundamental need for the community to utilise rare and common variation results for target discovery.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session I

PB4507 Hypotonia: The Co-Existence and Patterns of Human Phenotype Ontology Terms in Neonatal and Pediatric Patients.

Authors:

Z. Liu[1], M. K. Higashi[1], y. liu[1], K. Wain[1], N. P. Melanie[1], P. Kruszka[1], A. scheurer-Monaghan[2]; [1]GeneDx, Gaithersburg, MD, [2]Western Michigan Univ., Kalamazoo, MI

Background: Congenital hypotonia is a relatively common diagnosis in the newborn period. Diagnosing an underlying cause for neonatal hypotonia is challenging since the classic presentation of "floppy infant" is associated with a large differential diagnosis and a myriad of tests and procedures for evaluation. Recent studies have concluded that exome sequencing as a first-tier test for neonates with unexplained hypotonia is warranted, but this has not been implemented universally into clinical care. Ontology-based search methods interrogating phenotypes associated with hypotonia during and beyond the neonatal period may yield new insights into clinical outcomes. Intuition into phenotype associations over the lifespan may aid the clinical adoption of exome sequencing as a first-tier test for neonatal hypotonia.**Objective:**This study's purpose was to develop a knowledge graph describing the association between hypotonia and other HPO terms within age group cohorts in neonatal and pediatric patients.**Methods:**Using a de-identified historical dataset of genetic testing data, the hypotonia cohort was identified via keyword search for "hypotonia" in any HPO field. These fields were previously converted from clinical notes. Cytoscape software developed a knowledge graph to map co-existing HPO terms at the patient-level. R package wordcloud2 and Stats was used for enrichment analysis (comparing hypotonia to those without) and to develop word clouds stratified by age categories. **Results:**Among 222,525 cases referred for exome or genome sequencing, a cohort of 51,912 cases (23.3%) were identified with hypotonia. From newborn (1-30 d) to pediatric (1-4 y), the total number of phenotypes associated with hypotonia increase by 201.0%. Enrichment analysis revealed that the main phenotypes changed across age cohorts, with seizures most common in neonatal hypotonia (35%) and global developmental delay (GDD) most common in pediatric hypotonia (93%). Co-existence analysis revealed 979 HPO pairs from 232 HPO terms are present in 1% or more of the patients. Rare phenotypes only coexist with hypotonia but are not associated with any other phenotypes (N=51, 22.0%). A constellation of phenotypes (34.1%) are shared jointly between patients that exhibit both GDD and hypotonia. **Conclusions:**Our goal was to make available pairwise HPO correlations for 51,912 hypotonia cases. We observed changes in the pattern of reported HPO terms between age categories of patients with hypotonia, which provides evidence for complex comorbidity relationships. These findings support the implementation of exome sequencing into diagnostic workflows for infants with hypotonia.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session II

PB4508 Identification and characterization of genomic variationsfrom population scale whole genome sequencing inIndia

Authors:

B. Kahali[1], K. Subramanian[1], S. CHAKRABORTY[2], A. Rajesh[1], S. Vinod[1], S. Jani[1], N. Kalita[1], S. Adhya[1], TLSA study team, SANSCOG study team; [1]Ctr. for Brain Res., Indian Inst. of Sci., Bangalore, India, [2]Indian Inst. of Sci., Bangalore, India

India's population of 1.3 billion comprising >4,500 well defined ethnic groups could have numerous distinct genetic variations resulting from the country's rich genetic and social diversity, multiple waves of migration over the past thousands of years, and unique socio-cultural practices followed by extant population groups. However, representation of Indians in human genetic studies has been limited so far. Thus, population-specific reference genome datasets and genome-wide association studies in Indian population are warranted. Here, we present a comprehensive analysis of SNVs and InDels in 800 individuals from the TATA Longitudinal Study of Aging (TLSA) and Srinivaspura Aging Neurosenescence and Cognition (SANSCOG) population-based cohorts. We identify and characterize about 30 million SNVs and 4.5 million InDels from 40X short-read whole genome sequencing. The rare variants hugely outnumber the common variants, and about two-thirds of our identified SNVs and InDels are present in <1% individuals in our dataset. 3% of identified variations are present in coding regions. Out of ~12,000 frameshifts, ~4000 are rare. On average, each individual carry 20 variants intolerant to protein coding loss of function changes. We discover ~36,000 structural variants (SV), consisting of ~24,550 deletions, ~2900 duplications, ~8700 insertions, and 13 inversions. 1.4% of these SVs disrupt coding regions. 0.23% of total SVs are highly conserved and exhibit high intolerance to loss-of-function. The identified SNVs, InDels and SVs exhibit specificity to population groups. We elucidate the population structure inferred from coding and non-coding variants. We constructed a reference imputation panel using our discovery set variations merged with 1000Genomes SAS genotypes, and observe increased accuracy (~5%) compared to other standardized imputation panels. We have conducted rare-variant kernel-based association studies for commonly occurring age-associated phenotypes, highlighting novel suggestive associations. Our findings ameliorate the representation of South Asian populations in worldwide genetic studies. This dataset will be a resource for designing and interpreting large scale association and functional genomic studies in Indian population that could pave the way for devising better diagnosis and treatment strategies for a substantial proportion of world population.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session III

PB4509 Impact of Masking the Reference on Germline Variant Calling utilizing GPU Accelerated BWA, HaplotypeCaller and Deepvariant

Authors:

P. Vats[1], A. Sethia[1], H. Clifford[1], G. VACEK[1], M. Gatzen[2], M. Samadi[1]; [1]NVIDIA, Santa Clara, CA, [2]Broad Inst. of MIT and Harvard, Cambridge, MA

The human reference genome is a foundational resource enabling advances in genomic research. It provides a reference framework for mapping and analyzing DNA sequences from individuals and populations. The major applications of the reference genome are mapping, genome assembly, variant calling, functional annotation, and interpretation of genetic variation in individuals. Here, we explore the impact of human reference on variant calling. Variant calling involves identifying the differences between the reference genome and a given sample's DNA sequence. This is crucial for detecting genetic variation that may be associated with disease or other population specific traits. However, the choice of reference genome can impact the accuracy of variant calling as different versions of reference genome with false duplication of specific genes can lead to ambiguous mapping and can affect the ability to call a variant. The GIAB consortium (Genome in a Bottle) recommends masking the duplicated regions to remove false duplication which can improve variant calling accuracy. In this study, we show the impact of using the masked reference genome which results in higher accuracy than the un-masked reference genome. The overall accuracy is higher than any of the current state-of-the-art methods used for SNP variant calling. In this study, we focus on the guideline to use Illumina alt masked GRCh38 reference genome compared to GRCh38 reference genome. We performed germline variant calling utilizing the GIAB HG001 30x, HG002 30x, HG003 45x, and HG004 52x whole genome sequencing (WGS) dataset. Alignment was performed using GPU (Graphical Processing Unit) accelerated BWA-MEM and variant calling was attained by running GPU accelerated HaplotypeCaller and Deepvariant; these GPU accelerated applications are part of the Parabricks v4.1 release. For Single nucleotide variants (SNVs) on HG001-HG004 dataset, the HaplotypeCaller achieved on average recall 99.27%, precision 98.84%, and F1 score of 99.05% using unmasked reference, whereas for masked reference the average recall 99.32%, precision 98.84% and F1 score of 99.08% respectively. Similarly, on HG001-HG004 dataset Deepvariant achieved an average recall of 99.38%, precision 99.87%, and F1 score of 99.63% using unmasked reference, and for masked reference the average recall 99.40%, precision 99.88% and F1 score of 99.64%. The results from the two variant callers on HG001-HG004 dataset indicate that the variant calling was improved while using Illumina alt masked reference genome GRCh38 versus using the non-masked GRCh38 reference genome.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session I

PB4510 Implementing precision medicine on a nationwide scale

Authors:

P. Johansen, M. Bøgehus Rasmussen, B. Nybo Jensen; Danish Natl. Genome Ctr., Copenhagen S, Denmark

The Danish National Genome Center (DNGC) is a government agency established by law in 2019 as described in the Danish strategy for personalized medicine. The aim of the DNGC is, in cooperation with a wide range of stakeholders, to implement, facilitate and expand the use of whole genome sequencing (WGS) for Danish patients with rare diseases and cancer and to build a national safe high-performance computer infrastructure to store and exhibit large genomic datasets to clinicians and researchers. Patient groups, eligible for WGS, have been chosen through two recommendation rounds. This has resulted in the selection of 17 patient groups from several different medical specialties and falls within the broader categories of rare diseases (14 groups) and cancer (4 groups). For each group DNGC has facilitated and hosted a national specialist network, that have been tasked with making recommendations and describing the clinical indications and criteria for the use of WGS for their patient group. The patient groups contain 88 distinct indications of patients eligible for WGS as a first-tier diagnostic test. Currently the DNGC in collaboration with regional healthcare facilities sequences app. 1,100 samples a month in a population base of app. 5,8 million. These samples include, WGS, total RNA sequencing of cancer samples (solid and hematological) as well as a germline WGS for tumor-normal and heritable disease analysis. The total number of samples in the Danish Genome database is currently app. 16,000. DNGC have recently implemented a national genomic interpretation platform where interpreters from the healthcare sector can gain access to the relevant software and bioinformatic tools. Interpreters have access to software for both germline and somatic analysis as well as more than thousands of bioinformatic packages. Three technical working groups advise the DNGC on developments for the interpretation platform. The access to this platform raises the level of collaboration, knowledge sharing and makes genomic interpretation more uniform on a national scale for the advantage of patients receiving clinical WGS. DNGC have facilitated a national governance structure that includes technical and clinical leadership on how to implement WGS into a national public healthcare system. This governance relies on clinical experts, health professionals and decision makers in the Danish healthcare system to come together, participate, collaborate and take ownership to elevate precision medicine for the benefit of Danish patients. The DNGC has received funding for 5 years by a 147.000.000$ grant from the Novo Nordisk Foundation

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session II

PB4511 Improved phasing and imputation accuracy using a harmonized resource of high-coverage diverse human genomes.

Authors:

L. Nkambule[1,2], Z. Koenig[1,2], M. T. Yohannes[1,2], J. K. Goodrich[2,3], H. A. Kim[1], X. Zhao[3,4,5], M. W. Wilson[3], G. Tiao[3], S. Hao[3,4,5], N. Sahakian[6], K. R. Chao[3], gnomAD Project Consortium, M. E. Talkowski[1,3,4,5], M. J. Daly[1,2,3,7], H. Brand[3,4,5], K. J. Karczewski[1,2,3,4], E. G. Atkinson[1,2,8], A. R. Martin[1,2,3,4]; [1]Stanley Ctr. for Psychiatric Res., The Broad Inst. of MIT and Harvard, Cambridge, MA, [2]Analytic and Translational Genetics Unit, Massachusetts Gen. Hosp., Boston, MA, [3]Program in Med. and Population Genetics, The Broad Inst. of MIT and Harvard, Cambridge, MA, [4]Ctr. for Genomic Med., Massachusetts Gen. Hosp., Boston, MA, [5]Dept. of Neurology, Massachusetts Gen. Hosp. and Harvard Med. Sch., Boston, MA, [6]Broad Genomics, The Broad Inst. of MIT and Harvard, Cambridge, MA, [7]Inst. for Molecular Med. Finland, Helsinki, Finland, [8]Dept. of Molecular and Human Genetics, Baylor Coll. of Med., Houston, TX

The 1000 Genomes Project (1kGP) and Human Diversity Genome Project (HGDP) resources have existed for more than a decade but have only recently been sequenced to high coverage. These resources are often used independently for genomic analyses, despite various efforts showing how invaluable larger and more diverse resources are. We have jointly-called the 1kGP and HGDP datasets to generate a harmonized high quality public haplotype reference panel, HGDP+1kGP, which consists of 4,091 high-coverage whole genomes. This harmonized resource contains almost twice as many variants as the 1kGP reference panel alone. Compared to the 1kGP reference panel, we show that our HGDP+1kGP panel achieves higher phasing accuracy across different ancestries, with more pronounced improvement in African ancestry, and also higher imputation accuracy, particularly for rare variants. Specifically, our reference panel outperforms the 1kGP panel for genotype imputation in African individuals in both low coverage and array data and also in both SNPs and INDELs. These phased haplotypes are freely available through Google Cloud Storage. We believe this resource will be invaluable in supporting genomic research particularly in diverse ancestry cohorts.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session III

PB4512 Improved validation of conditions for variant classifications in ClinVar

Authors:

M. Landrum[1], S. Chitipiralla[1], G. Brown[1], C. Chen[1], J. Hart[1], D. Hoffman[1], C. Liu[1], Z. Maddipatla[1], R. Maiti[1], J. Mitchell[1], T. Rezaie[1], G. RILEY[1], G. Song[1], J. Yang[1], G. Zhou[1], V. Lyoshin[1], K. Kaur[1], A. Russette[1], B. Kattman[2]; [1]NCBI/NLM/NIH, Bethesda, MD, [2]NCBI/NLM/NIH, Fort Collins, CO

The ACMG/AMP guidelines for the interpretation of sequence variants (Richards et al, 2015) are intended "to determine whether a variant in a gene with a definitive role in a Mendelian disorder may be pathogenic for that disorder. Pathogenicity determination should be independent of interpreting the cause of disease in a given patient." These guidelines are often cited in submissions to ClinVar as the assertion criteria, or classification rules, used by the submitter to classify variants. These rules are consistent with the most typical case in ClinVar, *i.e.* a germline variant classified for a single Mendelian disease. The ClinVar team also sees other less common, less straightforward scenarios in submitted data. First, some variants are classified for a set of clinical features not a disease; without additional information, it can be difficult to distinguish whether the data describes a patient's specific phenotype or a novel disease that currently is only defined by the set of features. Second, disease definitions are not always discrete, so a variant may be classified for several diseases that are considered parts of a phenotypic spectrum. Third, not all laboratories that submit to ClinVar indicate the disease for the classification; this leads to many variants classified for a condition that is "not provided" or "not specified". These scenarios reflect the complexity of both biology and data management for clinical genetics; however, they can create confusion and uncertainty for ClinVar users. The ClinVar team has added more checks to pre-submission validation to help submitters provide the clearest and highest quality data as the condition for the classification. First, variants that are classified for a set of clinical features are only accepted when the features define a novel disease; a patient's phenotype is described elsewhere in the submitted record (SCV). Second, if the condition is described by multiple identifiers from a disease database, such as multiple OMIM IDs, the submitter must provide the "explanation for multiple conditions". Options for the explanation are "uncertain", meaning that the variant is classified for several diseases that may represent a spectrum of disease and it's uncertain if the variant is associated with one or more of the diseases; and "co-occurring", meaning that the variant causes two distinct diseases (expected to be rare). Third, submitters will now get an error if they provide more than one SCV classified for a disease and for no disease ("not provided" or "not specified"). These validations are expected to reduce the number of variants with conflicting classifications and to clarify the rarer, more complicated cases.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session I

PB4513 Improving the imputation quality of the cohort in the Department of Veterans Affair's Million Veteran Program (MVP) by leveraging the TOPMed reference panel.

Authors:

Y. Shi[1], B. Gorman[2], D. Dochtermann[3], X. Gong[1], T. Webster[4], A. Mittal[5], M. Sullivan[6], J. Schmidt[7], P. Karnam[8], P. Devineni[8], C. Cai[9], P. Tsao[10], S. Pyarajan[11], on behalf of MVP program; [1]VA Boston Hlth.care System, Jamaica Plain, MA, [2]Booz Allen Hamilton, Lexington, MA, [3]Boston VA Hlth.Care, Boston, MA, [4]Thermo Fisher Scientific, Olga, WA, [5]ThermoFIsher Scientific, San Jose, CA, [6]Thermo Fisher Scientific, New York, NY, [7]Thermo Fisher Scientific, Santa Clara, CA, [8]Boston VA Hlth.Care System, Boston, MA, [9]VA Palo Alto Hlth.Care System, Palo Alto, CA, [10]VAPAHCS/Stanford Univ., Palo Alto, CA, [11]VA Boston Hlth.Care System, Jamaica Plain, MA

The Million Veteran Program (MVP) was initiated by the U.S. Department of Veterans Affairs (VA) to implement personalized medicine in the care of U.S. veterans. The current tranche of high-quality genotype data contains 664,479 highly diverse patients assayed at 668,029 unique markers in a custom Applied Biosystems™ Axiom™ MVP1.0 microarray. Here we described the improvement of imputation quality of the MVP cohort with the TOPMed reference panel over the combined African Genome Resources (AGR) and 1000 Genomes (IKG), which was validated by using 10,375 whole genome sequences.We first performed pre-imputation variant QC by removing variants with high missingness (>20%), that failed an ancestry-aware Hardy-Weinberg Equilibrium test, or were monomorphic. We used SHAPEIT4 (v.4.2.0) to phase the variants, followed by Minimac4 (v 1.0.2) to impute phased genotypes. We evaluated imputation quality with empirical R^2 at different MAF bins across ancestries. We performed a height GWAS and compared height association p-values for genome-wide significant GIANT exome array associations to further validate imputation quality. To investigate whether the empirical R^2 in TOPMed imputation is inflated, we converted 10,375 whole genome sequences (EUR, AFR, HIS) into genotype dosages and calculated true R square (r^2) between the imputed genotype dosages and the WGS dosages. The TOPMed reference panel improved imputation quality (empirical r^2) for both common and rare markers, across all ancestries (EUR, AFR, AMR, ASN) in the MVP cohort. The number of well imputed markers (empirical $R^2 >= 0.7$) in the rare spectrums were greatly increased: the 0.01%-0.1% MAF bin increased from 45.8 to 67.4% and the 0.1%-1% MAF bin increased from 80.38 to 93.21%. Imputation in non-PAR X chromosome was also significantly improved. Common variant associations were similar between 1000 Genomes and TOPMed reference panels. The mean empirical r^2 of common markers (0.01, 0.5] across the ancestry spectrum was high: AFR was the highest (0.98 ± 0.007) followed by EUR (0.96 ± 0.013), and AMR (0.95 ± 0.009), and ASN (0.84 ± 0.01). The average difference between true r^2 and the empirical r^2 across the MAF bins was 0.55 ± 0.05, -0.03 ± 0.005, 0.029 ± 0.002 for EUR, AFR, and AMR, respectively. The mean empirical r^2 of rare markers ((0.001,0.01]) was above 0.86 and had the same population order as the common markers. We could not evaluate ultra-rare variant (MAF<0.1%) for AFR and AMR due to limited number of NGS samples. In conclusion, the TOPMed reference panel significantly improved the imputation, and imputation quality of all four ancestries and AFR and AMR benefit the most from the panel.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session II

PB4514 Increase the training dataset for disease-specific variant prioritization by combining SNVs in related diseases

Authors:

Q. Liang, D. Kostka; Univ. of Pittsburgh, Pittsburgh, PA

Better characterizing and prioritizing non-coding genetic variants have the potential to facilitate insights into molecular mechanisms underlying complex disease. Previously, we developed a disease-specific variant prioritization method and applied it to 111 disease terms in the NHGRI-EBI GWAS Catalog. Here, we build on this approach by systematically evaluating if information sharing between different disease terms leads to better variant prioritization performance. Specifically, we exhaustively screen all pairs of disease terms in the NHGRI-EBI GWAS Catalog, combine annotated SNVs, and assess the effect of variant prioritization. We combine SNVs between terms with a range of disease-term-specific sample weights and determine performance improvement of variant prioritization (or the lack thereof) using nested cross-validation. We find, for example, that SNV prioritization performance as measured by average precision increased 30% for “multiple myeloma” when combined with disease term “multiple sclerosis”, and 17% when combined with disease term “systemic lupus erythematosus”. More generally, we find that it is possible to quickly identify suitable term pairs and achieve an average increase of SNV prioritization performance of 5% across the GWAS catalog. These experiments show that combining SNVs across disease terms can be beneficial for disease-specific variant prioritization and suggest that data sharing is a promising avenue for improving the performance of disease-specific models and expanding their applicability to a wider range of diseases.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session III

PB4515 Initial modified ACMG criteria for *RPGR* variant curation from the ClinGen X-linked inherited retinal disease gene Variant Curation Expert Panel

Authors:

K. Worley[1], W. Hankey[2], M. Mero[1], R. Ayyagari[3], M. Cheetham[4], E. De Baere[5], V. Chong[6], N. Faya[7], K. Goetz[8], L. Haer-Wigman[9], H. M. Hussain[10], K. Kämpjärvi[11], R. Koenekoop[12], K. Lee[13], L. Meng[1], M. Nachury[14], P. Sieving[15], J. Stout[1], L. Sullivan[16], G. Wang[17], M. Wang[1], C. Zeitz[18], R. Lewis[1], R. Chen[1], ClinGen X-linked Inherited Retinal Disease VariantCuration Expert Panel; [1]Baylor Coll. of Med., Houston, TX, [2]UNC Sch. of Med., Univ. of North Carolina, Chapel Hill, NC, [3]Univ California San Diego, San Diego, CA, [4]UCL Inst. of Ophthalmology, London, United Kingdom, [5]Ghent Univ. and Ghent Univ. Hosp., Ghent, Belgium, [6]Janssen R & D, LLC, South San Francisco, CA, [7]Cincinnati Children's Hosp., Cincinnati, OH, [8]NEI Office of Data Sci. and Hlth.Informatics, Natl. Eye Inst., Bethesda, MD, [9]Radboud Univ. Med. Ctr., Nijmegen, Netherlands, [10]Baylor Coll. Med., Houston, TX, [11]Blueprint Genetics, OY, Keilaranta, Finland, [12]McGill Univ Hlth.Ctr., Montreal, QC, Canada, [13]Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, [14]Stanford Univ, Stanford, CA, [15]Ophthalmology & Vision Sci., Univ. of California Davis, Davis, CA, [16]Univ Texas HSC Houston, Houston, TX, [17]Invitae, San Francisco, CA, [18]Sorbonne Univ. Inst. de la Vision Dept. of Genetics, Paris, France

The X-linked Inherited Retinal Disease Variant Curation Expert Panel (XLIRD VCEP) was established in 2020 as part of the Clinical Genome Resource (ClinGen), to address the need for consistent variant interpretation in inherited retinal disease (IRD) genes with an X-linked inheritance pattern. These genes account for ~15% of all IRD cases reported in large cohorts and several are targets for genetic therapeutics. *RPGR* is one of the most commonly mutated genes in retinitis pigmentosa.

The XLIRD VCEP implements the ClinGen variant curation practices that have been recognized by the U.S. FDA to assess individual variants in seven X-linked genes (*CACNA1F*, *CHM*, *NPD*, *OFD1*, *RPGR*, *RP2*, and *RS1*). By combining experience from clinical and laboratory researchers with expertise in phenotypes and molecular mechanisms of IRD, a set of variant curation rules customized to each gene is specified to assess the pathogenicity of the variants. Variants in the seven genes will be scored systematically following these rules and categorized into five classes, including pathogenic, likely pathogenic, variant of uncertain significance (VUS), likely benign, and benign, according to ACMG/AMP guidelines.

We established the XLIRD VCEP, with membership and curation protocols available on the ClinGen website (https://www.clinicalgenome.org/affiliation/50086/). Starting with *RPGR* (OMIM 312610), in which variants account for 9% of rod-cone and cone-rod dystrophies, we generated gene-specific curation rules. Performance of various computational prediction tools are evaluated, and thresholds have been defined for allele frequencies in population databases consistent with benign or disease-causing variant effects. To facilitate the curation process, in addition to published variants, private variant data from genetic testing laboratories are also collected. Pilot curation exercises performed will be evaluated to refine the initial rules, which will then be applied to curate variants in the *RPGR* gene. This process will be repeated for the other XLIRD genes, and the curations reviewed iteratively to refine the processes further.

Systematic and consistent curation of variants from FDA-recognized gene-specific guidelines supported by sharing genomic data and expertise will evaluate the current classification of variants and decrease the numbers of VUS within these XLIRD genes. This will improve the specificity and accuracy of the molecular diagnoses of patients with variants in these genes and increase the value of genetic testing as a diagnostic tool and a guide for patient eligibility for genetic therapies for these inherited retinal disorders.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session I

PB4516 Insights into digenic mechanisms through DIVAs, an Explainable-AI phenotype-driven approach for digenic variant interpretation.

Authors:

F. De Paoli, G. Nicora, E. Rizzo, I. Limongelli, S. Zucca; enGenome srl, Pavia, Italy

Background
The challenge of identifying the genetic cause of Rare Diseases is even more demanding in the context of non-Mendelian inheritance. Due to the lack of guidelines for digenic variant interpretation (DVI), Artificial Intelligence (AI) tools could greatly support this process and Explainable-AI (XAI) allows further investigation of digenic mechanisms. We curated a dataset of pathogenic digenic combinations and we expanded DIVAs, a phenotype-driven AI-based approach for DVI, to dissect the digenic mechanisms into True Digenic or Composite (TD/CO) versus Dual Molecular Diagnosis (DM).
Methods
We gathered almost 800 pathogenic combinations from public databases, internal cases and manual curation and translated the phenotypic description reported in the corresponding study or provided by clinicians into a set of Human Phenotype Ontology (HPO) terms. DIVAs has been trained on *bona fide* data and integrated with an additional layer of XAI that exploits Shapley values to subclassify digenic combinations by mechanism. We tested the prioritization ability of DIVAs and other DVI tools (ORVAL, DiGePred and DIEP) on 9 confirmed digenic WES cases collected within international collaboration. DIVAs was then applied on a cohort of undiagnosed patients affected by disorders of the immune system provided by the Clinical Immunogenomics Research Consortium Australasia (CIRCA). Furthermore, a cohort of unsolved samples belonging to the Rare Genome Project (RGP) of the Broad Institute and 12 patients affected by Shwachman-Diamond syndrome (SDS) provided by the University of Pavia were analyzed.
Results
On a test set of about 400 combinations, our approach shows 80% sensitivity in pathogenic prediction. On the same data, the XAI approach used to dissect the digenic mechanism reports an accuracy of 91%, while ORVAL has almost 70% accuracy. On the WES data, DIVAs pinpoints the causative digenic combinations in the first 25 positions in 77% of the cases while the other DVI solutions show lower prioritization performances (11%). Furthermore, DIVAs identifies the genetic diagnosis of a patient affected by SDS caused by a *SBDS* mutation with a second mutation on *KMT2A* gene explaining additional phenotypes, including developmental delay, horseshoe kidney and bone abnormalities. The *SBDS-KMT2A* combination was correctly subclassified as a DM (PMID:35893049). CIRCA and RGP results review is still ongoing.
Conclusion
We improved DIVAs by including an XAI layer that predicts the digenic mechanism of pathogenic combinations. The integration of an oligogenic approach, exploiting biological networks, is ongoing.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session II

PB4517 *Integrating PacBio and Illimina producdes nearly complete* phased human genome references for six trios from Middle Easten ancestries

Authors:

M. Ghorbani[1], S. Moosa[1], Z. Siddig[1], R. Farhad[1], I. Alazwani[1], R. Mathew[1], M. Al Hashmi[1], A. El Khouly[1], S. Tomei[1], K. Fakhro[2], E. Eichler[3], Y. Mokrab[1]; [1]Sidra Med., Doha, Qatar, [2]Sidra Med. and Res. Ctr., Doha, Qatar, Qatar, [3]Univ of Washington, Seattle, WA

The genetic diversity of the Middle Eastern region is poorly captured in the various reference genomes published to date including the pangenome form HPRC. Genomic analysis of SNV and SV calls for both disease association studies and rare gene discovery has been shown to significantly improve with the use of population-specific genomic references and annotations. Based on six families (2 Qatari families, and one family from Sudan, Jordan , Syria and Afghanistan), we built de novo reference genomes using a combination of PacBio Hifi reads (30-50x), and Illumina short-read sequencing (30X). For 12 (father and mother) individuals from these trios, we generated high quality haplotype-resolved phased assemblies using hifi-asm and for 6 childs we used trio-hifi-asm to generate phased assemblies. Haplotypes were resolved accurately using independent assessment of assemblies by parental inherited short reads k-mers. Also, high Quality Value (QV) values in the range of 44-59 were achieved for the haplotypes. In terms of contiguity, we obtained N50 values between 50 and 100 Mb and a maximum contig length of more than 130 Mb. We managed to achieve telomere to telomere assembly of chromosomes 6,8,11,12,18 for three individuals combined (one contig per chromosome). Our best assembly achieved T2T assembly of chromosome 8 and 12 for both haplotypes. These high quality genome references will provide valuable resource for population specific gene discovery, variant calling, and disease association studies for underrepresented region of Middle East.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session III

PB4518 Integrative Analysis of miRNA and mRNA (MIMR): a web tool for identification of hidden pathological mechanisms

Authors:

d. kim, Y. Ko; Hankuk Univ. of Foreign Studies, Seoul, Korea, Republic of

MicroRNAs (miRNAs) are small non-coding RNAs that intricately regulate target gene products, resulting in the inhibition of gene expression. Although these miRNAs play crucial roles in essential biological processes, encompassing immunity, metabolism, and cell death, their specific impacts on diseases remain unknown. Recent studies have been focused on the integration of miRNA and mRNA expression to reveal the underlying biological pathways and mechanisms responsible for disease manifestation. Traditional approaches are involved in miRNA target relationships incorporated from TargetScan, PicTar, or PITA, and mRNA and miRNA expression data are utilized to identify disease-associated pathways, However, these methods consider only direct interactions between proteins or genes and may miss significant indirect regulatory effects, leading to the failure of identification of hidden pathological mechanisms.To address these limitations, we have proposed the novel approach for integrative analysis of miRNA and mRNA expression data and developed MIMR (Integrative Analysis of miRNA and mRNA), a web-based application that leverages Random Walk with Restart (RWR) algorithm. MIMR incorporates both direct and indirect interactions by harnessing protein-protein interaction (PPI) networks and experimentally validated mRNA-miRNA target interactions. MIMR provides extensive results, including pathological pathways associated with a specific disease, interactive network diagrams representing the mRNAs and miRNAs. Our method applied to mRNA-miRNA expression data for nephrotic syndrome patients and successfully identified the biological pathways related to the manifestation of symptoms in patients with nephrotic syndrome. In summary, MIMR offers a deeper insights into the hidden mechanisms of diseases and identify potential therapeutic strategies through integrated analysis of miRNAs and mRNAs.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session I

PB4519 † Introduction to the Health and Medical Research Resources Information center in Korea National Institute of Health:focusing on Clinical & Omics Data Archive (CODA)

Authors:

Y. Lee, J-E. Kim, H. Chai, H-Y. Park; Korea NIH, Cheongju-si, Korea, Republic of

The Clinical & Omics Data Archive (CODA) is established in Aug, 2016 by Korea National Institute of Health (KNIH). CODA is a central data repository for integrating and sharing Korean multi-omics data produced from the research projects sponsored by the Ministry of Health and Welfare (MoHW). It aims to construct a trust based archive for bio-medical infrastructure. It collects various biomedical data (epidemiological information, health records etc al) and multi-omics data (microarray, whole-exome, whole-genome, transcriptome, metabolome). To date, an unprecedented amount of data (about 226,955 samples, 2,972TB) from national research projects are collected in CODA. We have collected genetic data (Whole Genome sequencing data, >30X) from more than 20,000 Koreans. With these 20,000 samples, we are working on cataloging genetic variations specific to the Korean population and the results will include information about genomic variations, such as single nucleotide variants, insertions, deletions, and structural variations. It will be a valuable resource for researchers and medical professionals studying genetics and genomics by providing a comprehensive and accurate dataset of genetic variants that can be utilized for a variety of purposes, including disease research, precision medicine, and population genetics studies. CODA had shared qualified data to researchers for data-driven studies. It also provides analysis infrastructure with various pipelines for facilitating analysis. We expected to contribute to the advancement of personalized medicine and enhance the understanding of the genetic basis of diseases, ultimately leading to improved diagnosis, treatment, and prevention strategies specific to the Korean population.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session II

PB4520 Investigating the complex genomic structure of the low-affinity *FCGR* locus under different copy number states in healthy donors using long-range sequencing.

Authors:

S. Frampton[1], J. Forster[1], M. Armand[1], H. Parker[1], K. Hussain[1], C. Hargeaves[1], R. Foxall[1], R. Smith[1], D. Bryant[1], S. Beers[1], M. Cragg[1], J. Gibson[2], J. Strefford[1]; [1]Univ. of Southampton, Southampton, United Kingdom, [2]Univ of Southampton, Southampton, United Kingdom

Therapeutic monoclonal antibodies rely on FcγRs for efficacy. The low-affinity FcγR genes (*FCGR2A, FCGR2B, FCGR2C, FCGR3A, FCGR3B*) are in a highly polymorphic 200k locus, challenging to analyze with short-read methods. We used long-range technologies (Oxford nanopore (ONT) and Bionano (BN) for accurate study of this locus in poorly defined copy number regions (CNRs).

Using a healthy human cohort (n=22) with different *FCGR* SNPs and CNV states, ONT adaptive sampling and cas9-mediated enrichment (chr1:161490622-161690621) enabled SNP and SV calling, phasing, de novo assembly, and methylation calling. *FCGR* maps from ONT data were compared to BN optical maps from matched samples.

ONT sequencing (mean coverage 200x, 35Kb N50), identified a total of 16,326 high-quality SNPs and indels across 2,296 unique genomic locations, averaging 742 variants per sample. Most variants were found in intergenic regions (60.5%), followed by intronic regions (26.7%). Variants were also observed in 3'-UTRs (3.3%), downstream of genes (2.9%), upstream of genes (2.8%), exons (1.8%), splicing sequences (0.8%), and 5'-UTRs (0.8%).

The intergenic region between *FCGR3B* and *FCGR2B* exhibited the highest polymorphism frequency, accounting for 22.5% of all variants (n=3673). Among the genes, *FCGR2C* had the highest number of variants (34.5%, 2036 total), closely followed by *FCGR3B* (31.9%, 1881 variants). Interestingly, *FCGR3B* had the highest polymorphic index (variants/gene length) and proportion of homozygous variants (51.5%), while the remaining genes had 20-35% homozygous variants.

Most of the variants (88.6%) were associated with an rs ID and we identified 114 variants (5%) that were shared across all 22 samples. *FCGR2C* had the highest count of variants (148, 7.3%) without rs IDs. Successful phasing of 92.8% of the heterozygous variants enabled genotyping of clinically-relevant SNPs and identification of extended haplotypes.

Eight polymorphic short tandem repeat (STR) regions were detected that have the potential to be of functional relevance to FcγR expression. Leveraging Bionano and ONT technologies, we refined the locations of breakpoints for CNR1 (5.4kb), CNR2 (14.5kb), CNR3 (7.9kb), and CNR4 (4.7kb). Notably, we reported a novel case of CNR4 duplication involving the entirety of *HSPA7* and *FCGR3B* with its breakpoints within the proximal region of *FCGR2C* and *FCGR2B.*

In conclusion, long-range technologies offer a powerful solution to characterize *FCGR* genomic variants. This enhances our understanding of the region's regulation, potentially leading to targeted therapeutic approaches for cancer patients and improved patient stratification.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session III

PB4521 † JOB : Japan Omics Browser provides integrative visualization of multi-omics data.

Authors:

Y. Takahashi[1], Q. Wang[2], Y. Okada[3], Japan COVID-19 Task Force; [1]Osaka Univ., Suita, Japan, [2]The Univ. of Tokyo, Tokyo, Japan, [3]The Univ.Tokyo / Osaka Univ. / RIKEN, Tokyo, Japan

In recent years, a wealth of human genome, transcriptome, proteome and other omics data has become available. Integrative analysis of different omics layers is effective in elucidating disease mechanisms. Here, we present the Japan Omics Browser (JOB), which enables visualization of regulatory effects of millions of variants to nearby genes. JOB offers visualization of per-variant regulatory effects in the human blood at mRNA and protein level distinctively, quantified from statistical fine-mapping of mRNA-expression quantitative loci (eQTL) and protein QTLs (pQTLs) in 1,405 samples from the Japan COVID-19 Task Force (JCTF ; extended from Wang QS. et al. Nat Commun 2022). In addition, JOB shows per-tissue regulatory effect prediction score (EMS), trained via multi-task learning utilizing hundreds of features (Wang QS. et al. Nat Commun 2021). More than 10% of the putative regulatory variants available in the browser are validated through massively parallel reporter assay (MPRA), where the full data over 20,000 variants are planned to be added to the browser by the time of presentation. This will be accompanied with further additions of omics data, including micro RNA expression QTLs and fine-mapped cell type-specific eQTLs from single cell RNA-seq experiments, all aiding functional interpretation of trait associated variants at single variant resolution. We summarize a set of examples demonstrating the utility of JOB. First is the example of rs74953707, where all the EMS, MPRA, and eQTL signal colocalize to clearly highlight the regulatory effect on gene *FAM189B* (posterior inclusion probability : PIP=1.0). Notably, this variant is rare (MAF~=0.003) in European and the association has not been reported in other major database, highlighting the value of JOB. Second is the example of rs35083095, which shows high pQTL signal and affects genes *SETMAR* (PIP=0.78) and *SUMF1* (PIP=0.73) through changes in protein expression levels, providing further characterization of its previously known eQTL effect. Finally, as an example of interpreting complex trait-associated locus with JOB, we highlight the regulatory effect of rs10411704 on *CD22* expression (50< fold enrichment of EMS and pQTL PIP=0.37) co-localizing with hematopoietic traits (PIP=0.28) in the UK Biobank. JOB, an intuitive and interactive web tool friendly for both experts and non-experts, is freely accessible at (https://optical-hexagon-384504.an.r.appspot.com/), on desktop, smartphones or any other devices.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session I

PB4522 Landscape of Colorado Center for Personalized Medicine Research Freeze 3

Authors:

N. Rafaels[1], S. Chavan[1], M. Fisher[1], J. Shortt[1], N. Pozdeyev[1], H. Smith[2], M. Lin[1], E. Todd[2], K. Crooks[1], C. Gignoux[1]; [1]Univ. of Colorado Anschutz Med. Campus, Aurora, CO, [2]Univ. of Colorado Anschutz Med. Campus, AURORA, CO

IntroductionThe Colorado Center for Personalized Medicine (CCPM) has recently completed a research freeze incorporating 73,346 individuals genotyped or imputed at 46,264,610 sites from 49,372 samples from whole exome sequencing (WES) and 24,079 samples genotyped on the Multi-Ethnic Genotyping Array (MEGA). Filtering of the WES and MEGA data to the American College of Medical Genetics (ACMG) Recommendations for Secondary Findings revealed 1793 and 522 patients, respectively, identified for confirmatory genotyping. GWAS summary statistics are available for 1563 phecodes derived from electronic health records (EHR), as well as imputed blood type and HLA type. **Methods**Briefly, 49,496 samples were sequenced using hg38 Twist comprehensive exome kit with additional non-exonic regions sequenced at low coverage (GxS) to yield genomic backbone coverage sufficient for whole-genome imputation. After removing 124 samples for quality control, we performed variant QC on 10,452,242 and 1,414,694 variants from WES and GxS, respectively, keeping 9,741,141 and 1,139,510 variants, respectively. We then merged 24,079 non-overlapping samples genotyped on MEGA (QC methods: https://www.medrxiv.org/content/10.1101/2022.06.09.22276222v1) keeping only sites genotyped in both datasets with concordant alleles, after which principal component analysis was performed. We then performed imputation against TOPMed reference panel, restricting to variants with MAF>0.00005 or inclusion on the Illumina Global Diversity Array to achieve sufficient overlap, leaving 1,853,339 autosomal variants in 2 batches. After filtering to variants with R2>0.55, we then merged the two RGC imputed batches with previously imputed data from MEGA using IMMerge keeping variants in common. We then re-inserted genotype calls from RGC that passed QC for imputed variants in common, reset R2=1 and recalculated allele frequency. **Conclusions**Our current research freeze consists of one of the largest Biobank GWAS datasets spanning the entire United States, and genotyping samples from additional whole exomes and GWAS array is underway. In the future, we hope to enrich our database with PRS, MetaXcan results, set based tests, as well as mine whole exome data for CHIP, telomere length, and more. We collaborate openly on numerous initiatives including the Global Biobank Meta-Analysis Initiative, the PAGE Study, and the PRIMED Consortium. We provide a list of best practices for merging array and NGS-based datasets as calibrated for biobanks and other institutional resources.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session II

PB4523 Leveraging distributed cloud infrastructure to analyze genetic data at scale.

Authors:

M. Khanna, B. Alexander; Pfizer, Cambridge, MA

The quantity and quality of genetic data has been increasing at an exponential rate over the last few decades. These increases present new challenges for both the storage and analysis of genetic datasets. Systems such as High-Performance Computing environments allow for increased computational power for genetic analyses but are ultimately bottlenecked by their shared and limited resources. Further, computational algorithms are increasing in complexity, but parallelization can be difficult and time-consuming to implement. Variability in algorithm implementation, underlying libraries, and hardware stacks can lead to artifacts and unreproducible results. Therefore, a new paradigm is required to meet the demands of these challenges. The Genetic Informatics Platform provides a framework for scientists to ingest genetic data and to build reproducible workflows utilizing the power of cloud infrastructure. Our cloud-based approach addresses these challenges by providing the following solutions: dynamic scalability and distributed computing; improved performance through the utilization of multiple database types; and workflow reproducibility and transparency through containerized algorithms. Using this framework, we have ingested 35,813 public genome wide association studies, standardized to a common human genome build and integrated into a distributed database cluster. These datasets reference 32,845 distinct traits, across a total of 72 populations. Further, we provide a user interface to query and visualize this database through interactive plots. Additionally, containerized algorithms can retrieve data from cloud services and local locations. This cloud framework enables researchers to create reproducible workflows that scale both vertically and horizontally, meeting the demands of the future of genetic datasets and analysis.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session III

PB4524 † Long-read sequencing of 1000 Genomes Project samples to catalog normal patterns of human genome structural variation.

Authors:

J. Gustafson[1], K. Hoekzema[2], M. Galey[3], J. Goffena[3], N. Damaraju[4], S. B. Gibson[2], W. De Coster[5,6], C. Reeves[7], A. Helland[7], A. Raval[7], A. Snead[8], C. Gonzaga-Jauregui[9], The 1000 Genomes ONT Sequencing Consortium, M. Zody[7], F. Sedlazeck[10,11,12], M. Loose[13], M. Jain[14], E. Eichler[2,15,16], D. E. Miller[16,3,17]; [1]Univ. of Washington, Seattle, WA, [2]Dept. of Genome Sci., Univ. of Washington, Seattle, WA, [3]Dept. of Pediatrics, Univ. of Washington, Seattle, WA, [4]Inst. for Publ. Hlth.Genetics, Univ. of Washington, Seattle, WA, [5]Applied and Translational Neurogenomics Group, VIB Ctr. for Molecular Neurology, VIB, Antwerp, Belgium, [6]Applied and Translational Neurogenomics Group, Dept. of BioMed. Sci., Antwerp, Belgium, [7]New York Genome Ctr., New York, NY, [8]Dept. of Biology, New York Univ., New York, NY, [9]Intl. Lab. for Human Genome Res., Univ. Natl. Autónoma de México, Queretaro, Mexico, [10]Human Genome Sequencing Ctr., Baylor Coll. of Med., Houston, TX, [11]Dept. of Molecular and Human Genetics, Baylor Coll. of Med., Houston, TX, [12]Dept. of Computer Sci., Rice Univ., Houston, TX, [13]DeepSeq, Sch. of Life Sci., Univ. of Nottingham,, Nottingham, United Kingdom, [14]Dept. of Bioengineering, Dept. of Physics, Northeastern Univ., Boston, MA, [15]Howard Hughes Med. Inst., Chevy Chase, MD, [16]Brotman Baty Inst. for Precision Med., Univ. of Washington, Seattle, WA, [17]Dept. of Lab. Med. and Pathology, Univ. of Washington, Seattle, WA

Over 50% of individuals with a suspected Mendelian condition remain undiagnosed after a comprehensive clinical evaluation, limiting their ability to benefit from precision or N-of-1 therapies. This has led to growing interest in using new technologies such as long-read sequencing (LRS) in challenging unsolved cases to identify precise molecular diagnoses. However, use of these new technologies is hindered by the limited availability of comprehensive datasets from healthy controls, which are crucial for filtering, categorizing, and prioritizing identified variants. Per genome, LRS-based approaches identify approximately 25,000 structural variants (SVs) larger than 50bp, including insertions, deletions, duplications, inversions, and translocations. This is more than three times the number of SVs detected by short-read genome sequencing (srGS). Filtering and prioritizing variants identified by LRS becomes challenging without an appropriate annotation dataset,.
The 1000 Genomes Project was an international endeavor to identify patterns of human variation among diverse populations, using samples consented for broad sharing of both materials and data. A range of technologies, including microarray, exome sequencing (ES), and srGS have been employed to characterize these samples, and the data generated have been used for variant filtering and prioritization both in the research and clinical settings. Building upon that project, the 1000 Genomes ONT Sequencing Consortium is an international effort to sequence at least 500 samples (~40% African, 40% Asian, 10% European, and 10% American ancestry) from the 1000 Genomes Project using the Oxford Nanopore Technologies (ONT) platform. Our goal is to better characterize normal genome-wide SV patterns as well as to identify variants in regions of the genome difficult to evaluate using srGS (e.g. hard-to-map regions).
Here, we present preliminary data from the first 250 genomes sequenced to 30x coverage with an average read N50 of 50 kbp. Our analysis includes both alignment and de novo assembly-based approaches. Shasta-based de novo assembly resulted in contig N50 values of more than 30 Mbp and enabled phased variant calling with DeepVariant and phased SV calling using both Sniffles2 and cuteSV. We used joint SV callsets from this dataset to filter and prioritize SVs in unsolved clinical cases. The data generated as part of this project are publicly available and analysis efforts are open to all. We anticipate that this dataset will be of broad utility to the human genetics community, and lead to discovery of disease-causing variants among individuals being sequenced with the ONT sequencing platform.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session I

PB4525 Long-read sequencing of hundreds of human brains provides insight into the impact of structural variation and methylation

Authors:

K. Billingsley[1], K. Daida[1], M. Meredith[2], P. Alvarez Jerez[1], R. Genner[1], L. Malik[3], B. Baker[1], M. Mastoras[2], m. abdelhalim[1], M. Jean[2], R. Lorig-Roach[2], M. Asri[2], S. Behera[4], K. Shafin[5], T. Pesout[6], T. Pesout[2], P. Carnevali[7], J. Yang[8], A. Rhie[1], K. Miga[9], M. Nalls[10], W. Timp[11], A. Phillippy[12], M. Chaisson[8], X. Reed[1], M. Jain[13], F. Sedlazeck[14], M. Kolmogorov[1], B. Paten[9], C. Blauwendraat[1]; [1]NIH, Bethesda, MD, [2]UCSC, santa cruz, CA, [3]NIH, bethesda, MD, [4]Baylor Coll. of Med., Houston, TX, [5]Google Hlth., Mountain View, CA, [67]Chan Zuckerberg Initiative,, Redwood City,, CA, [8]Univ. of Southern California, Los Angeles, CA, [9]UCSC, Santa Cruz, CA, [10]datatechnica, glen echo, ME, [11]Johns Hopkins Univ., Baltimore, MD, [12]NIH/NHGRI, Bethesda, MD, [13]Univ. of California Santa Cruz, Santa Cruz, CA, [14]Baylor Coll. Med., Houston, TX

Structural variants drive gene expression in the human brain and are causative of many neurological conditions. However, most existing genetic studies have been based on short-read sequencing methods which poorly capture these regions of the genome. Long-read sequencing significantly increases our power to detect disease-associated and functionally relevant structural variants but has not been a feasible option for large-scale genomic projects because it was low throughput and too costly. Here, we leverage a new scalable wet-lab protocol and computational pipeline for Oxford Nanopore Technologies (ONT) and apply it to neurologically normal control samples from the North American Brain Expression Consortium (NABEC) cohort. Through this work, we present a publicly available long-read resource from the frontal cortex of 222 human brain samples (average N50 ~30kb and 44X coverage). We discover 86,440 structural variants, consisting mainly of insertions (n=46,434) and deletions (n=38,602). Utilizing matched expression datasets for these samples including CAGE-seq and bulk and single-cell RNA-seq we apply quantitative trait locus (QTL) analyses and identify structural variants that impact gene expression in post-mortem frontal cortex brain tissue. Further, we determine haplotype-specific methylation rates of millions of CpGs and with this data identify cis-acting structural variants. In summary, these results highlight that long-read sequencing at population scale can identify disease-relevant regulatory loci that were inaccessible using previous technologies. We believe this new resource will provide a critical step toward understanding the biological effects of genetic variation in the human brain. Expanding on this, as part of the NIH's Center of Alzheimer's disease and Related Dementias (CARD) long-read sequencing initiative, we are currently applying this framework to sequence a total of 4000 human case and control brain samples to catalog structural variants in Alzheimer's and related Dementias.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session II

PB4526 Mass General Brigham Biobank - Insights from a New England Healthcare Biobank

Authors:

S. Koyama[1,2], Y. Wang[2,1], K. Paruchuri[2], M. Uddin[1], S. J. Cho[1], S. Urbut[1], S. Haidermota[2], W. Hornsby[2], R. Green[3], B. M. Neale[2], J. W. Smoller[2], M. Lebo[4], E. W. Karlson[2], P. T. Ellinor[2], A. R. Martin[2], P. Natarajan[2]; [1]The Broad Inst. of MIT and Harvard, Cambridge, MA, [2]Massachusetts Gen. Hosp., Boston, MA, [3]Harvard Med. Sch., Boston, MA, [4]Mass Gen. Brigham Lab for Molecular Med., Cambridge, MA

Mass General Brigham (MGB) is an integrated healthcare system in the Greater Boston area, Massachusetts, United States, serving 1.5 million patients annually. Within MGB, we analyzed genetic and phenotypic data from over 65,000 patients in a volunteer patient cohort with the goal of discovering the interplay between disease and genomics with a particular emphasis on understanding the impact of ancestral diversity in real-world medical practice through comprehensive phenotyping and genotyping.

We used a network-based clustering genetic algorithm within the MGB Biobank to reveal a detailed population structure, including substructures within European ancestry. These population structures mirror the sequential immigration to the Greater Boston area throughout American history. The genetic ancestral clusters we identified exhibited distinct geographical distributions, allele frequencies of pathogenic variants, and polygenic load. These findings suggest that granular ancestral assessment has the potential to enhance personalized care for patients.

The use of the MGB Biobank offered a notable advantage over general population biobanks, as it demonstrated a relative enrichment for a breadth of uncommon conditions. Using genome-wide phenome-wide association analysis, we associated genotypes and 1,416 PheCodes identifying 114 Bonferroni-significant associations ($P < 1.6 \times 10^{-11}$). We identified several highly penetrant variants such as rs35004220 and Hereditary hemolytic anemias, rs72660908 and Rhesus isoimmunization in pregnancy, that were consistent with prior studies while expanding our understanding of the associations between specific genetic variants and their corresponding phenotypic manifestations.

Our research highlights the power of large-scale, unbiased analyses within a hospital-based biobank to understand the complex interplay between genotypes, geographical location, and phenotypes, paving the way for tailored interventions.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session III

PB4527 Matched Annotation from NCBI and EMBL-EBI (MANE): Towards completion of MANE Select set and beyond.

Authors:

S. Pujar[1], J. Loveland[2], A. Astashyn[1], R. Bennett[2], A. Berry[2], E. Cox[1], C. Davidson[2], O. Ermolaeva[1], C. Farrell[1], R. Fatima[2], T. Goldfarb[1], J. Gonzalez[3], D. Haddad[1], M. Hardy[2], T. Hunt[4], J. Jackson, D[1], V. Joardar[1], M. Kay[2], V. Kodali, K[1], K. McGarvey[1], J. Mudge[2], M. Murphy[1], S. Rangwala[1], F. Thibaud-Nissen[1], A. Vatsan[1], C. Wallin[1], D. Webb[1], A. Frankish[2], F. Cunningham[2], T. Murphy[1]; [1]NCBI/NLM/NIH, Bethesda, MD, [2]EMBL-EBI, Hinxton, United Kingdom, [3]EMBL-EBI, Bethesda, United Kingdom, [4]EMBL/EBI, Hinxton, United Kingdom

Accurate and consistent genome annotation is a prerequisite to leveraging genomic sequence information for various applications. In clinical genomics where annotation products are used to report clinical variants, the lack of a reporting standard may cause confusion and inconsistency in reporting and clinical interpretation of variants. Matched annotation from NCBI and EMBL-EBI (MANE), a collaboration between National Center for Biotechnology Information (NCBI) and European Molecular Biology Laboratories-European Bioinformatics Institute (EMBL-EBI), provides a representative transcript (MANE Select) and the associated protein for every protein-coding gene to serve as a universal standard for variant reporting and for display in commonly used genomic resources. For a small number of genes where the MANE Select alone is not sufficient to report all known clinical variants, a second transcript called MANE Plus Clinical is provided after consultations with clinical experts. Each MANE Select and Plus Clinical transcript represents an exact match in the exons of a RefSeq transcript and the corresponding Ensembl transcript such that two identifiers can be used synonymously. MANE transcripts are intended to be stable and will not be updated to cause a change in the sequence accession or version, except in rare instances when egregious errors are identified. MANE transcripts are chosen based on biological criteria such as transcript expression and coding region conservation and perfectly match GRCh38. MANE v1.1 released in April 2023 covers almost 99% of protein-coding genes including all genes in the ACMG SF v3.1 set. Eleven MANE Select transcripts are annotated on a patch representing an alternate locus because the gene could not be accurately represented on the chromosome. Five MANE Plus Clinical transcripts were also added in this release. This year, we are working to add the remaining protein-coding genes and achieve near-100% coverage. MANE data is available on FTP and multiple RefSeq and Ensembl resources. Detailed information about the MANE dataset is available at https://www.ncbi.nlm.nih.gov/refseq/MANE/. We welcome your feedback or requests on specific MANE transcripts at MANE-help@ncbi.nlm.nih.gov. As a next step, we plan to include non-coding genes in MANE. We have begun work towards converging on annotation of a set of non-coding genes associated with disease phenotypes and hope to include them in a MANE release this year. The work is funded in part by the National Center for Biotechnology Information of the National Library of Medicine. Additional support is from Wellcome Trust-WT200990/Z/16/Z, EMBL-Core-Funds and NIH-U24HG007234.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session I

PB4528 MedGenome's genomics solutions for large-scale genetics studies

Authors:

K. Suryamohan[1], R. Gupta[2], M. Corbo[1], R. Menon[3], M. Muraleedharan[1], E. Ramesh[1], H. Gowda[1], H. G.H[1], V. Ramprasad[4], S. Seshagiri[1]; [1]MedGenome Inc., Foster City, CA, [2]MedGenome Labs Ltd., Bangalore, India, [3]MedGenome Labs Pvt ltd, Bangalore, India, [4]MedGenome Inc., Bengaluru, India

Genomics has revolutionized biomedical research, transforming our understanding of human health and disease. MedGenome, a leading provider of sequencing and bioinformatics solutions, offers a comprehensive range of services, including whole genome and whole exome sequencing, transcriptome sequencing, liquid biopsy, immune repertoire analysis, and single-cell sequencing. Our proprietary solutions effectively support pre-clinical research, biomarker discovery, and early drug development pipelines for academia, pharmaceutical and biotechnology companies. With state-of-the-art laboratories in India and the US, our team of highly skilled and experienced scientists can process a wide array of sample types using established best practices, ensuring the generation of high-quality sequencing data. MedGenome has developed an array of cutting-edge bioinformatics tools and optimized wet lab workflows to meet the ever-evolving needs of research and diagnostics. We have created a user-friendly, cloud-based analytics platform that facilitates scalable genomic data analysis and can generate publication-ready figures and reports. This interface was specifically designed for ease-of-use and does not require any prior programming or bioinformatics expertise. MedGenome has spearheaded the development of several proprietary technologies including VarMiner, an AI-enabled variant interpretation software, and HiTmAb, an innovative high throughput antigen-specific monoclonal antibody discovery platform that harnesses the power of single-cell B-cell receptor (BCR) sequencing for identification of antigen-specific B-cell antibodies.MedGenome has also positioned itself as a frontrunner in human genetics research. Through collaborations with over 500 hospitals in India, we routinely generate large-scale genomic data to identify novel human genetic variants to enable drug discovery and diagnostics. As a founding member of the GenomeAsia 100K project, we have curated an extensive catalog of population-level genomic sequencing data from across South Asia, thus creating a vital resource for identifying novel genetic risk factors and rare variants that are underrepresented in public databases. As part of our ongoing efforts, we have successfully launched several clinical tests, including Kardiogen, a Polygenic Risk Score test used to determine an individual's risk of developing coronary artery disease. Our expertise in diagnostics, sequencing and bioinformatics positions us at the forefront of biomedical research, driving groundbreaking advancements and empowers scientists to address complex genomic questions.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session III

PB4530 MitoNet1.0: A genetic interaction network of the human mitochondrion.

Authors:

N. Mikolajewicz[1,2], S. Masud[1,2], D. Thomson-Taylor[1,2], K. Chan[1], K. Brown[1], O. Sizova[1], A. Habsid[1], R. Climie[2], M. Billman[3], C. Myers[4], B. Andrews[2], M. Costanzo[2], C. Boone[2], J. Moffat[1]; [1]The Hosp. for Sick Children, Toronto, ON, Canada, [2]Univ. of Toronto, Toronto, ON, Canada, [3]The Univ. of Bonn, Bonn, Germany, [4]Univ. of Minnesota, Minneapolis, MN

Background. The human mitochondrion is a complex organelle involved in numerous functions beyond its indispensable role in aerobic respiration. This complexity has resulted in challenges in defining what exactly a mitochondrial gene is. Various strategies (e.g., protein-protein interactions, protein localization, bioinformatics) have resulted in complementary, yet discordant mitochondrial catalogs, including gene ontology (GO), DepMap, GeneDX, and MitoCarta, among others. ***Methods***. We performed >200 genome-wide CRIPSR/Cas9 screens using a diverse library of isogenic mutant HAP1 cells to identify genetic interactions. Library-side co-essentiality (Pearson correlation and topological overlap matrices) and latent features identified through machine learning (ensemble classifier) were used to annotate mitochondrial genes. Mitochondrial gene recovery was benchmarked using existing mitochondrial catalogs and summarized using area under the precision-recall curve (AUPRC). Novel mitochondrial genes predicted by MitoNet1.0 were functionally validated in orthogonal CRISPR/Cas9 screens and *in vitro* assays. ***Results***. We present MitoNet1.0, a genetic interaction-based network of human mitochondrial genes. We identified >1500 known and novel mitochondrial genes with an AUPRC of 0.718 (DepMap), MitoCarta (0.663), 0.512 (GO), and 0.309 (GeneDx). MitoNet1.0 recovered core mitochondrial processes, like OXPHOS/ETC pathway, as well as associated mitochondrial processes, like intrinsic apoptosis. We highlight several novel mitochondrial genes as case examples demonstrating the utility of MitoNet1.0 for improved gene annotation efforts. ***Conclusions***. MitoNet1.0 is a relational mitochondrial network defined using genetic interaction co-similarities profiles. It represents an invaluable resource that can be used to query the function and relationship of candidate mitochondrial genes.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session I

PB4531 Mucopolysaccharidosis Type VII (MPS VII): *GUSB* gene variant database

Authors:

S. Daugherty[1], T. Hamazaki[2], M. Hegde[3], R. Giugliani[4], T. MacKenzie[5], M. Kiel[6], O. K. Japalaghi[1], B. Lianoglou[5], D. Marsden[1], D. McKnight[7], J. L. Merritt, II[1], N. Tatiele[1], L. Pollard[8], V. Rangel Miller[1], T. Sparks[9], N. Miller[1]; [1]Ultragenyx Pharmaceutical Inc., Novato, CA, [2]Osaka Metropolitan Univ., Osaka, Japan, [3]PerkinElmer, Waltham, MA, [4]HCPA/UFRGS, Porto Alegre, Brazil, [5]UCSF Sch. of Med., San Francisco, CA, [6]GENOMENON, Ann Arbor, MI, [7]Invitae Corp., San Francisco, CA, [8]Greenwood Genetic Ctr., Greenwood, SC, [9]Univ. of California, San Francisco, San Francisco, CA

Background: Mucopolysaccharidosis type VII (MPS VII) is an ultrarare, autosomal recessive, lysosomal storage disease caused by beta-glucuronidase (GUSB) enzyme deficiency. Patients present variably with fetal effusions, skeletal dysplasia, dysmorphology, and cardio-pulmonary signs. MPS VII is diagnosed clinically in association with GUSB enzyme deficiency and/or in association with two disease-associated variants in the *GUSB* gene. Urinary glycosaminoglycans are elevated in patients with MPS VII. MPS VII is the most common lysosomal storage disorder associated with non-immune hydrops fetalis (NIHF). Molecular testing is used to determine the etiology of NIHF in order to reduce risk of poor postnatal outcomes.
Methods: We tabulated data from multiple clinical laboratories, ClinVar, and from a consolidated literature search to develop a comprehensive landscape of variants found in the *GUSB* gene.
Results: As of 17 May 2023, we report 318 unique *GUSB* variants distributed as: 221 (70%) missense, 21 (7%) nonsense, 21 (7%) splicing, 16 (5%) frameshift, 9 (3%) UTR, 7 (2%) intronic, 6 (2%) CNV, 3 (<1%) deletion, 1 (<1%) insertion, 2 (<1%) start-loss, and 11 (3%) synonymous. 314 variants were associated with clinical classifications distributed as 65 Pathogenic, 39 Likely Pathogenic, 209 Variants of Uncertain Significance (VUS), and 1 conflicting variant. Of the variants in this study, 110 were recurrently encountered having been identified from at least two of the sources: ClinVar, clinical labs, and literature review. Of the remaining variants, 138 were uniquely identified from ClinVar, 32 from clinical labs and 38 from literature review.
Conclusions: We recently found that approximately half of the patients with MPS VII had one or more VUS. These data highlight the value of timely submission of novel and previously observed variants to public databases in association with clinical and biochemical evidence. Cataloguing and disseminating information about *GUSB* variants is essential to improve timely and accurate diagnosis of this ultrarare disease. Moreover, clinical study of *in utero* fetal therapy for MPS VII relies heavily on molecular diagnosis of MPS VII. With the advent of genome-based newborn sequencing, timely reporting and classification of MPS VII disease associated variants is also necessary to ensure an accurate diagnosis. These data highlight the value for rapid reporting of novel and previously observed variants to public databases for more ready access by clinicians to promote timely, accurate diagnosis of MPS VII.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session II

PB4532 Navigating the Path to Diagnosis: An Efficient Workflow for AADC Deficiency Assessment.

Authors:

N. Liu[1], J. Odom[1], C. Gijavanekar[1], T. Wilson[1], Q. Sun[2], L. Emrick[2], R. Sutton[2], K. Hyland[3], N. Blau[4], K. Glinton[1], S. Elsea[1]; [1]Baylor Coll. of Med., Houston, TX, [2]Baylor Coll. Med., Houston, TX, [3]Med. Neurogenetics Lab., a LabCorp company, Atlanta, GA, [4]Univ. Children s Hosp. Zurich, Zurich, Switzerland

Aromatic L-amino acid decarboxylase deficiency (AADCD) leads to a combined deficiency of neurotransmitter biosynthesis, resulting in a broad and heterogeneous phenotypic spectrum including developmental delays, hypotonia, autonomic dysfunction, and movement disorder with/without oculogyric crises in infancy. Early diagnosis is necessary to allow prompt therapeutic interventions. However, diagnosis is often delayed or missed due to the limitation of historical diagnostic approaches that require targeted analyte testing of cerebrospinal fluid and the rarity of the disorder, which necessitates access to specialized clinical expertise. Furthermore, genomic data alone are often insufficient for diagnosis and require additional testing to assess the pathogenicity of the variants identified. Toward improving diagnostic processes, a comprehensive review of the published literature on AADCD was conducted to document all reported AADCD cases to date. We then retrospectively assessed clinical and research genomic (>20,000 exomes) and metabolomic (>6,000 plasma & urine samples) databases and curated 343 unique AADCD cases to create a comprehensive database of metabolo-genomic and phenotypic variation to support efficient diagnostics for AADCD. This comprehensive review identified 102 *DDC* variants representing 117 unique genotypes. The presence of only 30 of these disease associated *DDC* variants in ClinVar further highlights the diagnostic challenge. AADC enzyme testing and targeted testing for 3-O-methyldopa (3-OMD) were commonly used to confirm diagnosis; however, results may be confounded by treatment with dopamine or vitamin B6 supplementation. Using untargeted metabolomics, our laboratory showed that biomarker profiles from both plasma (3-OMD, VLA, 3-methoxytyramine sulfate, dopamine-3-O-sulfate, VMA) and urine (3-OMD, VLA, vanilloylglycine, VMA) effectively distinguish affected individuals from carriers, and from patients taking dopamine medications and/or vitamin B6 supplements. This supports the efficacy of untargeted metabolomic screening for early diagnosis of AADCD. Current guidelines for AADCD diagnosis are invasive, costly, burdensome, and limit access to care. For AADCD and other inborn errors of metabolism with non-specific clinical presentations, implementing a non-invasive metabolo-genomics diagnostic workflow early in clinical evaluation will shorten the diagnostic journey, alleviate concerns regarding misdiagnosis, and importantly, will decrease health disparities and broaden access to effective treatment.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session III

PB4533 NCBI clinical resources to support genetic testing result interpretation and clinical action

Authors:

A. Malheiro, M. Kane, M. Hoeppner, T. Venkatapathi, B. Kattman; NCBI, NLM, NIH, Bethesda, MD

Expanded access to genetic testing prompts the need to educate the whole spectrum of the healthcare community about phenotypes with a genetic component. Healthcare providers are faced with genetic test results from clients that undertook consumer-initiated options or that participated in research studies like the All of Us program and need to discuss clinical implications. A suite of clinical resources at NCBI exists to support healthcare professionals who receive non-clinical testing results that may impact a variety of areas of clinical care. These reports may indicate risk alleles for cancer predisposition, complex diseases, genetic diseases, carrier status for prenatal planning, or variants that impact drug metabolism. Clinicians may need to confirm the test results via a suitable clinical genetic test. The NIH Genetic Testing Registry (GTR) is a free resource with over 75,000 clinical tests from more than 500 laboratories in 48 countries. Clinicians can search for tests by gene or disease, filter results by lab location and CLIA certification, then review the test details including methodology, clinical utility, and analytical validity. The clinician may need to research diseases or phenotypes that have been uncovered during testing. Information including diagnosis, prognosis, and treatment for many conditions is available in MedGen. MedGen compiles disease terminology, definitions, clinical features, and highly relevant literature from PubMed and the NCBI Bookshelf, including clinical practice guidelines. MedGen links to authoritative sources for genetic disease information, including OMIM and OrphaNet. For pharmacogenomics, NCBI has a dedicated resource: Medical Genetics Summaries (MGS), which includes concise, structured reviews on genetic variants and drug responses along with recommendations from authoritative sources including the FDA, Clinical Pharmacogenetic Implementation Consortium (CPIC), and PharmGKB. Each expert-reviewed chapter discusses one drug and uses standardized terminology. MedGen, GTR, and MGS link to molecular resources like ClinVar as well as patient-focused resources like MedlinePlus Genetics and NCATS Genetic and Rare Diseases to support clinicians in test interpretation and patient education. By making relevant, data-rich information readily available, NCBI aims to better equip healthcare providers to incorporate genetics into their patients' care and be well informed when faced with genetic test results they did not initiate themselves. This work was supported [in part] by the National Center for Biotechnology Information of the National Library of Medicine (NLM), National Institutes of Health.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session I

PB4534 NCBI Variant Databases for Identification and Analysis of Common and Somatic Variants

Authors:

T. Hefferon, Q. Wang, H. Zhang, J. Garner, G. Shekhtman, R. Villamarin-Salomon, L. Phan; NIH/NLM/NCBI, Bethesda, MD

Accurate identification and filtering of somatic mutations, including single nucleotide variants (SNVs) and structural variations (SVs), are crucial for precision medicine in genetic diseases such as cancer. Common genetic variation databases like dbSNP and NCBI ALFA have been instrumental in this process. Additionally, the incorporation of the Database of Genomic Structural Variation (dbVar) enhances the analysis of somatic SVs.
dbSNP catalogs genetic variations, allowing differentiation between germline and somatic mutations. Comparing genomic data with dbSNP enables the exclusion of germline variants, focusing on pathogenic somatic SNVs. The database's extensive variant collection and population frequency annotations aid in identifying clinically relevant somatic SNVs. NCBI ALFA, within dbSNP, provides curated data on somatic mutations, particularly in cancer, improving the accuracy of somatic mutation analysis. It integrates information from sources like The Cancer Genome Atlas (TCGA), facilitating the identification and filtering of cancer-specific somatic mutations. In addition to dbSNP and NCBI ALFA, dbVar is crucial for analyzing somatic SVs. It catalogs SVs associated with various diseases, including insertions, deletions, duplications, and chromosomal rearrangements. Incorporating dbVar data enhances the identification and filtering of somatic SVs implicated in disease development.
By integrating dbSNP, NCBI ALFA, and dbVar, researchers achieve comprehensive analysis of somatic mutations (SNVs and SVs) for precision medicine. These databases serve as reliable references for identifying and filtering somatic mutations, ensuring accurate interpretation and facilitating personalized treatment strategies.
In conclusion, the combination of dbSNP, NCBI ALFA, and dbVar databases significantly enhances the analysis of somatic mutations in precision medicine. Leveraging these resources enables the identification and filtering of somatic SNVs and SVs, leading to improved patient outcomes and a deeper understanding of genetic diseases.
Acknowledgment: This work was supported by the National Center for Biotechnology Information of the National Library of Medicine (NLM), National Institutes of Health.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session II

PB4535 New developments in the RD-Connect GPAP facilitate research on diagnosis and gene discovery for rare diseases

Authors:

I. Gut[1], M. Bayes[1], A. Blanco[1], G. Bullich[1], A. Corvo[1], S. Chaves[1], J. Dieguez[1], M. Fernandez-Callejo[1], A. Garcia[1], L. Kraatz[1], S. Laurie[1], H. Lochmuller[2], L. Matalonga[3], I. Paramonov[4], A. Papakonstantinou[1], G. Parra[1], D. Pico[1], S. Medrano[1], I-D. Teoc[1], R. Thompson[2], R. Tonda[1], O. Torres[1], J-R. Trotta[1], G. Vazquez[1], L. Zalatnai[1], D. Piscia[1], S. Beltran[5]; [1]CENTRO NACIONAL DE ANALISIS GENOMICO (CNAG-CRG), Barcelona, Spain, [2]Univ. of Ottawa, Ottawa, ON, Canada, [3]CNAG-CRG, Barcelona, Spain, [4]Centro Natl. de Analisis Genomico, Barcelona, Spain, [5]Centro Natl. de Análisis Genómico (CNAG), Barcelona, Catalonia, Spain

The RD-Connect Genome-Phenome Analysis Platform (GPAP) is an IRDiRC recognised resource that facilitates collation, discovery, sharing, and analysis of standardised genome-phenome data within a collaborative environment. The platform aims to facilitate Rare Disease (RD) research on diagnosis and gene discovery to non-bioinformaticians. The RD-Connect GPAP is an integral part of the European Joint Programme on Rare Diseases (EJP-RD) Virtual Platform and is a key component of the EU Solve-RD project, which has contributed to the diagnosis of hundreds of patients with previously inconclusive genomic tests. The RD-Connect GPAP securely hosts pseudonymized data from over 28,000 patients and relatives. Pheno-clinical data may be batch submitted or entered through user friendly interfaces. The data is standardized with the Human Phenotype Ontology (HPO), the Orphanet Rare Disease Ontology (ORDO), and OMIM. Raw or aligned sequencing data can be submitted in FASTQ, BAM or CRAM format. Data can be exported as GA4GH VCF and Phenopackets, among others. The submitted datasets are made accessible to the rest of the validated users after a maximum of 6 months, although justified extensions might be granted. The RD-Connect GPAP is connected to both the Beacon and the MatchMaker Exchange networks.
The RD-Connect GPAP is in continuous development. As part of these efforts we have recently launched a new graphical user interface (GUI) to improve general usability, as well as data submission, collation, analysis and interpretation. The new and intuitive GUI is fully integrated with the other RD-Connect GPAP modules, such as data management, pheno-clinical information, cohort analysis and the integrated IGV Browser. Among other features, the genomics analysis module now better supports reproducible research, with extended functionalities to save and share studies, and to label and tag variants of interest. New annotations such as REVEL, Splice-AI and InterVar ACMG classification have been added. Cohort analysis has been simplified so that users can create *in silico* cohorts and run gene discovery analysis across hundreds of patients with only a few steps.
The RD-Connect GPAP hosted by the Centro Nacional de Análisis Genómica (Barcelona, Spain) and is available at https://platform.rd-connect.eu. We welcome registration and data submission from international RD researchers complying with EU General Data Protection Regulation.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session III

PB4536 NIA Genetics of Alzheimer's Disease Data Storage Site (NIAGADS): 2023 Update.

Authors:

L. Armus[1,2], A. Kuzma[1,2], O. Valladares[1,2], H. Nicaretta[1,2], J. Manuel[1,2], P. Gangadharan[1,2], Z. Katanic[1,2], M. Kirsch[1,2], Y. Ren[1,2], N. Saravanan[1,2], A. Wilk[1,2], Y. Zhao[1,2], L. Qu[1,2], M. Moon[1,2], A. Lerro-Rose[1,2], P. Keskinen[1,2], J. Cifello[1,2], W. Horng[1,2], S. Tate[1,2], F. Robbins[1,2], H. White[1,2], L. Carter[1,2], S-Y. Chou[3], A. Naj[1,4], E. Greenfest-Allen[2,5], P. P. Kuksa[1,2], Y. Leung[1,2,6], W-P. Lee[1,2], G. Schellenberg[1,6], L-S. Wang[1,2,6]; [1]Penn Neurodegeneration Genomics Ctr., Dept. of Pathology and Lab. Med., Philadelphia, PA, [2]Inst. for BioMed. Informatics, Philadelphia, PA, [3]Dept. of Economics, Lehigh Univ., Bethlehem, PA, [4]Dept. of Epidemiology, Univ. of Pennsylvania Perelman Sch. of Med., Philadelphia, PA, [5]Dept. of Genetics, Philadelphia, PA, [6]Inst. on Aging, Philadelphia, PA

Background: NIAGADS is a national genomics data repository that facilitates access of genotypic and sequencing data to qualified investigators for the study of the genetics of Alzheimer's disease (AD) and related neurological diseases. Collaborations with large consortia and centers such as the Alzheimer's Disease Genetics Consortium (ADGC), Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium, the Alzheimer's Disease Sequencing Project (ADSP), and the Genome Center for Alzheimer's Disease (GCAD) allow NIAGADS to lead the effort in managing large AD datasets that can be easily accessed and fully utilized by qualified investigators in the research community.
Methods: NIAGADS is supported by National Institute on Aging (NIA) under a cooperative agreement. All data derived from NIA funded AD genetics studies are expected to be deposited in NIAGADS or another NIA approved site. NIAGADS manages a Data Sharing Service (DSS) that facilitates the deposition and sharing of genomic data and association results with approved users. In addition, researchers may freely use the NIAGADS Alzheimer's Genomics Database (www.niagads.org/genomics/) to search annotation resources that link published AD studies to AD-relevant sequence features and genome-wide annotations.
Results: As of June 2023, NIAGADS houses 99 datasets comprised of >172,000 samples including GWAS, sequencing, gene expression, annotations, deep phenotypes, and summary statistics. Qualified investigators can retrieve ADSP sequencing data with ease and flexibility through the NIAGADS DSS. To date, the ADSP and other contributing studies have completed whole exome sequencing (WES) of 20,503 samples and whole genome sequencing (WGS) of 36,361 samples. Raw WES and WGS files, quality controlled VCF files, and phenotype data files are available via qualified access. The next round of sequencing currently underway will generate around 25,000 additional genomes to be released in late 2023.
Conclusion: NIAGADS is a rich resource for AD researchers, with the goal of facilitating advances in Alzheimer's genetics research by sharing diverse datasets from a myriad of projects and institutions for secondary analysis by qualified researchers. Datasets, guidelines, and features are available on our website at https://www.niagads.org.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session I

PB4537 Omics Research in the Environmental influences on Child Health Outcomes (ECHO)-wide Cohort Study (EWC)

Authors:

G. Page, ECHO Program; RTI Intl., Atlanta, GA

The goal of the Environmental influences on Child Health outcomes (ECHO) Program (www.echochildren.org) is to understand the effect of a broad range of early environmental and omics influences on child health and development. ECHO is dedicated to both learning what factors affect child health and to finding ways to enhance it. ECHO has a collaborative research design comprising 69 cohorts in 31 consortia and was funded by the National Institutes of Health in 2016 and is renewed through at least 2030. Surveys, interviews, standardized examinations, laboratory analyses, and medical record abstraction are used to obtain information in five main outcome areas: pre-, peri-, and post-natal outcomes; neurodevelopment; obesity; airways; and positive health. Exposures include place- (e.g., air pollution, socioeconomic status), family- (e.g., parental mental health), and individual-level (e.g., diet, genomics) factors. The inclusion of children at different life stages allows for the evaluation of time-varying effects of exposures and critical periods of development. As of May 26, 2023, the EWC has data from 67,757 children and 35,559 have been consented for the EWC data and biospecimen collection. While the EWC was being implemented, the extant omics data that had been generated by the individual cohorts were aggregated and harmonized across cohorts. Extant genome wide methylation data are available on 8,387 samples from 5,990 participants including 614 parent-offspring pairs. Many participants have multiple tissues' data (n = 1,077 participants) or the same tissue at different ages (n = 1,742 participants). Extant genetic data are available on 5,794 participants from 13 cohorts. Data were generated on multiple single-nucleotide polymorphism (SNP) microarray platforms with 370,150 to 1,730,897 SNPs per sample. All the data have been cleaned to the same standard and mapped to the human genome build 38 and imputed using the TOPMed (Version r2 2020) reference panel. 11,681 microbiome samples from 5682 EWC participants, primarily 16S, are also available. This includes paired samples from 650 mother-infant dyads. Thousands of EWC participants have multi-omics data. Additional genetic, epigenetic, microbiome and metabolomics data on the EWC participants are being generated. ECHO investigators are conducting numerous omics-based studies on the EWC participants including studies of genetics of early childhood growth, robustness of epigenetic clocks across age and tissue, and effects of air pollution on the epigenome. The EWC is a powerful collaborative resource for studying child health and development whose power and utility will continue to increase over time.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session II

PB4538 Pan Genomic Analyses and GWAS of *Mycobacterium fortuitum from different isolation sources*

Authors:

B. Yakubu; Kwame Nkrumah Univ. of Sci. and Technology, Kumasi, Ghana

Mycobacterium fortuitum is a specie of rapidly growing non-tuberculous mycobacteria that can be found in soil, water, and various environmental sources including endoscopes, and prosthetic devices. *M. fortuitum* causes infections in humans, especially in those with compromised immune systems, such as individuals with cystic fibrosis, cancer, or HIV/AIDS. However, treatment of *M. fortuitum* infections involving a combination of antibiotics sometimes fail due to development of novel strains, virulent/antibiotic resistant genes and inadequate knowledge on the specific genes which have strong association with its virulence. To unravel the interplay of these genomic features of the bacterium on disease onset, Pangenome and GWAS analyses was conducted using Roary pangenome pipeline, including Prokka, Abricate, SCOARY. Downstream statistical analyses and visualizations were also performed using R-programming, Uniprot Ugene, Phandango and google sheets. Five different clusters were constructed along a nine-layer phylogenetic tree for similarity and evolutionary analyses. A total of 22526 pan genes were identified with 41.99% of them strongly associated with the identified traits. Mirage of both virulent and AMR genes found whiles some obsolete antibiotics were detected. In conclusion, we noted that *M. fortuitum* bacterium is quite diverse with the potential of increasing the diversity in future, relatively virulent and may continue to resist a number of Antibiotics.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session III

PB4539 PangyPlot: A Visualization Tool for Exploring Pangenomes

Authors:

S. Mastromatteo, F. Lin, K. Keenan, S. Chirmade, B. Thiruvahindrapuram, Z. Wang, D. Roshandel, J. Whitney, S. Allana, L. J. Strug; The Hosp. for Sick Children, Toronto, ON, Canada

A persistent issue with traditional linear reference genomes is that they fall short in capturing the full genetic diversity present in populations. Dependency on a singular reference sequence introduces reference bias — under-representation of alleles that differ from the reference genome — leading to a skewed and inaccurate understanding of genetic variation. Pangenomes (often implemented as graph genomes) overcome this limitation by incorporation of all genetic variations and structural variations found in a population into a single unified representation. Recent improvements in long-read sequencing technologies have reduced the complexity of assembling reference-quality genomes, making the construction of a graph genome viable.

Use of a graph genome in practice is challenging because the landscape of supporting bioinformatics tools remains sparse. Particularly, visualizing the structure of a graph genome is unintuitive despite being a critical part of linear genome-based analysis. Software such as Bandage and Sequence Tube Maps can visualize local regions of a graph but are less effective with complex structures and lack a robust annotation system.

Here we present PangyPlot, a novel browser-based tool designed to facilitate the exploration of pangenome structures. PangyPlot provides a dynamic and interactive visualization of graph genome structure at different resolutions. Users can view complex structural variation and zoom in to annotated SNP-level details. Unlike existing tools, PangyPlot incorporates annotations within the visualization as well as the ability to collapse complex structures and highlight paths along the graph corresponding to haplotypes.

We evaluated PangyPlot using the Human Pangenome Reference Consortium (HPRC) data along with 101 hybrid genomes assembled using Pacific Biosciences long-read sequencing and linked-reads (average N50=45Mb, L50=20) from individuals with cystic fibrosis (CF). We demonstrate the potential of the tool by visualizing the *PRSS1-PRSS2* CF modifier locus that contains a common 20 kb deletion polymorphism as well as the *SLC9A3* locus that is dense with large tandem repeats. Both regions impact CF severity and are difficult to characterize using traditional linear reference-based approaches.PangyPlot provides powerful and intuitive visualization for exploring pangenome structures. Leveraging graph-based representations provide deeper visual insights into genetic diversity and identification of genomic features that would be commonly missed due to reference bias. A version of PangyPlot that visualizes HPRC and our 101 genomes is being made available as a public web application.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session I

PB4540 PhenoDB and CAVATICA: Container-Based Annotation and Analysis

Authors:

L. Vail[1], S. Griffith[1], R. Martin[2], M. Brown[3], D. Miller[4], Y. Guo[5], A. Heath[3], Y. Zhu[6], N. Sobreira[2]; [1]Johns Hopkins Genomics, Baltimore, MD, [2]Johns Hopkins Univ, Baltimore, MD, [3]Children's Hosp. of Philadelphia, Philadelphia, PA, [4]Childrens Hosp. of Philadelphia, Philadelphia, PA, [5]Children s Hosp. of Philadelphia, Philadelphia, PA, [6]The Children s Hosp. Of Philadelphia, Havertown, PA

PhenoDB, a web-based tool, provides methods to annotate and analyze the variants of patients and their family members for links to diseases. CAVATICA is a cloud-based data analysis platform that allows users to run various analyses over their genomic data in a HIPPA complaint environment.
To provide access to some of the PhenoDB analysis tools in CAVATICA, the PhenoDB, CAVATICA, and D3b/CHOP teams have collaborated to create a CAVATICA app that provides a baseline variant analysis extracted from PhenoDB.
CAVATICA apps are built by encapsulating functionality into Docker containers, which run using the Common Workflow Language (CWL) to collect parameters from users. The encapsulated PhenoDB workflow was split into two containers so that each step can be run independently.
The first step is the ANNOVAR annotation, which is the input for PhenoDB analysis. As parameters, it accepts the user's choice of assembly (Hg19 or Hg38), the VCF files to be annotated, and the ANNOVAR reference data. The second step is a PhenoDB analysis based on the output of the first step. The parameters for the second step are: the characteristics of each individual included in the analysis (affected status, relationship to the proband, sex, and annotated VCF file), inheritance patterns, minor allele frequency, and RefGene gene location (variant type).
The PhenoDB app on CAVATICA provides a graphical user interface to configure the analysis, which can be run using any file stored on CAVATICA. All jobs executed on CAVATICA are retained, can be edited, and rerun. For collaboration, users can share their CAVATICA projects with other users.
This initial collaboration between the PhenoDB, CAVATICA, and D3b/CHOP teams is a starting point that we expect will lead to further PhenoDB functionality being added to CAVATICA. The creation of containers encapsulating PhenoDB analysis functionality also provides a gateway for collaborations on other cloud-based genomics analysis platforms.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session II

PB4541 Polygenic Interactive Discovery System: an interactive knowledgebase for exploring complex genetics networks

Authors:

M. Milojevic[1], R. Z. Marandi[1], B. Sherman[2], W. Chang[2], J. D. Lundgren[1], D. D. Murray[1], C. R. Macpherson[1], P. Leung[1]; [1]Ctr. of Excellence for Hlth., Immunity and Infections (CHIP), RigsHosp.et, Copenhagen, Denmark, [2]Lab. of Human Retrovirology and Immunoinformatics, Frederick Natl. Lab. for Cancer Res., Frederick, MD

Genome Wide Association Studies (GWAS) are used to assess the impact of single nucleotide polymorphisms (SNPs) on a specific phenotype. While widely used and beneficial, this method has limitations, namely strict control for type 1 error at the expense of type 2 error and an inability to account for pleotropic effects at the gene or pathway level. Further, GWAS output is not easy to explore in connection with existing literature to identify supporting/contradicting studies and develop new hypotheses. To address these concerns, we developed the Polygenic Interactive Discovery System (PIDS), an interactive genetics knowledgebase that facilitates the study of genetic effects on disease with the knowledge of relevant molecular pathways and prior studies. Data from Ensembl, GWAS Catalog, ClinVar, PGS Catalog, and Reactome were extracted and merged into a single, searchable data structure to form the knowledgebase. Users of PIDS are also able to upload their own SNP level association data. Output from any of the resources outlined above is mapped to the network generated from the knowledgebase, which can then be explored and added to using an interactive R shiny interface. The current version of PIDS contains a comprehensive biological network with 19496 genes and biologically relevant small molecules, connected by 353,889 edges. All resources were harmonized to a combined total of 326,948 SNPs with single-nucleotide effects studied across 4901 studies, 1775 PRS studies describing pleiotropic effects for 390,235,794 rsID-gene pairs. SNP co-ordinates were cross-checked with Ensembl's human gene list (N = 40,785 genes, genome build: GRCh38) and variation file for rsID matching. To showcase PIDS, networks were created for two existing datasets and explored through the graphical interface. First, we created a network using an existing polygenic risk score (PRS) for Type 1 diabetes and mapped results from a separate GWAS study, observing commonality between these two studies. Second, we created a network from an existing HIV GWAS and mapped other GWAS exploring the same phenotype and other viral infections, showing that genetic variation in HLA-B is important across these independent studies. In conclusion, PIDS is a versatile tool that leverages data from five distinct databases to complement genetic association studies with knowledge of biological function and enhance them by connecting with existing studies in related areas. The system aids the interpretation of SNPs at the resolution of one or more genes and may be used to explore functions of SNP at a gene and pathway level by providing a deeper, more connected view for overall interpretation and hypothesis generation.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session III

PB4542 Practical recommendations for TOPMed metabolomics data.

Authors:

F. Ockerman[1], L. Zhou[1], E. Drzymalla[1], T. Alkis[2], M. Grove[2], B. Yu[2], L. Raffield[1]; [1]Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, [2]Univ. of Texas HSC at Houston, Houston, TX

The Trans-Omics for Precision Medicine (TOPMed) program expects to soon release over 90,000 samples with broad-spectrum metabolomic data, representing over a dozen studies. However, investigators using this resource face potential challenges in pre-processing and integrating data across studies. Differing metabolomic platforms and analysis centers may cause technical variation. Likewise, missing metabolite values may vary in their distribution and source between studies. Consistent protocols for pre-processing and integration are thus necessary to unlock the potential of this rich resource. We compare several strategies and offer recommendations for the TOPMed community, with the goal of guiding and facilitating future genetic and phenotype-specific analyses.

As a pilot phase, we are currently analyzing data from 25,058 participants from diverse case-control and population-based cohort studies, including 15,633 participants from 3 cohort studies on the Metabolon platform and 9,425 participants from 5 cohort studies on the Broad/BIDMC platform. This dataset includes 1,730 named metabolites, including 364 metabolites measured in at least some cohorts across both platforms. With within-study rank-based inverse normal transformation, we demonstrate that estimates of age-metabolite associations are highly concordant ($r > 0.999$), and generally consistent with the existing literature, between pooled and inverse variance meta-analyzed data, although 36 metabolites are significant only in the meta-analysis. Most named metabolites had very low missingness in our dataset, and we found that metabolite associations with age and sex were highly consistent across all missingness imputation strategies (zero, min, half-min, k-nearest neighbors, random forest, quantile regression imputation of left censored data). We recommend replacing missing values with zero in metabolites characterized as xenobiotics. For other metabolites, we will compare imputation strategies with an analysis of metabolite quantitative trait loci (mQTLs).

In summary, we find largely consistent results in pooled and inverse variance meta-analysis. We recommend inverse-normal transformation to enable integration between studies. We recommend left-censored imputation for xenobiotics and will soon release recommendations for imputation in other metabolites. To aid investigators, we will release scripts for implementing these recommendations. Such pre-processing steps are necessary to optimize power in cross cohort metabolomic analysis, including planned QTL studies.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session I

PB4543 Profiling senescent cells in snRNA-seq data.

Authors:

M. Koutrouli[1], J. Bendl[2], D. Lee[3], P. Fnu[4], G. E. Hoffman[5], V. Haroutunian[6], J. F. Fullard[4], J. J. Lars[1], P. Roussos[5]; [1]Univ. of Copenhagen, Copenhagen, Denmark, [2]Icahn Sch. of Med. at Mount Sinai, New York City, NY, [3]Ctr. for Disease Neurogenomics and the Icahn Inst. for Data Sci. and Genomic Technology and the Dept.s of Psychiatry and Genetics and Genomic Sci., Icahn Sch. of Med. at Mount Sinai, New York, NY, USA, New York, NY, [4]Ctr. for Disease Neurogenomics, Icahn Sch. of Med. at Mount Sinai, New York, NY, [5]Icahn Sch. of Med. at Mount Sinai, New York, NY, [6]Dept. of Psychiatry, Icahn Sch. of Med. at Mount Sinai, New York, NY

Aging is a complex phenomenon with many factors, or hallmarks, contributing to its advancement. These hallmarks include cellular senescence, genomic instability, telomere attrition, epigenetic alterations, loss of proteostasis, deregulated nutrient sensing, mitochondrial dysfunction, stem cell exhaustion, and altered intercellular communication. Among these potential causes, cellular senescence stands out as having a particularly significant role in the aging process and age-related diseases, including Alzheimer's disease.Cellular senescence is a process in which cells permanently stop dividing and enter a state of permanent growth arrest. It is triggered by a variety of stressors, including DNA damage, telomere shortening, oxidative stress, oncogene activation, and changes in chromatin structure. Senescent cells (SNCs) can secrete pro-inflammatory molecules, chemokines, and growth factors that are thought to influence the surrounding cellular environment, ultimately leading to a range of age-related pathologies, including Alzheimer's disease.Excitatory neurons are one of the most abundant types of neurons in the brain and play a critical role in cognitive functions. Recent studies have shown that cellular senescence also occurs in excitatory neurons, leading to their dysfunction. The accumulation of senescent excitatory neurons is believed to contribute to the development and progression of Alzheimer's disease.To gain a deeper understanding of the role played by senescent excitatory neurons in Alzheimer's disease, our research aimed to identify and analyze these neurons within the PsychAD cohort. This cohort comprises a vast collection of 8,342,664 cells obtained from 1,495 unique donors. Our specific focus was on establishing biomarkers and characterizing excitatory senescent neurons in both Alzheimer's disease and healthy brain tissue.Our preliminary findings have revealed a significant correlation between the concentration of excitatory senescent neurons and Alzheimer's disease. By characterizing these senescent excitatory neurons, we hope to shed light on the underlying mechanisms that lead to their dysfunction and demise in the context of Alzheimer's disease as well as identify potential therapeutic targets for intervention and treatment.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session II

PB4544 Quantifying the incredible pleiotropy of microRNAs: focusing ChatGPT on the literature

Authors:

M. McGeachie[1], S. Begum[1], R. Sharma[2], K. R. Wang[3]; [1]Brigham & Womens Hosp., Boston, MA, [2]Harvard Med. Sch. & BWH, Boston, MA, [3]Roxbury Latin, Boston, MA

MicroRNAs (miRNAs) are recognized as key regulatory factors in numerous severe diseases, with the same miRNA often involved in several diseases simultaneously or being identified as biomarker for dozens of separate diseases. While of evident biological importance, miRNA pleiotropy remains poorly understood, and a robust measure of pleiotropy could greatly aid in understanding the broader role miRNAs play in health and disease. To this end, we introduce miRAIDD (miRNA Artificial Intelligence Disease Database), a comprehensive database of miRNA-disease causal associations constructed using OpenAI's ChatGPT model and describe its implications for understanding the general pleiotropy of miRNAs in human disease.

We employed ChatGPT to read over 60,000 abstracts from PubMed mentioning miRNAs and used it to determine if the abstract described evidence of a causal role for the miRNA in the condition or trait described therein. We aggregated these results into miRAIDD, totaling 85,454 annotations, 51,410 unique papers, and 23,346 unique papers that demonstrate a causal miRNA-disease relationship; these represent a marked increase over existing human-curated miRNA-disease collections. The performance of the AI model in generating the database was found to be comparable to human annotators and existing miRNA-disease association databases.

We combined miRAIDD with other publicly available sources of data on miRNA to identify factors influencing the pleiotropy of miRNAs, defined as the number of diseases a miRNA is reported to have caused. The most important factor influencing pleiotropy of a particular miRNA is the number of publications featuring the miRNA. The number of tissues a miRNA has been observed in and the number of validated and predicted gene targets were also significant predictors. After adjusting for publications, the date of discovery, evolutionary conservation, and number of gene targets predicted by computational algorithms (but, importantly, not the number of experimentally validated gene targets) were the strongest indicators of number of diseases caused. We identified over 100 miRNAs causing more than 50 distinct diseases each, which suggests that many miRNAs have substantial pleiotropy.

While much has been written about the potential future impact of strong AI models like ChatGPT impacting biomedical research, our investigation represents one of the first such implementations. The miRAIDD database is freely available for download, with the anticipation that it will further expedite research about miRNA publications, results, and pleiotropy.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session III

PB4545 † RefSeq Annotation and Curation of the T2T-CHM13 Human Genome Assembly

Authors:

T. Goldfarb, S. Pujar, J. Jackson, V. S. Joardar, V. K. Kodali, K. McGarvey, M. Murphy, A. Vatsan, D. Webb, F. Thibaud-Nissen, T. D. Murphy; Natl. Ctr. for Biotechnology Information, Natl. Library of Med., NIH, Bethesda, MD

The RefSeq project at the National Center for Biotechnology Information (NCBI) manually curates the publicly available reference genomes to represent a standard reference set of well supported and non-redundant reference sequences used by the medical and scientific communities. A large focus of RefSeq curation has been on the human reference genome, GRCh38. While the availability of this sequence has revolutionized science and medicine during the last two decades, the technology used at the time of its assembly resulted in some gaps and repeat regions that could not be properly sequenced, and therefore regions of the genome that could not be annotated, curated, and represented.
Recently, a new human assembly was released by the Human Pan Genome Consortium (HPRC), T2T-CHM13, using long-read sequencing technologies that overcomes some of the limitations of sequencing complex and repeat regions of the genome that cannot be sequenced using clone-based techniques. RefSeq annotation of the T2T-CHM13 assembly was made available in April 2022 and includes 82,862 curated RefSeq transcripts with experimental evidence for the annotation drawn from 9.7 billion RNA-seq reads, nearly 83 million PacBio and Oxford Nanopore long transcriptome reads, 8.6 million ESTs, and 345,700 GenBank cDNAs and corresponding proteins. This annotation is expected to provide valuable information and transcript sequences on some of the regions missing in the GRCh38 assembly. Preliminary analysis of the annotation suggests that there are more than 400 protein-coding genes uniquely annotated on T2T-CHM13, including important disease-related genes. Presented here are the results of RefSeq curation of the T2T-CHM13 annotation, including new RefSeqs for genes that are novel to the T2T-CHM13 assembly compared to the GRCh38 assembly. In addition, RefSeq has curated and represented transcripts found on T2T-CHM13 that represent alternate alleles of genes, including differences in coding potential (e.g. unitary pseudogenes), polymorphic splice sites, and some cases in which the T2T-CHM13 allele represents the major allele found within some populations. RefSeq will continue curation of the T2T-CHM13 assembly and expect that this effort will help users get the maximum use and benefit of this assembly in years to come.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session I

PB4546 Repository of Alzheimer's Disease and Related Dementia (RADR) and its application to diverse populations

Authors:

R. Vaza, A. Elmas, H. Wu, V. Subramaniam, A. Renton, A. Goate, P. O'Reilly, K-l. Huang; Icahn Sch. of Med. at Mount Sinai, New York City, NY

Alzheimer's disease and related dementias (ADRD) present significant public health challenges, and understanding their genetic underpinnings is crucial for developing effective interventions and risk prediction strategies. However, existing genetic resources are largely fragmented, limiting the development of genomic medicine for ADRD. We introduce the Repository of Alzheimer’s Disease and Related Dementia Variants (RADR), a curated database of 3530 rare genetic variants associated with ADRD. The database standardizes and integrates data from various sources, including Alzforum, ClinVar, and DNA-seq studies. Additionally, we explore the distribution and pathogenicity of genetic variants within RADR with a focus on pathogenic, likely pathogenic, and risk factors/modifiers. We demonstrate the application of RADR to two large-scale biobanks, BioMe (n = 30813) and the UK Biobank (n = 201,136), for genotype-phenotype correlations with ADRD-related phenotypes. The findings highlight incomplete penetrance, enrichment of certain variants, and potential population-specific differences. Overall, RADR provides a comprehensive and unified resource for researchers and clinicians studying ADRD, facilitating genetic variant analysis and the potential to aid in the development of personalized approaches for diagnosis, prevention, and treatment.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session II

PB4547 Resources for archiving and analysis of human data in Japan

Authors:

Y. Kodama[1], A. Fukuda[1], T. Tokimatsu[1], J. Mashima[1], T. Hachiya[2], M. Ishii[2], M. Kawashima[3], N. Mitsuhashi[3], T. Fujisawa[1], T. Tanjo[1], O. Ogasawara[1]; [1]Natl. Inst. of Genetics, Mishima, Shizuoka, Japan, [2]Genome Analytics Japan Inc., Tokyo, Japan, [3]Database Ctr. for Life Sci., Chiba, Japan

The Japanese Genotype-phenotype Archive (JGA, https://www.ddbj.nig.ac.jp/jga) as a centralized repository in Japan has been maintained jointly by the National Bioscience Database Center (NBDC) (https://humandbs.biosciencedbc.jp) and DDBJ. NBDC is responsible for the ethical guidelines and its Data Access Committee (DAC) grants data submission and data access requests to JGA. DDBJ accepts metadata, genotype and phenotype data associated with data samples, and assigns stable, unique accession numbers prefixed with ‘JGA’. JGA studies are indexed at the DDBJ Search (https://ddbj.nig.ac.jp/search). As of 5 June 2023, 339 studies (585,632 samples and 980 terabytes of data files) are available. DDBJ also provides a secured supercomputing environment for analyzing personal genome data. The secured system is connected with the JGA server through a high-speed network, and users can smoothly download and analyze JGA datasets as in a cloud environment. For the convenience of users, we pre-process raw genome sequencing data deposited at JGA by using the workflow "jga-analysis" (https://github.com/ddbj/jga-analysis), and provide alignment and variant call data along with the original raw data from JGA. Allele frequencies of analyzed variants are available at the DBCLS TogoVar (https://togovar.org), which is a database for variant annotation with allele frequencies among Japanese populations, without restriction.. We also provide the NBDC-DDBJ imputation server to enable researchers to execute imputation analysis by simply uploading their data and selecting an imputation panel dataset in the web interface. Researchers can use Japanese imputation panel datasets distributed by JGA upon a NBDC DAC approval.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session III

PB4548 Shared loci between post-traumatic stress disorder and cardiovascular phenotypes using EHR and heart imaging data from more than 1 million participants

Authors:

J. SHEN[1,2], W. Valentim[2,3], D. Koller[2,4], M. B. Stein[5], J. Gelernter[6], R. Polimanti[2], G. A. Pathak[2]; [1]Children's Hosp. of Soochow Univ., Suzhou, China, [2]Yale Univ. Sch. of Med., New Haven, CT, [3]Faculdade de Med. da Univ.e Federal de Minas Gerais, Belo Horizonte, Brazil, [4]Univ. of Barcelona, Catalonia, Spain, [5]Univ. of California San Diego, San Diego, CA, [6]Yale Univ Sch. of Med., West Haven, CT

Objectives: The aim of this study is to investigate the shared loci between post-traumatic stress disorder (PTSD) and cardiovascular disease (CVD) phenotypes, utilizing electronic health records (EHR) and heart imaging data, to gain a deeper understanding of the underlying mechanisms. Methods: We analyzed 76 autosomal loci associated with PTSD from the recent genome-wide association study (GWAS; N=1,307,247) of the Psychiatric Genomics Consortium (PGC). CVD EHR phenotypes (circulatory system -141 phecodes; endocrine/metabolic - 105 phecodes) were obtained from the Million Veteran Program (MVP; N=458,061). Cardiac and aortic structural and functional imaging traits were derived from UK Biobank (UKBB, N=26,893). We estimated global and locus-level genetic correlations (rg) between PTSD and CVD traits, using LDSC and LAVA respectively. To estimate probability of shared causality between PTSD and CVD traits we applied the coloc method. For replication, we also tested PGC-PTSD data with respect to phecodes from UKBB (N=420,531) and similar phenotypes available from other GWAS consortia (N=915,868). Statistical significance was defined considering false discovery rate (FDR-q<0.05). Results: Significant globalrg were found between PTSD and circulatory phecodes (100 traits, globalrg 0.107~0.661), endocrine/metabolic phecodes (53 traits, globalrg 0.072~0.551), and heart imaging (8 traits, globalrg 0.207~0.190). Of the 76 PTSD-associated loci, we observed 33 loci to be genetically correlated with 59 traits (absolute [localrg] - 0.159 ~ 1). Considering loci with evidence of local PTSD-CVD rg, we investigated shared causality using probability of (i) having the same causal variant for the two traits [H4-PP>80%]; and (ii) two LD-linked causal variants for the two traits [H3-PP>80%]. We found 20 CVD traits that shared the same causal SNP with PTSD at 12 loci, and 27 CVD traits where causal variants were different but LD-linked at 9 loci. Of these results, we found colocalizations for 18 traits at 12 loci that replicated in UKBB phecodes or similar phenotypes available from other GWAS consortia. We found eight traits including coronary atherosclerosis, atrial fibrillation, obesity, diabetes, and ischemic heart disease at four loci (3p21.31,7p22.3,17q21.32, 19p13.11) that shared causal variants with PTSD in all three cohorts. Conclusion: Overall, our study provides new insight by prioritizing specific loci that share potential causal variants between PTSD and several CVD conditions. These loci will be further investigated as drug repurposing targets and their molecular function, which may explain their shared mechanism.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session I

PB4549 Spit take: Improving high molecular weight DNA extraction from shelf-stable saliva for long-read sequencing.

Authors:

L. Morina[1], J. Hosea[1], C. Montano[2], W. Timp[1]; [1]Johns Hopkins Univ., Baltimore, MD, [2]Natl. Human Genome Res. Inst., Bethesda, MD

Single-molecule, long-read sequencing technologies have enormous potential in population genetics and epigenetics. Single-molecule platforms (PacBio/ONT) can generate reads long enough to span low-complexity genomic regions and distinguish covalently modified nucleotides directly on individual DNA molecules. However, these methods require high DNA input, and retaining high molecular weight (HMW) DNA fragments is required to best leverage these platforms' advantages. These limitations are particularly relevant when dealing with limited-quantity patient samples, whether whole blood, buffy coat, or saliva. In recent years, saliva samples have seen an increase in use for large-scale genetic studies. This is partly because saliva samples are ideal for clinical studies - collection does not require a trained individual, samples are stable at room temperature, and collection tubes can be shipped for collection. However, this leads to a unique set of challenges: contamination, an uncharacterized DNA degradation rate for room temperature samples, and a variable cellular composition within the sample. Our goal is to overcome these obstacles so that saliva can be more widely used for single-molecule sequencing of high molecular weight (HMW) DNA in large-scale genetic studies. We are developing a DNA sequencing protocol specifically tailored for HMW DNA from saliva, primarily for use with the Oxford Nanopore Technologies (ONT) PromethION platform. The protocol, which includes DNA extraction, shearing and size selection, and library prep, has been tested on a medley of saliva samples gathered within the past five years. So far, we have tested two HMW DNA extraction methods, two size selection methods, and two shearing methods and sequenced all samples with PromethION. In the future, we plan to assess three more extraction methods. We also plan to process whole blood and buffy coat samples gathered around the same time as the saliva samples for comparison. The initial saliva samples exhibited varying bacterial contamination but were sequenced regardless. Contamination was evaluated by visual inspection or after sequencing by percent mapping to the human genome by *minimap2*. We have established a basic filtering pipeline to accompany the final protocol using alignment (*minimap2)* and k-mer based classification (*kraken2)*. This experimental and computational workflow will be useful for the wider community to apply to saliva specimens, a convenient and cost-effective source of DNA for sequencing.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session I

PB4552 Telmisartan reduces T2D incidences in electronic health records

Authors:

A. Tiwari[1], Q. Wang[1], Q. Wei[1], R. Chen[1], Y. Tan[1], Y. Yan[1], B. Li[1], X. Zhong[2]; [1]Vanderbilt Univ., Nashville, TN, [2]Vanderbilt Univ. Med. Ctr., Nashville, TN

Background: Essential hypertension and Type 2 Diabetes (T2D) are common conditions that often coexist. Some medications treating blood pressure have been found to also affect diabetes due to shared targets. Our previous work suggested that telmisartan, an anti-hypertensive medication, may protect against T2D as the drug reversed the disease signature of T2D based on an integrative analysis of T2D genetics and drug-induced perturbation experiments. **Objective:** In this project, we use electronic health records (EHR) from Vanderbilt University Medical Center (>3 million de-identified individuals) to validate the potential therapeutic effect of telmisartan on T2D.**Methods:** We first constructed a telmisartan drug user cohort including 40,241 adults diagnosed with essential hypertension, and two types of control cohorts: 1) non-telmisartan user cohort, and 2) comparator cohorts of hypertensive patients taking other anti-hypertensive medications. Each control cohort was matched by age, gender, body mass index (BMI), and disease comorbidities to the telmisartan cohort using propensity score. We then estimated the hazard ratios (HRs) and 95% CIs using stratified Cox proportional hazard models, adjusting for age at diagnosis, gender, BMI, and disease comorbidities. **Results:** Among the hypertension patients, we found that usage of telmisartan was associated with a reduced risk of T2D (HR 0.48; 95% CI 0.44 to 0.52; P<2e−16) compared to the non-telmisartan user control cohort. Telmisartan was also significantly associated with a decreased risk of T2D for all three comparator cohorts (clonidine (HR 1.33; 95% CI 1.22 to 1.45; P<1.07e-10), metoprolol (HR 1.22; 95%CI 1.12 to 1.34; P<6.33e-06), and timolol (HR 1.29; 95% CI 1.14 to 1.47; P<8.8e-05)). **Conclusions:** Our analysis of EHR data provides further evidence supporting the protective effect of telmisartan against T2D. Further validation such as randomized clinical trials is needed. **Keywords:** hypertension drug, T2D, drug repurposing, electronic health records, insulin resistance, clinical outcomes.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session II

PB4553 The Alzheimer's Disease Sequencing Project - Discovery, Discovery Extension and Follow Up Study ADSP-FUS: *APOE* genotype status and demographic characteristics across datasets.

Authors:

P. Mena[1], A. Zaman[1], K. faber[2], L. Adams[1], J. Inciute[1], P. Whitehead[1], T. Foroud[3], D. Reyes-Dumeyer[4], A. Kuzma[5], H. Issen[6], A. Naj[7], E. Martin[8], C. Dalgard[9], G. Schellenberg[10], L-S. Wang[11], R. Mayeux[12], B. Vardarajan[13], J. Vance[1], M. Cuccaro[14], B. Kunkle[1], M. Pericak-Vance[8]; [1]Univ. of Miami, Miami, FL, [2]Natl. Centralized Repository for ADRD, Indianapolis, IN, [3]Indiana Univ Sch Med, Indianapolis, IN, [4]Columbia Univ., New York City, NY, [5]Univ Pennsylvania, Philadelphia, PA, [6]Univ. of Pennsylvania, Glenside, PA, [7]Univ. of Pennsylvania, Philadelphia, PA, [8]Univ. of Miami Miller Sch. of Med., Miami, FL, [9]USUHS - TAGC, Bethesda, MD, [10]Univ Pennsylvania Sch Med, Philadelphia, PA, [11]UPENN, Philadelphia, PA, [12]Columbia Univ, New York, NY, [13]Columbia Univ., New York, NY, [14]John P. Hussman Inst. for Human Genomics, Miami, FL

Background: The ADSP is a National Institute on Aging (NIA) initiative focused on identifying genetic risk and protective variants for Alzheimer Disease (AD). Initial phases (Discovery and Discovery Extension) were predominantly non-Hispanic Whites of European Ancestry (NHW-EA). The ADSP expanded the population diversity in the Follow Up Study (ADSP-FUS) where existing ethnically diverse and unique cohorts underwent whole genome sequencing (WGS). The recent phase, ADSP-FUS 2.0: The Diverse Population Initiative, continues focusing on non-European populations, Hispanic/Latino (HL), non-Hispanic Black with African Ancestry (NHB-AA), and Asian populations.
Methods:ADSP cohorts consist of studies of AD, dementia, and age-related conditions. Clinical classifications are assigned based on standard criteria and derived from clinical measures and history, as well as additional neuropathologic data. In addition to production of WGS, APOE genotyping is available for all ADSP samples.
Results: The ADSP currently consists of 40 cohorts comprised of ~36,300 individuals, with plans to sequence >120,000 individuals from diverse ancestries. Genotyping, sequencing, and clinical adjudication has been performed on 36,361 participants (cases N=12,133, median age=72; cognitively-unimpaired(CU) individuals N=17,116, median age=74; ADRD N=7,112, median age=71). Mean ages for cases and controls vary across cohorts, from 57.0+5.6 to 86.5+4.2 for cases and 63.3+7.8 to 90.0+0 for controls. 61% of participants are female that are distributed as follows: cases (60.3%), CU(63.7%), and ADRD (55.8%). The most prevalent APOE genotype is APOE ?3/?3 (% by cases/CU for ?2/?2=0.2,0.5; ?2/?3=4.7, 9.4; ?2/?4=2.2, 1.8; ?3/3=41.2, 63.2; ?3/?4=40.9, 23.0; 4/4=10.8, 2.1). These proportions differ by ethnicities, with the highest for APOE ?4/?4 observed in Amerindian participants (18.3%) and the lowest in NHW-EA participants (1.7%).
Discussion: The results provide an overview of clinical features in ADSP cohorts. The continued growth of the ADSP-FUS 2.0 is central to the ADSP. Expanding in size and diversity, this genomic resource, available via NIAGADS, will be integrated with ADSP programs focused on phenotype harmonization, association analyses, functional genomics, and machine learning. In concert with these programs, the ADSP-FUS 2.0 will accelerate the identification and understanding of potential genetic risk and protective variants for AD across all populations with the target of developing new treatments that are globally effective.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session III

PB4554 The Association to Function Knowledge Portal enables forward- and reverse-genetics analyses of human genetic and functional genomic data

Authors:

T. Nguyen[1], M. Costanzo[2], M. Brandes[1], M. Duby[1], Q. Hoang[1], D. Jang[1], R. Koesterer[1], P. Kudtarkar[3], A. Moriondo[1], O. Ruebenacker[1], P. Smadbeck[1], S. Ying[3], K. Gaulton[4], M. Boehnke[5], J. Flannick[6], N. Burtt[7]; [1]The Broad Inst. of MIT and Harvard, Cambridge, MA, [2]Broad Inst. of MIT & Harvard, Cambridge, MA, [3]Univ. of California San Diego, San Diego, CA, [4]UCSD, San Diego, CA, [5]Univ. of Michigan, Ann Arbor, MI, [6]Boston Children s Hosp., Boston, MA, [7]Broad Inst., Cambridge, MA

Genome-wide association studies (GWAS) provides valuable insights into the genetic architecture of diseases; however, most functional effects of GWAS signals remain unknown, hindering the understanding and treatment of diseases. We have created the Association to Function Knowledge Portal (A2FKP; a2fkp.org) to address this obstacle on a large scale. The unique values of the A2FKP lie in the inclusion of 669 phenotypes in 11 disease systems across 11 ancestries, the integration of 8 data types (genetic and functional genomic) totaling more than 4,500 datasets, 8 bioinformatic methods that allow for integration and visualization of these data, and the capacity for both single-trait and multi-trait analysis. The A2FKP connects closely with a sister resource, Common Metabolic Diseases Genome Atlas (CMDGA; cmdga.org) that houses more than 7,200 epigenomic and functional genomic datasets. In addition, we have developed 8 community portals tailored to disease-specific research communities. The A2FKP includes tools for multi-trait analysis on both the variant level and the gene level. In a forward genetics approach, on the variant level, the Signal Sifter allows for clustering genetic associations by sets of traits, accounting for linkage disequilibrium, which can reveal novel pathways underlying complex traits. Users can search for variants associated with multiple traits, and filter the results based on direction of effect and strength of associations. Similarly, on the gene level, the Gene Finder enables assembling sets of genes with significant gene-level associations for a custom list of traits. It queries gene-level associations from both common and rare variant data, Human Genetic Evidence Calculator scores calculated by a novel statistical method that evaluates the strength of evidence for associations, tissue-specific gene expression levels, and effector gene predictions curated by experts. Starting with a list of traits, users can filter the results based on any of those criteria. In a reverse genetics approach, using the Variant Sifter, researchers can explore genomic regions of interest to prioritize potentially causal variants and genes. They can choose one or more phenotypes and opt to see relevant credible sets, both contributed by our collaborators and curated from the literature, and multiple types of epigenetic annotations for 48 tissue types, including accessible chromatin, chromatin states, binding sites, and target gene predictions. To build the A2FKP, we work directly with disease communities who provide their expertise, data and methods; we welcome future collaborations.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session I

PB4555 The Baylor College of Medicine Center for Precision Medicine Models

Authors:

D. Lanza, J. Rosenfield, S. Yamamoto, M. Wangler, H. Bellen, O. Kanca, J. Rogers, M. Roth, R. Darshoori, V. Ramanathan, U. Ramamurthy, J. Posey, S. Nagamani, M. Dickinson, J. Seavitt, L. Burrage, A. Milosavljevic, J. Heaney; Baylor Coll. of Med., Houston, TX

Precision animal models are critical tools for interpreting variants of uncertain significance and for pre-clinical investigation of personalized medicine approaches. To assist clinicians, researchers, patient support groups, and large research consortia (*e.g.*, Undiagnosed Diseases Network and GREGoR) with the development of precision animal models of undiagnosed, rare and Mendelian disorders, we have established the Baylor College of Medicine (BCM) Center for Precision Medicine Models (CPMM). Our center combines the expertise of fly, mouse, and nonhuman primate modeling programs with the clinical, gene discovery, and informatics expertise at BCM. Variants nominated for modeling through the Center website are reviewed for the feasibility and applicability of creating a precision fly or mouse model that can address the clinical question with consideration of existing model, variant, and clinical data. The Center also searches a database of genome and exome data from nonhuman primates in National Primate Research Centers to identify potential models with spontaneous mutations relevant to the nominated variant. To support the pathogenicity of a nominated variant the Center can query the Baylor Genetics clinical exome database for similar variants, perform RNA-seq using patient-derived cells, or reanalyze the patient exome or genome data. Nominations not accepted for modeling are accompanied by a clinical review summary including suggestions for other collaborations or existing modeling data that would support variant pathogenicity. If a nomination is accepted for modeling, Center members work closely with the nominator to develop a modeling plan that will address critical clinical questions. The Center leverages the resources and infrastructure of the Undiagnosed Diseases Network Model Organism Screening Center's Fly Core and Knockout Mouse Phenotyping Project (KOMP2) site at BCM to generate and characterize its precision fly and mouse models. To date, the Center has reviewed 70 nominations encompassing over 111 variants and accepted 18 variants for modeling in mouse, 19 variants for modeling in fly, and 2 variants for modeling in both organisms. The precision models currently being generated by CPMM are helping to uncover new disease mechanisms, reveal new disease biomarkers and phenotypes, and identify and test potential new therapeutic strategies for patients.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session II

PB4556 The Brain Gene Registry: A collaborative resource of the Intellectual and Developmental Disabilities Research Centers (IDDRC's) to promote advances in translational science.

Authors:

V. Lanzotti[1], A. Gupta[2], I. Y. Oh[2], D. Baldridge[1], M. Thompson[1], L. Kaster[2], A. Abbacchi[1], A. Lai[1], A. Sveden[3,4], S. Mahida[4], M. Chopra[3,4], S. Srivastava[3,4], H. Hazlett[5], E. R. Riggs[6], J. M. Savatt[6], J. Piven[5], M. Sahin[3,4], C. Gurnett[1], J. Constantino[7], P. R. O. Payne[2], Brain Gene Registry Consortium, B. Minor[2]; [1]Intellectual and Dev.al Disability Res. Ctr., Washington Univ. Sch. of Med., St. Louis, MO, [2]Inst. for Informatics, Data Sci. & Biostatistics Washington Univ. in St. Louis, St. Louis, MO, [3]Rosamund Stone Zander Translational NeuroSci. Ctr., Boston Children's Hosp., Boston, MA, [4]Dept. of Neurology, Boston Children's Hosp. Intellectual Disability and Res. Ctr.; Harvard Med. Sch., Boston, MA, [5]Dept. of Psychiatry, Univ. of North Carolina Intellectual and Dev.al Disability Res. Ctr., Chapel Hill, NC, [6]Autism & Dev.al Med. Inst., Geisinger, Danville, PA, [7]Ctr. for Behavioral and Mental Hlth., Children's Hlth.care of Atlanta, Dept.s of Psychiatry and Pediatrics, Emory Univ., Atlanta, GA

Monogenic disorders account for a significant proportion of population-attributable risk for intellectual and developmental disabilities (IDD), but the data necessary to establish a causal relationship between a given genetic variant and a neuropsychiatric disorder is often lacking. Recognizing this scientific roadblock, investigators from 13 Intellectual and Developmental Disabilities Research Centers (IDDRC's) formed a consortium, and using their collective resources, created a Brain Gene Registry (BGR); a repository which pairs genomic data from participants with variants in putative brain genes with high-quality neurophenotypic data. Phenotypic profiles are developed using a combination of data elements extracted from participants' electronic health records (EHR) and through a battery of standardized assessments, collectively referred to as the Rapid Neurophenotypic Assessment Protocol (RNAP). The RNAP was developed by this group and adapted for virtual administration; it comprises information on cognitive, neuropsychiatric, and sensorimotor function, including symptoms of dysmorphology, and neurodevelopmental conditions such as attention deficit hyperactivity disorder (ADHD) and autism spectrum disorder (ASD). BGR participants also co-enroll with GenomeConnect which enables the co-registered information to be assimilated in ClinVar. As of June 2023, the BGR contains data on 378 participants who are 55% male, 79% White, 7% Asian, 5% Black, and 11% Hispanic/Latine. The registry contains data for >80 genes with cohorts of more than 10 participants in CACNA1A, DNMT3A, SLC6A1, SETD5, and MYT1L, and 38 genes with cohorts between 3 and 10 participants. Most variants are *de novo;* 43% are classified as Pathogenic, 20% are Likely Pathogenic and 37% are Variants of Uncertain Significance, according to ACMG criteria. Cognitive standard scores on the Developmental Profile 4 (DP4) are shifted relative to a neurotypical population, with most participants scoring between 40 and 100 (*M= 67.69, SD=9.75*). EHR data show that symptoms of the muscle are the earliest and most prevalent diagnoses, followed over time by speech and language disorder, autism and ADHD. BGR-derived data has been used to accelerate and enrich curation for 15 brain genes evaluated by ClinGen's BGR ID-ASD Gene Curation Expert Panel. The BGR consortium welcomes referrals of participants with VUS or newly described variants in IDD genes and encourages the use of the evolving BGR data set to advance discoveries in translational science. Referrals, collaboration and data access requests can be submitted directly through the BGR website (https://braingeneregistry.wustl.edu/).

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session III

PB4557 The Ferrome: A curated list of iron-related genes with brain-specific sex differences in expression.

Authors:

K. Sharma[1], N. R. Wheeler[2], W. S. Bush[2,3], A. R. Kallianpur[1,4]; [1]Dept. of Genomic Med., Cleveland Clinic, Cleveland, OH, [2]Dept. of Population and Quantitative Hlth.Sci., Case Western Reserve Univ. Sch. of Med., Cleveland, OH, [3]Cleveland Inst. for Computational Biology, Case Western Reserve Univ., Cleveland, OH, [4]Dept. of Molecular Med., Cleveland Clinic Lerner Coll. of Med. of Case Western Reserve Univ., Cleveland, OH

Iron is essential for myelination, neurotransmitter synthesis, and energy production, and its dysregulation in the brain is likely to influence neurophysiology, cognition, and social behavior. Iron stores/utilization and brain iron transport differ by sex due to heightened demands during the female reproductive years. Furthermore, for unknown reasons, cognitive/neurodegenerative and neuropsychiatric disorders often show a sex predilection. This study assessed sex differences in brain expression of critical iron-related genes using a newly created resource. We manually curated an up-to-date, comprehensive database of 822 iron-regulatory and iron-metabolic genes and genes regulated by iron or that encode proteins dependent on iron for their function, which we collectively term the "ferrome". This database includes genes involved in iron-uptake/transport (336), iron-sulfur-cluster biogenesis and mitochondrial function (103), and essential oxidoreduction reactions (185); ferroptosis, a type of iron-mediated programmed cell death, involves at least 198 genes. Using a 100 KB window flanking each gene and the TOPMed genomic reference panel, ferrome genes were found to contain approximately 3 million common genetic variants. Based on data from the Genotype-Tissue Expression (GTEx) project, 715 ferrome genes (87%) are expressed in one or more brain regions. The liver, an organ central to systemic iron regulation, is also a major producer of iron-related transcripts and protein products, with 663 (81%) of genes expressed. Utilizing a recent analysis of sex-biased gene expression within GTEx, we performed gene-set enrichment analyses to determine whether ferrome genes are more or less likely to show differences in expression in the brain by sex. Of 20,013 genes expressed in the brain, ferrome genes were 1.21 times more likely to exhibit sex-biased gene expression ($p = 0.038$) than non-iron-related genes; 164 ferrome genes (20%) showed sex-biased expression in brain. No significant enrichment was observed, however, for sex-biased gene expression in the liver ($p = 0.304$). Top pathways enriched in these 164 sex-biased ferrome genes included: Metabolism (including biological oxidations and iron uptake/transport), Neurodegenerative Disorders, Ferroptosis, and Oxidative Phosphorylation. The ferrome and its genetic variants will be useful for targeted analyses of the role of iron-related genes in many neurological and neuropsychiatric phenotypes. A surprising proportion of these genes shows significant brain expression differences by sex, suggesting that such targeted analyses may lead to novel sex-specific associations in brain disorders.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session I

PB4558 The German Human Genome-Phenome Archive (GHGA) - A national infrastructure for secure archival and community-driven analysis of omics data

Authors:

C. Mertes[1,2,3], A. Benet-Pages[4,5], J. Winkelmann[4,3], J. Gagneur[1,2,3], on behalf of the GHGA Consortium; [1]Technical Univ. of Munich, Munich, Germany, [2]Munich Data Sci. Inst., Munich, Germany, [3]Inst. of Human Genetics, Sch. of Med., Technical Univ. of Munich, Munich, Germany, [4]Inst. of Neurogenomics, Helmholtz Ctr. Munich, Oberschleissheim, Germany, [5]Med. Genetics Ctr. Munich (MGZ), Munich, Germany

The German Human Genome-Phenome Archive (GHGA) is building a secure national omics data infrastructure, aiming to enable the secondary use of human omics data for research purposes. It is part of the German National Research Data Infrastructure (NFDI) and will serve as the national hub of the Federated European Genome-Phenome Archive (FEGA). GHGA strives to provide the necessary computing infrastructure, an ethical-legal framework to handle omics data in a data-protection-compliant and FAIR (Findable, Accessible, Interoperable, Reusable) manner, a harmonized metadata schema, and standardized workflows to uniformly process the incoming data. GHGA will be based on cloud computing infrastructures managed in a network of data generators. Utilizing the Global Alliance for Genomics and Health (GA4GH) standards, researchers will have controlled access to raw and processed sequence data.

We will showcase the first set of tools GHGA has developed to support our communities: a FEGA-compliant metadata model that links omics data with experimental and phenotypic information to make data traceable, and tools to help overcome challenges such as legacy consent forms, consent to share secondary data, and GDPR compliance. Furthermore, GHGA has co-developed workflows (together with the nf-core community) for data analysis, benchmarking, statistical analysis and data visualization.

Initially focusing on stakeholders that drive the national efforts for research and clinical sequencing at scale (cancer, rare and infectious diseases), GHGA will enable cross-project analysis and hence promote new collaborations and research.

This project is funded by the Deutsche Forschungsgemeinschaft (DFG) - project number 441914366 (NFDI 1/1).

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session II

PB4559 The Incontinentia Pigmenti Genetic Biobank: an example of fruitful partnership with patients

Authors:

M. Ursini[1], A. Pescatore[1], E. Spinosa[1], C. Casale[1], M. Lioi[2], F. Fusco[1]; [1]IGB ABT CNR, Naples, Italy, [2]Basilicata Univ., Potenza, Italy

The Incontinentia Pigmenti Gene Biobank (IPGB) is a facility that stores and manages biological samples from patients with the rare genetic disease Incontinentia Pigmenti for use in research. The IPGB research program collects blood, urine and saliva samples from participants. The IPGB research program seeks to construct a cohort studies to improve health of participants by characterizing the disease natural histories, identifying risk factors and revealing biomarkers. The biobank supports the collection, analysis, storage, and distribution of the biosamples that the program collects for research purposes. The IPGB program aims to make the research results accessible to participants, and it is developing new approaches to generate, access, and make data available to approved researchers.The IPGB is powered by participants that come from a group generally overlooked in biomedical research. Participants have been involved in defining the IPGB program since its inception. IPGB has built trust with patients and their families through continuous collaboration with patient associations (IPASSI ONLUS https://incontinentiapigmenti.it/). The association support and participate into the research and improve the quality of data collection.Participants are digitally registered through the IPGB website (https://www.igb.cnr.it/ipgb/). After completing the consent forms and registration, participants are provided with several basic health surveys and, where appropriate, the surveys can be enhanced to include additional questions as they have been developed in the event of the COVID pandemic.Consent and surveys are available in English, Italian, French and Spanish, and all the contents for the participants are provided in a simple and immediate language. IPGB has a high level of privacy and security standards developed in accordance with the BBMRI-ERIC consortium of which IPGB is a member. A particular effort has been put into transparency. The IPGB was created to allow participants to have access to their own data and most research test results. We are developing protocols for the return of genetic, laboratory tests, to participants; this development is guided by feedback from participants and internal and external experts.We plan to follow participants' health and outcomes for decades, to enable research that provides a better understanding of the disease, which in turn would support accurate molecular diagnoses, disease prevention strategies, treatment selection, and development of targeted therapiesThe IPGB research protocol has been published previously (Fusco et al. EJHG. 2019).

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session III

PB4560 The Mouse Peroxisome Research Resource for Basic and Preclinical Research

Authors:

J. Hacia[1], L. l. Arayan[2], N. Braverman[3], A. Zuberi[2]; [1]Univ of Southern California, Los Angeles, CA, [2]The Jackson Lab., Bar Harbor, ME, [3]MUHC-RI, Montreal, QC, Canada

Peroxisomes are a hub of cellular metabolism and signaling pathways and play essential roles in the development and functions of mammalian organ systems. Inherited defects in genes involved in peroxisome assembly, structure, replication, and functions are responsible for a group of multisystemic monogenic disorders that include X-linked adrenoleukodystrophy and Zellweger spectrum disorder. Based on genome-wide association studies (GWAS) and lipidomic data, peroxisome activities have been proposed to contribute to the pathophysiology of common disorders including diabetes, cancer, Alzheimer's disease, and infectious diseases. Despite their relevance to human health, the limited number of publicly available mouse models and antibody resources to investigate peroxisome biology and test targeted therapies for peroxisomal disorders has hindered progress within the translational science spectrum. Here, we discuss the NIH-funded Mouse Peroxisome Research Resource (MPRR) that provides a central resource for mouse models and antibodies for peroxisome research. The MPRR works with the peroxisome basic and preclinical research communities and patient advocacy groups to accelerate the creation, distribution, and proper use of high-impact mouse models and immunological reagents. It solicits for and ensures that all deposited mouse models are placed on standardized genetic backgrounds to control for the presence of genetic modifiers that influence phenotypes of interest, such as disease severity. It also produces novel high-priority mouse models with defined genotypes on standardized genetic backgrounds, cryopreserves them, and distributes them to the public. The MPRR assists in the targeted phenotyping of these mouse models, including the measurement of relevant peroxisome metabolite levels in tissues of interest. Moreover, it produces, publicizes, and distributes validated antibody reagents for peroxisome research, including characterizing relevant mouse models. Overall, the MPRR seeks to develop and distribute well-annotated resources to accelerate basic science research and the preclinical testing of therapeutic interventions for disorders caused by peroxisome dysfunction.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session I

PB4561 The NHGRI Sample Repository for Human Genetic Research: biospecimens and a new multi-variant search tool for PGx and PRS

Authors:

A. Liautaud, D. Kusic, G. Calendo, J. Madzo, S. Sampson, J. Schulz, L. Willis, E. Horton, M. Mitchell, N. Gharani, L. Scheinfeldt; Coriell Inst. for Med. Res., Camden, NJ

The NHGRI Sample Repository for Human Genetic Research (NHGRI Repository) housed at the Coriell Institute for Medical Research facilitates studies of human genetic and genomic variation by establishing, characterizing and distributing a large (over 3,700) and diverse publicly available collection of renewable biospecimens obtained from human populations living around the world, including biospecimens associated with the HapMap and 1000 Genomes Projects as well as the Human Pangenome Reference Consortium. Participants that have generously donated to the NHGRI Repository consented to their biospecimens and associated data being used for a wide range of general research and to broad data sharing of largescale genomic data. Through the 1000 Genomes Project, the majority of this collection has been characterized with publicly available whole genome sequencing data and other large-scale genomic data. We recently developed a user-friendly and integrated multi-SNP search tool that allows catalog users to dynamically query 1000 Genomes Project whole genome sequencing data (30x). A set of up to 100 variants at a time can be searched by individual variant (rsid). The multi-SNP search returns an interactive table of each subject id, population descriptor, genotype for each SNP included in the search, and sex that can be filtered or exported to an excel file. This tool is intended to support researchers interested in identifying biospecimens with particular multi-variant profiles such as those used in pharmacogenetics (PGx) and polygenic risk assessments. For example, a researcher interested in identifying cells or DNA that carry one or two copies of the CYP2C19*2 allele can simultaneously search the three haplotype defining SNPs (rs12769205, rs4244285, rs3758581) according to PharmVar. Similarly, researchers interested in identifying cells or DNA that carry high or low numbers of polygenic risk alleles can search up to 100 variants at a time; for example, 23 of the 34 variants that define the ROOT PRS model for breast cancer are included in the WGS dataset and can be searched together in a single batch. This tool adds to the growing number of web-based genomic search tools affiliated with the NHGRI Repository, including a Gene search tool and a Gene Expression search tool. These tools are intended to enable researchers a fast and simple way to identify biospecimens with the most appropriate variation profiles for their research goals. Our Genomic Data Search tools can be accessed at https://www.coriell.org/1/Browse/Genomic-Data-Search, and more information about the NHGRI Repository can be found at https://catalog.coriell.org/NHGRI.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session II

PB4562 The NIH Comparative Genomics Resource: Maximizing the impact of research organisms to human health.

Authors:

V. Schneider[1], N. A. O'Leary[1], S. H. Rangwala[2], F. Thibaud-Nissen[1], T. L. Madden[1], T. D. Murphy[1]; [1]NIH/NLM/NCBI, Bethesda, MD, [2]NCBI/NLM/NIH, Bethesda, MD

NCBI is developing the NIH Comparative Genomics Resource (CGR) to maximize the impact of eukaryotic research organisms and their genomic data to biomedical research. Comparative genomics involves the evaluation of genomic features from diverse organisms to systematically explore and evaluate biological relationships. Because all biological processes rely on genes, proteins, and the biochemical pathways they belong to, and these fundamental elements are shared across the tree of life, this approach can inform nearly any aspect of biology- and help identify new model organisms that may be used to address issues in human health. However, the relevant data and applications for the growing number of sequenced organisms are often siloed, challenging their use. Additionally, assembled genomes often contain contaminating sequences and lack annotations that inform comparisons. CGR facilitates reliable comparative genomics analyses for all eukaryotic organisms through community collaboration and an NCBI genomics toolkit. The toolkit offers high-quality genomics-related resources impactful to users' research, including interconnected databases with access points enabling seamless navigation of NCBI content and interoperable data and tools that can integrate into users' workflows. We will present our recent developments on several of these resources, including a new public cloud-ready tool suite for removing foreign sequence contamination, a new web-based graphical tool supporting the visual comparison of two genomes based on alignments, and NCBI Datasets, which provides web and programmatic interfaces that help users search, browse, and download genome-related data. We will illustrate the value of CGR in the context of health-related workflows. We will also demonstrate how connecting community-supplied genome-related content to the organism-agnostic CGR can make it easier for researchers to use data from sources and organisms they may not have known about, and how community feedback has impacted CGR development. Through CGR, we can assist researchers in using genomic data available from thousands of organisms to answer questions involving human health.
This work was supported by the National Center for Biotechnology Information (NCBI) of the National Library of Medicine (NLM), National Institutes of Health (NIH).

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session III

PB4563 The NINDS Human Genetics Resource Center: publicly available resource of large and clinically well-characterized cells and DNA

Authors:

L. Scheinfeldt[1], A. Green[2], A. Amberson[2], S. Sampson[2], T. Cutts[2], R. Zhang[3]; [1]Coriell Inst. for Med. Res., Philadelphia, NJ, [2]Coriell Inst. for Med. Res., Camden, NJ, [3]NIH/NINDS, Rockville, MD

Established in 2002, the National Institute of Neurological Disorders and Stroke (NINDS) Human Genetics Resource Center is a public biorepository of DNA and lymphoblastoid cell lines available to promote neurogenetics research. Specimens donated by over 42,000 individuals can be accessed through an online catalog (http://catalog.coriell.org/NINDS), and over 70,000 catalog samples have been distributed to researchers around the world. The collection includes samples from individuals diagnosed with cerebrovascular diseases (N>12,800), Parkinsonism (N>5,600), motor neuron diseases (N>2,500), Epilepsy (N>6,100), Tourette syndrome (N>4,200), Dystonia (N>3,800), and neurologically normal controls (N>7,500). Each patient biospecimen is annotated with standardized clinical and demographic common data elements (CDEs) specifically designed for each disease subcollection (https://catalog.coriell.org/1/NINDS/Downloads2/Clinical-Data-Forms). This resource thereby facilitates large, multi-site studies of genetic risk factors with consistent clinical data that would not otherwise be possible. Indeed, over 900 publications have used samples and data from the NINDS Human Genetics Resource Center, and these publications have been cited over 90,000 times. These publications include genome-wide association studies of Parkinson's disease, ALS, Epilepsy, Stroke, Tourette Syndrome, and Dystonia. This collection, therefore, has made a substantial impact on research around the world.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session I

PB4564 The PhenX Toolkit: Recommended measurement protocols to standardize data collection for COVID-19 studies

Authors:

D. Williams, M. C. Krzyzanowski, J. Abedini, P. Ceger, B. Cook, I. Glaze, W. Huggins, C. Ives, M. Mandal, H. Pan, T. Reeve, G. Swank, P. West, L. G. Wood, X. Wu, T. Hendershot, C. M. Hamilton; RTI Intl., Research Triangle Park, NC

The PhenX (Phenotypes and eXposures) Toolkit is a web-based catalog of recommended measurement protocols to assist with the design of studies with human participants and to facilitate cross-study data integration and analyses. PhenX protocols are recommended by domain experts using a consensus-based process that includes community input. The PhenX Toolkit currently includes measurement protocols across 30 research domains (e.g., Demographics) that add breadth to the Toolkit and six collections (e.g., Social Determinants of Health) that add depth to the Toolkit. The Toolkit site has over 4,700 registered users and has been recommended for use as a NIH Common Data Elements (CDEs) resource in over 600 NIH Funding Opportunity Announcements (FOAs).
In 2020, to assist NIH with their response to the COVID-19 pandemic, the PhenX Toolkit began releasing protocols into the COVID-19 Protocol Library to facilitate sharing of COVID-19 surveys. The goals were to reduce the proliferation of additional COVID-19 related surveys and to identify protocols that could be recommended for use and release into the PhenX Toolkit. Initially, 20 protocols from the COVID-19 Library were included in six COVID-19 Specialty Collections. The collections were decided based on the results of a crowd-sourcing effort and guidance from the PhenX Steering Committee, enabling "rapid release" of these protocols into the Toolkit. Use of these protocols helps investigators understand the symptoms and treatments of COVID-19 in the context of social determinants of health, the impact of the pandemic on mental health and behavior, and the effects of infection on health and daily living. Later, the recognized need for protocols describing the emerging longer-term effects of the COVID-19 pandemic led to the development of a Long COVID collection. The Long COVID Collection contains 25 protocols that address Long COVID in the context of an individual’s quality of life, daily living, comorbidities and COVID impact, multisystem inflammatory syndrome, physical functioning, vaccination, and symptoms.
The COVID-19 Variable Compare Tool (VCT) enables investigators to compare COVID-19 Specialty protocols and COVID-19 Library instruments. The VCT allows investigators to explore survey questionnaires and questions within the questionnaires, which are referred to as "variables". Through the interface, an investigator may perform a keyword search of variables, view a side-by-side comparison of questionnaires, or generate a heatmap visualization that shows the variable-based similarities between multiple questionnaires.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session II

PB4565 The POPGEN project: building a French reference panel of genomes

Authors:

A. Herzig[1], G. Marenne[1], M. Guivarch[1], A. Saint Pierre[1], T. Ludwig[1,2], I. Alves[3], D. Christian[3], R. Redon[3], J-M. Sebaoun[4], L. Gressin[4], J-C. Beaudoin[4], V. Morel[4], B. Fin[5], C. Besse[5], R. Olaso[5], D. Bacq[5], V. Meyer[5], F. Sandron[5], A. Ferrane[6], A. Boland[5], H. Blanché[4], M. Zins[7], J-F. Deleuze[4,5], G. Le Folgoc[1], E. Génin[1,2]; [1]Univ. Brest, Inserm, EFS, UMR 1078, GGB, IBSAM, F-29200, Brest, France, [2]CHU Brest, F-29200, Brest, France, [3]Nantes Université, CHU Nantes, CNRS, INSERM, l'institut du thorax, F-44000, Nantes, France, [4]Fondation Jean Dausset, CEPH, Paris, France, [5]Université Paris-Saclay, CEA, Ctr. Natl. de Recherche en Génomique Humaine (CNRGH), 91057, Evry, France, [6]Inst. de Santé Publique, Pôle de recherche clinique, INSERM, F-75013, Paris, France, [7]Université de Paris Cité, Université Paris-Saclay, UVSQ, Inserm, Cohortes Epidémiologiques en Population, UMS 11, F-94807, Villejuif, France

Background/Objectives: The POPGEN project was launched as part of the French genomic medical initiative to build a catalogue of variants found in the different regions of metropolitan France and provide allele frequencies in order to help filter out neutral variants from patient genomes.
Methods: Individuals from the population cohort Constances were asked to complete a questionnaire on birthplaces and birth years of their parents and grandparents. Based on their answers, 15,000 individuals were selected to cover the different regions of metropolitan France and were posted saliva collection kits. Genotyping was successful for 9,772 individuals and 4,000 individuals were selected for whole-genome sequencing. Different methods were used to study fine-scale population structure and rare variants were imputed using public reference panels enriched by 856 whole genomes from the FranceGenRef project.
Results: We demonstrate the fine-scale population structure of French populations and show how it relates to geography. Using these results, we show how the performance of imputation panels can vary across the territory; driven by patterns in haplotype sharing. We also investigate the important impact for downstream genetic epidemiological study designs.
Conclusions: This study proposes a design to sample individuals from the general population to create reference panels that could help improve imputation accuracy for geographically clustered variants. The POPGEN project will contribute to the "Genome of Europe" project.
Grant: French Ministry of Research PFMG2025 and ANR IA-10-LABX-0013 FranceGenRef

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session III

PB4566 † The World's Largest Methylation Data Cohort: The Million Veteran Program increases its multi-omics capability by profiling 45,460 on a methylation array to provide novel insights and to serve as a rich data resource.

Authors:

K. Markianos[1], P. Schreiner[1], B. Despard[1], S. Gautam[1], B. Barnes[2], W. Zhou[3], B. Gorman[1], P. Devineni[1], Y. Shi[4], P. Tsao[5], S. Pyarajan[1]; [1]VA Boston, Boston, MA, [2]Illumina, Inc., San Diego, CA, [3]Ctr. for Computational and Genomic Med., The Children's Hosp. of Philadelphia, Dept. of Pathology and Lab. Med., Univ. of Pennsylvania, Philadelphia, PA, [4]VA Boston Hlth.care System, Jamaica Plain, MA, [5]VAPAHCS/Stanford Univ., Palo Alto, CA

The Department of Veterans Affairs (VA) Million Veteran Program (MVP) is one of the world's largest databases of genomic and clinical data. MVP has already proven to be a valuable precision medicine research resource. It has the potential to develop into a multi-omic resource with significant implications for veteran health research. To this end, we introduce the new MVP methylation dataset. We describe quality control and analysis of 45,460 methylation samples released by the Million Veteran Program using the Illumina Infinium Methylation EPIC Array (~850k genomic sites). We developed a Dockerized, high-throughput SeSAMe-based quality control and preprocessing pipeline incorporating common preprocessing steps that calculate metrics describing the reliability of methylation detection, which are necessary to validate probes and define sample inclusion. In addition, we developed a high-precision fingerprinting process that allows unambiguous matching of methylation assays to genotype and sequencing assays from the same individuals. We present demographic distributions of the data set (ancestry, gender) and we perform exemplary analysis: HIV+ age acceleration and Body Mass Index (BMI) EWAS. We compare our analysis with published literature and comment on the sensitivity of the results to technical covariates such as DNA storage length. This data release represents the largest methylation data set to date. Our results illustrate the potential of this dataset for elucidating the role of epigenetic variation in human disease, particularly when combined with MVP's large genomic and clinical datasets.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session I

PB4567 UCSC Genome Browser: Instantly lift annotations between human genomes.

Authors:

L. Nassar, B. J. Raney, G. P. Barber, J. Casper, H. Clawson, C. Fischer, J. Navarro Gonzalez, A. S. Hinrichs, C. M. Lee, G. Perez, M. Haeussler, W. J. Kent; UC Santa Cruz, Genomics Inst., Santa Cruz, CA

With the rapid increase in sequencing accessibility, combined with consortium projects such as T2T and HPRC, we are seeing an exponential trend in the number of human genomes not only becoming available but actively researched. Unfortunately, many databases are singularly mapped to the reference such as GRCh38/hg38, leaving these new assemblies with sparse annotations.

Here we present liftOver-on-the-fly, a process by which annotations from one assembly can be quickly displayed on another. To do this we leverage our annotation lifting pipeline, which we have operated for over a decade, to lift any desired annotations from one assembly, such as hg38, to any number of target genomes including T2T and HPRC assemblies. The liftOver tool works best when lifting between closely related assemblies, so human-to-human lifts are often 100% matches.

LiftOver-on-the-fly is still in active development, and the types of data that can be lifted, as well as the origin and target assemblies, are expanding. We anticipate that the demand for features such as this will continue to increase and annotation transfers such as these will eventually become commonplace.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session II

PB4568 Understanding the role of structural variants in Alzheimer's Disease: Building a reference imputation panel across 4,699 samples to drive SV-GWAS

Authors:

M. Byrska-Bishop[1], A. Corvelo[1], A. O. Basile[1], W. E. Clarke[1,2], U. S. Evani[1], R. Musunuri[1], A. Runnels[1], B. Vardarajan[3,4], G. Narzisi[1], M. C. Zody[1]; [1]New York Genome Ctr., New York, NY, [2]Outlier Informatics Inc., Saskatoon, SK, Canada, [3]Taub Inst. for Res. on Alzheimer's Disease and the Aging Brain, Coll. of Physicians and Surgeons, Columbia Univ., New York, NY, [4]Dept. of Neurology, Coll. of Physicians and Surgeons, Columbia Univ. and the New York Presbyterian Hosp., New York, NY

The role of structural variants (SVs) in Alzheimer's disease (AD) is not well understood. While the majority of AD research has focused on the contribution of single nucleotide variants (SNVs) and short insertions and deletions (INDELs), SVs have received comparatively less attention. SVs refer to larger-size (>50 bp) alterations in the genome that have the potential to impact gene expression and disrupt regulatory elements, thus likely contributing to disease state. Understanding the role of SVs in AD could potentially shed light on novel mechanisms and therapeutic targets for this devastating neurological disorder. Here, we describe variant calling and generation of a reference imputation panel consisting of SNVs, INDELs, and SVs (large insertions and deletions) discovered across 4,669 Illumina WGS samples from 700 multiplex AD families (EFIGA and NIA LOAD) and unrelated case-control samples (ROS/MAP). We called and genotyped 105,874,088 SNVs and INDELs using GATK HaplotypeCaller. In parallel, we produced a call set containing 123,094 SVs by combining: 1) Illumina calls discovered using Manta, MELT, and Absinthe across 4,669 AD samples; 2) Oxford Nanopore (ONT) calls discovered using Sniffles2 across 4 AD samples; and 3) genotyping on the AD samples of an external PacBio call set produced by the Human Genome Structural Variation Consortium from 32 diverse human genomes. All SVs were genotyped at the cohort-level using either Paragraph or Pangenie. To make the variants accessible for future association studies, we generated an integrated haplotype-resolved reference imputation panel. First we filtered the SNV/INDEL and SV call sets, selecting non-singleton sites in Hardy-Weinberg equilibrium with low missingness rate which resulted in 64,386,896 small variants and 82,594 SVs. Next, we phased the SNV/INDEL call set using statistical phasing, as implemented in the SHAPEIT4 software. We then used the phased SNV/INDEL panel as a scaffold onto which we phased the SVs. We computed the switch error rate (SER) of SNV/INDELs using ONT-based phasing information, available for 4 of the AD samples. The SER ranged from 0.11% to 0.52%, indicating high phasing accuracy. Similarly, to evaluate the accuracy of SV phasing, we computed their flip rate relative to flanking SNVs, against the ONT-based phased call sets. The flip rate of SVs ranged from 0.48% to 0.88%, suggesting high accuracy of SV phasing as well. Finally, we show examples demonstrating how the reference panel including SVs can be used to accurately impute genetic variants across the size spectrum and enable discovery of new associations improving our understanding of the genetic architecture of AD.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session III

PB4569 Unveiling hidden transcripts using a T2T human reference genome

Authors:

J. Kim[1], E. Tseng[2], A. Phillippy[3], A. Rhie[4]; [1]Seoul Natl. Univ., Seoul, Korea, Republic of, [2]Pacific BioSci.s, Menlo Park, CA, [3]NIH/NHGRI, Bethesda, MD, [4]NIH, Bethesda, MD

A complete human reference genome (T2T-CHM13v2, hereby CHM13) was recently released, including an additional 200 million base pairs of sequence. It corrects thousands of structural errors and unlocks the most complex regions of the human genome. With the inclusion of numerous genes not present in the widely used GRCh38 reference, there is a chance to discover novel transcripts that have been excluded from prior analyses.In this study, we utilized long-read RNA sequencing data derived from the HG002 cell line as a biological replicate for the identification and annotation of full-length isoforms in the CHM13 reference. To evaluate the applicability of CHM13 as a reference for transcriptome data, we compared mapping statistics between the CHM13 and GRCh38 references. We found 4,900 transcriptome reads that mapped to CHM13 but not GRCh38. These reads mapped to 62 regions on CHM13, among which 19 loci (30.6%) were found to be in non-syntenic regions with respect to GRCh38. The remaining transcripts mapped to sequences that exhibited better concordance to CHM13, which could be attributed to corrections made in CHM13 or better agreement between the HG002 and CHM13 haplotypes. Our results also showed that over 200,000 reads were mappable to CHM13 with a higher mapping quality compared to GRCh38. In contrast, only 900 reads were unmappable to CHM13 but alignable to GRCh38, encompassing 8 non-overlapping gene loci in total. Predominantly, most of these transcripts were found in the intronic regions of 4 genes and one exonic region of the hypervariable HLA-DQA1 gene.We conclude the more complete CHM13 reference improves overall transcript mappability, which is essential for accurate isoform identification and discovering potentially novel transcripts. Future work will include comparisons to a personalized, diploid HG002 reference as well as a pangenomic reference.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session I

PB4570 Unveiling the spectrum of genetic variants intheKorean population: whole genome sequencing of 8,100 Korean participants

Authors:

Y. Kim, N-H. Choi, M. Hwang, D. Shin, K. Yoon, H-Y. Park, B-J. Kim; Korea NIH, Cheongju-si, Korea, Republic of

Recent extensive sequencing projects have catalogued an immense number of genetic variants, offering potential targets for drug development and valuable resources for precision medicine research. However, the majority of the datasets consisted of individuals of European ancestry, highlighting the need for additional sequencing datasets encompassing diverse ethnicities. The Korean Reference Genome (KRG) project was initiated by the National Institute of Health, Republic of Korea, with the aim of identifying variants in the whole-genome sequences of approximately 20,000 Korean participants. During the interim phase of the Korean Reference Genome (KRG) project, approximately 8,100 Korean participants were subjected to sequencing with an average depth of 30x. The obtained reads were aligned to the hg19 reference genome, and variant calling was performed using DRAGEN™. The resulting 8.1K KRG dataset comprised around 46 million variants, with approximately 84.6% being single nucleotide polymorphisms (SNPs) and the remaining 15.4% being insertions and deletions (indels). Among these variants, 22.9% were classified as common (with a minor allele frequency, MAF, of 1% or higher), while 77.1% were categorized as rare (with an MAF below 1%). Within the 791K protein coding altering variants identified, the majority were rare, while 10.6% were common. The KRG project expanded the repertoire of predicted loss-of-function (pLOF), encompassing around 36K pLOF variants. Additionally, 2,276 variants were annotated as pathogenic or likely pathogenic based on information from the ClinVar or LOVD databases. The KRG significantly improved the accuracy of imputation for the Korean population. When imputing independent 499 Korean samples using different reference panels, including the KRG, the KRG demonstrated the highest level of accuracy with a mean imputation accuracy of r^2=0.996. Furthermore, the KRG offered extensive genomic coverage for common variants (MAF >= 1% & r^2>=0.8, 84.5%) and rare variants (MAF 0.1~1% & r^2>=0.8, 51.5%). This study represents a valuable genetic resource for the East Asian population along with a comprehensive imputation server for imputation analysis.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session II

PB4571 Utilizing targeted TSO500 ctDNA NGS assay and Seer Proteograph to study genome and proteome changes in various cancer types

Authors:

N. Prasad; Discovery Life Sci., Huntsville, AL

Liquid Biopsy in recent year have emerged as a valuable sample to understand the comprehensive genome and proteome in oncology patients. It provides a unique opportunity to diagnose, monitor and evaluate response to therapy using a non-invasive blood draw. New emerging Next Generation Sequencing (NGS) and proteomic technologies allow researchers to investigate key changes within the genome and proteome in these patients to enable patient centric care and treatment. We have utilized the TruSight Oncology ctDNA (TSO 500ct DNA) assay paired with Seer Proteograph, an automated liquid handling and nanoparticle-based enrichment technology assay along with SCIEX's Zeno TOF 7600 Mass Spec, to enable deeper genomic and proteomics profiling from plasma samples of oncology patients. This workflow was used to generate deep, unbiased genomics and proteomic data from a plasma sample sets from donors with various cancer types and stages. The NGS and proteomic data from these sample sets were then correlated to underscore the importance of using a multiomic approach to fully understand the underlying drivers of disease in oncology patients.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session III

PB4572 Validating gene-disease relationships in Charcot-Marie-Tooth disease through a robust curation effort: The Clinical Genome Resource Charcot-Marie-Tooth gene and variant curation expert panels.

Authors:

K. Latchman[1], J. Raposo[1], D. Bis-Brewer[2], A. Rebelo[3], K. Johnson[4], J. Medina[5], M. Dohrn[6], D. Beijer[7], L. Van de Vondel[8], G. Spurdens[1], I. Xu[7], R. Le Shay Araujo[1], H. Elloumi[9], R. Seyedsadjadi[10], A. Cortese[11], J. De Winter[12], D. Hermann[13], S. Mohan[14], T. Stojkovic[15], S. Scherer[16], S. Zuchner[5]; [1]Univ. of Miami Dept. of Human Genetics, Miami, FL, [2]Geisinger, Danville, PA, [3]Univ Miami, Miami, FL, [4]The Perelman Sch. of Med. at the Univ. of Pennsylvania, Philadelphia, PA, [5]Univ. of Miami Miller Sch. of Med., Miami, FL, [6]RWTH Aachen Univ., Aachen, Germany, [7]Univ. of Miami, Miami, FL, [8]Univ. of Antwerp, Antwerpen, Belgium, [9]GeneDx, Rockville, MD, [10]Harvard, Cambridge, MA, [11]Univ. Coll. London, London, United Kingdom, [12]Univ. of Antwerp and Antwerp Univ. Hosp., Brussels, Belgium, [13]Univ. of Rochester, Rochester, NY, [14]Univ. of North Carolina At Chapel Hill, Apex, NC, [15]Inst. of Myology, Paris, France, [16]Univ. of Pennsylvania, Philadephia, PA

The Clinical Genome Resource's Charcot-Marie-Tooth Gene Curation Expert Panel (ClinGen CMT GCEP) was established in December 2019 to evaluate the proposed gene-disease relationships of more than 90 genes reported to cause CMT. CMT is the most common inherited neurological disorder, affecting approximately 1 in 2500 individuals and the goal of the CMT GCEP is to use the ClinGen gene-disease clinical validity framework to evaluate the strength of evidence for these gene-disease relationships. ClinGen's curation process is uniquely recognized by the FDA and thus will contribute to regulatory considerations of gene-based diagnosis and therapies. To date, the CMT GCEP has completed an evidence-based classification of 55 genes and more than 40 genes have been classified as having definitive evidence for robust gene-disease relationships based on current literature. With more than 50% of reported CMT genes having been curated, we recently gained approval to begin a ClinGen CMT Variant Curation Expert Panel (CMT VCEP). The process of gene and variant curation is in accordance with ClinGen's gene-disease clinical validity framework whereby biocurators within the CMT GCEP and VCEP review the literature, score genetic and experimental evidence to support or refute a claim of gene-disease relationship, and make a final decision among an expert panel regarding the determining evidence category, such as definitive, moderate, or limited. Evidence supporting the gene-disease relationship include case-level data, co-segregation analyses, and functional experiments. A standardized pre-curation process (lumping or splitting criteria) to define the disease entity according to MonDO (Monarch Initiative) is followed when more than one phenotype is indicated for a gene. Pre-curations and curations are thoroughly discussed by the expert panels and final classifications, based on cumulative score and replication over time, are published on the ClinGen website. The CMT GCEP and VCEP are an internationally collaborative effort, comprised of recognized experts in Charcot-Marie-Tooth disease, and whose primary goals are to analyze and define the clinical relevance of known CMT genes, improving our knowledge of genomic variations, and providing valuable resources to further guide precision medicine and research.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session I

PB4573 VanDyPlot, an interactive gene-specific variant visualization interface.

Authors:

M. Marhabaie[1], S. Ramadesikan[1], I. Showpnil[1], U. Gurusamy[1], A. C. Herman[1], K. Voytovich[1], D. M. Gordon[1], H. C. Kuck[1], G. E. Lammi[1], R. K. Wilson[1,2], D. C. Koboldt[1,2]; [1]The Steve and Cindy Rasmussen Inst. for Genomic Med. at Nationwide Children's Hosp., Columbus, OH, [2]Dept. of Pediatrics, The Ohio State Univ. Coll. of Med., Columbus, OH

The advent of high-throughput sequencing and exponential growth of genetic testing has dramatically increased the catalog of human genetic variation in disease genes and led to the discovery of novel disease-causing genes. Variant interpretation, particularly for novel variants, remains a major challenge in both clinical and research settings. Understanding patterns of known disease-associated variants can play an important role in the pathogenicity classification of the variant. Thus, there is an increasing need for a robust user-friendly interface that can be used to visualize large datasets. Unfortunately, currently available tools are either not user-friendly or heavily geared toward cancer research, thus, do not have an exhaustive representation of germline genetic variants in genes associated with constitutional disease conditions. To fill this gap, we developed VanDyPlot, a React-based web application that utilizes Amazon Web Services to visualize variant information across genes. It features (1) a lightweight application that (2) encompasses all RefSeq transcripts (with MANE transcript indicator) and (3) can handle multiple tracks from diverse databases with (4) protein domain annotations and (5) various preset variant filters. The current release (accessible at https://vandyplot.igmdev.org) focuses on variants reported in the ClinVar database. As of May 2023, VanDyPlot contains information for more than 2,178,000 variants (317,827 pathogenic/likely pathogenic, 828,893 benign/likely benign, and 1,016,514 variants of uncertain significance) from more than 22,740 genes. Functional domains, motifs, and active sites of proteins are also annotated on the plot using the Pfam database and the UniProt knowledgebase. By visualizing genomic variation data in a compact and accessible way, VanDyPlot offers a comprehensive view of variation across a gene and helps to consolidate and understand patterns of disease-associated variants.

Session Title: Genetic, Genomic, and Epigenomic Resources and Databases Poster Session II

PB4574 We do know what we don't know: variant assessment across *All of US* long read efforts

Authors:

F. Sedlazeck[1,2], Y. Zhu[1], M. Izydorczyk[1], M. Mahmoud[3], L. Paulin[1], Z. Khan[3], G. Metcalf[3], D. Muzny[1], R. Gibbs[1], All of Us Research Program consortium; [1]Baylor Coll. Med., Houston, TX, [2]Rice Univ., Houston, TX, [3]Baylor Coll. of Med., Houston, TX

The All of Us Research Program seeks to generate whole genome data on at least 1M diverse participants in the US with clinical return of results. While this resource will provide the evidence needed to make clinical sequencing more applicable to non-Caucasian populations, it does not address the full spectrum of variation. The precision enabled by short reads does not extend to larger classes of variants (e.g. SVs, tandem repeats etc). These variants are documented to have a role in disease and impact more nucleotides than SNVs so additional methods to resolve are required.

All of Us recently launched an initiative to produce ~15,000 long read genomes using PacBio and Oxford Nanopore. This endeavor will enhance our understanding of the full spectrum of variation across understudied populations, and serve as a valuable annotation resource for many subsequent studies involving medically relevant and challenging genes (e.g *LPA*). While long read data does enable full variant characterization, using these data for population analyses requires new computational methods.

In response, we have developed tools to identify all classes of variants and enable their comprehensive representation. For SV calling, Sniffles outperforms other methods in speed and accuracy and can produce a fully genotyped population variant file that can be directly utilized in SV association studies. Further, Sniffles enables identification of mosaic SVs in bulk sequencing. Here, we found multiple candidates that we could already validate.

To improve our understanding for repeats in the genome, we have co-developed TRGT. TRGT can call complex alleles (e.g. in *FMR1* or *RFC1)* that are often inaccessible by other methods and known to be pathogenic. TRGT reports length designation with sequence resolved alleles, which are important for interpretation (e.g. fragile X syndrome). Further, we extended Truvari to operate in tandem repeat regions where the allelic representation is challenging. Truvari allows us to merge variant types and accurately compare them across samples.

In addition to resolving complex alleles, long read data creates methylation signals, which we leverage to improve phasing by up to 3-fold in length using MethPhaser. Phasing information is essential to estimate gene dosage and expression consequences.

These innovations will help to identify new gene targets for diseases in *All of Us* and other consortia. In my presentation, I will highlight multiple novel findings enabled by these methodologies and highlight their likely functional and phenotypic consequences.

Session Title: Mendelian Phenotypes Poster Session I

PB4575 17 years old Puerto Rican male with a Laminopathy disorder

Authors:

L. Bou-Padilla[1], S. Carlo[2], E. Albino[3], H. Roman[4], A. S. Cornier[5]; [1]Ponce Hlth.Sci. Univ., Ponce, PR, [2]Ponce Hlth.Sci. Univ., Ponce, PR, [3]Univ. of Puerto Rico - MSC, San Juan, PR, [4]Genetic section - San Jorge Children's Hosp., San Juan, PR, [5]Genetic Section - San Jorge Children s & Women s Hosp., San Juan, PR

Case of a 17-year-old Puerto Rican male was referred to Genetics at the age of 3 ½ with a CPK test of 1,200 mcg/L suggestive of muscular dystrophy. The patient was born to a 27-year-old (G1, P0, C1), and by the age of 5-6 months, motor delay was detected. DNA testing performed at the time was not informative, and a muscle biopsy was obtained, resulting in an unidentified type of muscle dystrophy. Further genetic testing of the LMNA gene demonstrated a de novo dominant negative mutation in the LMNA gene (c.125T>C; p.L42S). A mutation in LMNA results in early onset of Congenital Muscular Dystrophy (CMD) with axial weakness, spinal rigidity, and cardiac and pulmonary involvement, which are the most critical aspects of the condition. Over the years, he had irregular visits to the cardiologist, who recommended atenolol without specific cardiac involvement. At the age of 5 years, he developed respiratory failure for which he required a tracheostomy and assisted mechanical ventilation. At the age of 7 years, left ventricular hypertrophy was noted and has remained stable with medications. Cardiac MRI performed at 11 years of age revealed normal ventricular size and function, but there was mild right atrial enlargement. Holter Monitoring revealed frequent runs of ectopic atrial tachycardia and some episodes of first-degree A-V block, as well as frequent ventricular arrhythmias, including 12 runs of V-tach. By the age of 12 years, the patient developed atrial flutter/fibrillation with rapid ventricular response and requiring admission to the intensive care unit. Over the course of the admission, his electrocardiograms seemed to change from episodes of atrial tachycardia, fibrillation, and flutter with variable conduction. An implantable cardioverter-defibrillator (ICD) was inserted to prevent death. The genetic variant, p.Leu42Ser, involves the replacement of neutral and non-polar leucine with neutral and polar serine at codon 42 of the LMNA protein. This variant is absent in population databases (gnomAD). Similar variants have been observed in individuals with clinical features of LMNA-related conditions (Invitae), and it has been observed as de novo in at least one case. Predictive algorithms (SIFT, PolyPhen-2, Align-GVGD) all suggest that this missense mutation is likely to have a disruptive effect on protein structure and function. Given the patient's fulfillment of clinical criteria for a severely affected individual with laminopathy due to LMNA gene mutations, this variant is considered pathogenic.

Session Title: Mendelian Phenotypes Poster Session II

PB4576 A case of *MBTPS1*-related disorder due to compound heterozygous variants in *MBTPS1* gene: genotype-phenotype expansion and the emergence of a novel syndrome

Authors:

K. Liaqat[1,2,3], K. Treat[1,2,3], L. Mantcheva[1,2,3], A. Nasir[4], E. Conboy[1,2,3], F. Vetrini[1,2,3]; [1]Indiana Univ. Sch. of Med., Indianapolis, IN, [2]Dept. of Med. and Molecular Genetics, Indiana Univ. Sch. of Med., Indianapolis, IN, [3]Undiagnosed Rare Disease Clinic (URDC), Indiana Univ. Sch. of Med., Indianapolis, IN, [4]Dept. of Anesthesiology, Second Affiliated Hosp. of Zhengzhou Univ., Zhengzhou, China

MBTPS1 (NM_003791.4) encodes Site-1 protease (S1P) is a serine protease that functions sequentially with Site-2 protease (S2P) regulating cholesterol homeostasis and endoplasmic reticulum stress response. Bi-allelic loss of function variants in *MBTPS1* are currently associated with Spondyloepiphyseal dysplasia, Kondo-Fu type (MIM:618392). More recently, a few additional syndromic conditions have been associated with the *MBTPS1* disruptions, including cataract, alopecia, oral mucosal disorder, and psoriasis-like (CAOP) syndrome, and Silver-Russell-like syndrome (SRS). In this report, we describe a 14-year-old female with a complex medical history including early-onset cataracts, laryngomalacia, feeding dysfunction, white matter volume loss, inguinal hernia, joint hypermobility, retrognathia, speech delay and dysmorphic features. Additionally, features of ectodermal dysplasia including decreased sweating, heat intolerance, dysplastic nails, chronically dry skin, hair growth issues, and concerns for a mitochondrial disorder as fatigue and hypotonia, were also observed. Previous testing including *MECP2* sequencing, SRS and Angelman and Prader-Willi testing, metabolic screening, CMA karyotype analysis were all negative. To identify the underlying genetic etiology in this patient, whole exome sequencing (WES) analysis was performed, which identified compound heterozygous variants in the *MBTPS1* gene, c.2255G>T: p.(Gly752Val) from mother and a splice site variant c.2831+5G>T from father. The glycine at amino acid position 752 is highly conserved among various species and the 3D modeling predicts a slight local perturbation of the loop region in the mutant protein due to difference in interaction pattern in wild and mutant structure. The effect of the c.2831+5G>T was investigated by RNA-seq analysis and resulting Sashimi plot showed skipping of exon 21, predicted to result in frameshifting p.Ser901fs28* leading to non-sense mediated decay (NMD).To date only few studies have been published that described the *MBPTS1*-related disorders. Interestingly, we observed ectodermal dysplasia in the patient that further expands the phenotypic spectrum of *MBTPS1* gene related disorders. In a recent study, it has already been demonstrated that *MBTPS1* bi-allelic loss of functions causes CAOP syndrome due to mitochondrial dysfunction responsive to riboflavin (Vit.B12) treatment (PMID: 35362222). Of note, we have begun the treatment with Vit.B12 and our patient's hair has begun to grow back. This case expands our current knowledge of this rare syndromic condition and expands the genetic and phenotypic spectrum of the *MBTPS1*-related disorders.

Session Title: Mendelian Phenotypes Poster Session III

PB4577 A case report and review of the literature suggests that Turner syndrome associated with Xp deletions may cause increased autoimmune disease due to aberrant X chromosome inactivation in immune cells.

Authors:

A. Rudrapatna[1], **R. Slaughenhaupt**[1], J. Owens[2], A. Shillington[2], N. Sheanon[2]; [1]Univ. of Cincinnati, Cincinnati, OH, [2]Cincinnati Children's Hosp., Cincinnati, OH

Turner Syndrome (TS), caused by a partial or complete loss of an X chromosome, causes short stature, hypogonadotropic hypogonadism, and characteristic dysmorphic features. There are widely variable rates of autoimmune disease, hearing deficits, lymphedema, skeletal anomalies, congenital heart disease, and renal disease. Despite the relatively high frequency of TS and recognized increased risk of autoimmune disease, inflammatory bowel disease is seen in only 2-7% of TS patients. Here, we describe a 4-year-old girl with short stature, coarctation of the aorta, and early-onset inflammatory bowel disease at 3 months of age. A SNP microarray revealed 78.5% of cells with monosomy X and 22.5% of cells with an Xp22.33p11.22 53.4 Mb terminal deletion, consistent with a diagnosis of mosaic TS.

Small cohorts have demonstrated that, in comparison to monosomy X, patients with TS due to Xp deletions have a higher incidence of autoimmune disease (66.0% vs 31.5%) and inflammatory bowel disease (17.4% vs 7.2%). Individuals with an Xp deletion will generally have preferential inactivation of their partial X chromosome, though a small subset of cells may persist with expression of the partial X chromosome. In most organ systems, this small subset of cells expressing the partial X chromosome is likely not physiologically meaningful. However, immune cells have a special capacity for signal amplification that could allow for only a few cells expressing a partial X chromosome to cause disease. Comparing the Xp deletion of the reported patient to Xp deletions of other TS patients with early-onset inflammatory bowel disease suggested a potential critical region encompassing 400 genes. Of these, *FOXP3* and *POLA1* notably act as immune-supressing genes; loss of function of these genes can lead to increased inflammation via IL-10, NF-κB, and type 1 interferon signaling pathways. Thus, we propose that this paradoxical finding of a smaller deletion causing more severe autoimmune disease is driven by skewed X chromosome inactivation (XCi) in a small subset of immune cells that would occur in partial X chromosome deletions but not monosomy X.

Session Title: Mendelian Phenotypes Poster Session I

PB4578 A case series: phenotypic variations of male MECP2 disorders

Authors:

J. Lee[1], M. F. Barros Perez[1], R. Vela Sosa[1], C. Toribio-Dionicio[1], G. Fradkis[2], M. Banikazemi[3,4]; [1]Dept. of Pediatrics, New York City Hlth.and Hosp. Corp., Metropolitan Hosp. Ctr., NEW YORK, NY, [2]Dept. of Genetics, New York City Hlth.and Hosp. Corp., Metropolitan Hosp. Ctr., NEW YORK, NY, [3]Dept. of Genetics, New York City Hlth.and Hosp. Corp., Metropolitan Hosp. Ctr., New York, NY, [4]Dept. of Pediatrics, New York City Hlth.and Hosp. Corp., Metropolitan Hosp. Ctr., New york, NY

Background: Rett syndrome was previously considered to be exclusive to females and lethal in males. Recently, Rett syndrome with MECP2 variations has been recognized as a distinct entity, and MECP2 disorders are now classified into two separate categories based on gender. Males with MECP2 disorders exhibit three subtypes: severe neonatal encephalopathy, intellectual disability, as well as pyramidal signs, parkinsonism, and macroorchidism (PPM-X) syndrome. The approval of Trofinetide by the FDA in 2023 represents a significant therapeutic breakthrough for Rett syndrome, as it promotes neurological development in individuals with MECP2 mutations regardless of gender. Objective and methods: We identified male patients with MECP2 disorders and characterized their clinical courses. A cohort of 7,790,503 who were screened for male gender and a diagnosis of Rett syndrome with a reported MECP2 mutation. Results: A total of three male patients with reported MECP2 variation were identified. Two of these patients were maternal half-brothers aged 14 and 15 years. The genetic variation was identical between the half-brothers - a novel, frame shift mutation (c.1128_1156del29) in exon 4. Their mother was a carrier of this mutation and had borderline intellectual disability. The elder sibling presented with symptoms of the intellectual disability subtype of male MECP2 disorders. He was initially evaluated for global developmental delay at 2 years of age, which later progressed to profound intellectual disability and spastic cerebral palsy. Interestingly, the younger sibling's presentation was more suggestive of severe neonatal encephalopathy subtype. During infancy, he experienced severe global developmental delay, followed by respiratory insufficiencies after viral infections, and epilepsy at 5 years of age. The most recent case was a full-term neonate who developed a more severe form of neonatal encephalopathy. At birth, he presented with microcephaly and hypotonia. He required a gastrotomy tube insertion for feeding and a tracheostomy for central apnea by the age of 6 months. Genetic analysis identified a variation (c.317 G>A; p.R106Q) in exon 3 of the *MECP2* gene. The same variation has been previously reported in 22 female patients identified from RettBASE, of which 8 patients exhibited classical Rett syndrome. Conclusions: We report three patients of male MECP2 disorders who have not been previously reported in the literature. Recent therapeutic advancements provide hope for changing this devastating disease into a potentially treatable condition, making timely diagnosis even more crucial.

Session Title: Mendelian Phenotypes Poster Session II

PB4579 † A clinical knowledge graph-based machine learning framework to prioritize candidate genes for facilitating diagnosis of Mendelian diseases and rare genetic conditions.

Authors:

R. Gnanaolivu, G. Oliver, G. Jenkinson, N. Chia, E. Klee, C. Wang; Mayo Clinic, Rochester, MN

Background: Diagnosis of mendelian and rare genetic condition requires identification of clinical phenotypes-associated genetics findings and prioritization of likely disease-causing genes. With a large number of genetic variants detected from clinical next-generation sequencing (NGS), effective computational prioritization is desired to integrate with knowledge of patients' phenotypes.
Methods: We seek to develop a machine learning (ML) framework prioritizing candidate diagnostic-relevant genes for genetics conditions, leveraging a large-scale clinical knowledge graph (CKG) consisting of 19,405,058 nodes and 217,341,612 edges as relationships, curated from 24 databases and ten ontology data sources such as the Human Phenotype Ontology (HPO). The entire CKG is stored in a Neo4j graph database for easy querying, visualization, and computations. The developed framework includes two major ML modules: (i) prediction of reliable knowledge relationships using a graph embedding approach FastRP (Fast Random Projection), and (ii) gene-phenotype relevance ranking according to the Cosine similarity of gene and phenotype nodes in CKG, using the predictions and information content for each phenotype as weights. Multiple ML strategies were compared and optimized for both modules. The optimal framework was tested in an institutional rare disease cohort of 250 patient subjects with research consent. PCAN (Phenotype Consensus Analysis), an established gene-phenotype prioritization method without utilizing knowledge relationships, is compared with our proposed framework.
Results: Within the proposed ML framework, we evaluated four graph edge embedding operations (L1, L2, Average, and Hadamard), using three ML algorithms (XGboost, Naïve Bayes, Multilayer Perceptron) to evaluate link predictions. L1 operation with Naïve Bayes performed best with an AUPRC=0.92, followed by L2 with Naïve Bayes with an AUPRC=0.91. When applying the best models to an institutional rare disease cohort with NGS-detected genetic variants, our method ranked the causal gene within the top 40 out of 42,571 genes and 15,872 phenotypes in the CKG network for 91% of the cases, with 97 cases ranking the causal gene within the top 10. With PCAN only 20% of the cohort ranked top 40, and only 18 (7%) cases had the causal gene ranked higher than our model.
Conclusions: These results suggest leveraging knowledge graph methods for facilitating genetic condition diagnosis have value. The proposed ML framework is a scalable solution for integrating multiomics data modalities and identifying disease-causing pathway alternations for other complex diseases.

Session Title: Mendelian Phenotypes Poster Session III

PB4580 A cohort of Tunisian patients with Netherton syndrome: Molecular signature and Genetic investigations

Authors:

H. CHOUK[1,2], E. Petrova[3], L. Boussofara[4,5], N. Litaiem[6], H. Hammami[7], M. Mokni[8], B. Sriha[9], S. Miskinyte[3], A. Saad[4,2], M. Denguezli[5], A. Hovnanian[10], D. H'mida[4,2]; [1]Higher Inst. of Biotechnology of Monastir, Sousse, Tunisia, [2]Laboratry of Cytogenetics, Molecular Genetics and Reproductive Biology, Sousse, Tunisia, [3]Imagine Inst., Paris, France, [4]Faculty of medicine, Sousse, Tunisia, [5]Service of Dermatology, Farhat Hached, Sousse, Tunisia, [6]Service of Dermatology, Charles Nicolle, Tunis, Tunisia, [7]Service of Dermatology, Habib thameur, Tunis, Tunisia, [8]Service of Dermatology, La Rabta, Tunis, Tunisia, [9]Service of pathology, Farhat Hached, Sousse, Tunisia, [10]Imagine Inst. - INSERM URM 1163, Paris, France

Introduction & Objectives: Netherton syndrome (NS) is a rare autosomal recessive genodermatosis, it is actually defined by the clinical triad characterized by an atopic manifestation, scaly erythroderma (SE) or Ichthyosis Linearis Circunflexa (ILC) and a specific Hair shaft abnormality known as Trichorrhexis invaginata (TI). NS is caused by loss of function mutations in the *SPINK5* gene encoding kazal-type serine protease inhibitors (LEKTI) highly expressed in the stratified epithelium. We aim to analyze NS in Tunisian patients by examining their clinical, biological, histological, and genetic features, and identifying therapeutic targets through analysis of the inflammatory signatures in their skin. **Materials & Methods:** Our study focused on 17 patients from 13 unrelated Tunisian families with NS, who were referred to our department for genetic confirmation of their diagnosis. **Results:** All patients experienced neonatal onset of symptoms, with congenital ichthyosiform erythroderma in 13 cases at birth, and bullous in only one case. The scalp was the site of a diffuse scaly shell. TI, as well as atopic dermatitis, was noted in all patients. Immunostaining of inflammatory cytokines showed massive neutrophil and mast cell infiltrates in the lesion and non-lesion skin in both NS subtypes. A remarkable expression of Th2 cells in the dermis for the two subtypes was noted. Next-generation sequencing and Sanger sequencing of the *SPINK5* gene identified a known homozygous c.1888-1G>A mutation at the splice acceptor site of exon 21 in 7 patients. A homozygous c.2264dupA mutation at exon 24 was present in 2 patients. A homozygous c.2471delAAGA deletion in 2 related patients. A previously described homozygous c.2441+3delCAGT mutation at the splice donor site of exon 25 in one patient. A new nonsense pathogenic variant homozygous c.217G>>T at exon 4 in 2 patients. Furthermore, the last patient was compound heterozygous for a new c.2302G>T mutation, and a deep intronic variant. Haplotype analysis in NS patients by genotyping seven microsatellite markers flanking the *SPINK5* gene indicates that all carriers of the c.1888-1G>A mutation have the same haplotype. **Conclusion:** In conclusion, we present the largest cohort of Tunisian patients with NS, including 7 patients harboring the same mutation, potentially indicating a founder effect. We have also identified 5 pathogenic variants, including 2 novel variants, in a total of 10 patients. Our findings suggest that intra and inter-familial phenotypic variations observed in individuals with the same *SPINK5* mutation may be influenced by other epigenetic and environmental factors, highlighting the complexity of NS expression.

Session Title: Mendelian Phenotypes Poster Session I

PB4581 A *COL12A1* heterozygous mutation leads to bent bone dysplasia.

Authors:

A. Kot, J. Martin, D. Wachtell, C. Chun, J. Zieba, D. Krakow; UCLA, Los Angeles, CA

Bent bones are described in several skeletal disorders including campomelic dysplasia, bent bone dysplasia FGFR2 type, osteogenesis imperfecta, and hypophosphatasia - the underlying genetic causes of these disorders are well established. Herein, we describe a novel bent bone skeletal dysplasia resulting from a heterozygous missense mutation in *COL12A1*, encoding collagen XII. Collagen XII is expressed in the perichondrium, articular cartilage, and prehypertrophic zone. It interacts with collagen I and other ECM proteins to organize collagen fibrils in dense connective tissue and bone. Collagen XII variants are described in patients with phenotypes including Ehlers-Danlos syndrome and collagen VI related myopathy.

This novel bent bone case was a female at 21 weeks of gestation with nonconsanguineous parents of 2 unaffected children. Radiographic findings included distinctly bent appendicular bones, coronal clefts, underminerlized 12th ribs, and delayed pubis/calcaneus ossification. Trio exome analysis revealed 6 candidate de novo mutations, of which *COL12A1* was prioritized due to its expression in the skeleton. The mutation (NM_004370.5 c.4120A>G, p.Asn1374Asp) occurs 3 residues past a von Willebrand factor domain, which function in cell adhesion and ECM binding.

Histological analysis of the affected femur showed reduced collagen fibril organization with a mildly disorganized hypertrophic zone. Immunofluorescence studies from affected fibroblasts suggest the mutation leads to intracellular retention of collagen XII. As SOX9 expression is critical for chondrocyte differentiation, is expressed in the prehypertrophic zone, and SOX9 mutations cause campomelic dysplasia, we hypothesized SOX9 may be dysregulated in the affected cells and growth plate. Western blot analysis found Increased SOX9 in a primary chondrocytes and fibroblasts. Immunohistochemistry showed increased SOX9 expression throughout the growth plate with strong staining surrounding cartilage canals. This finding adds to the diverse set of molecules that produce bent bone dysplasias, highlights roles of the ECM in skeletogenesis outside of structural support, and implicates SOX9 expression as a novel mechanism in these non-campomelic dysplasia bent bone cases.

Session Title: Mendelian Phenotypes Poster Session II

PB4582 A comparative analysis of a rare condition with a skeletal phenotype - *FLNB* related spectrum: case series

Authors:

V. Plaiasu, B. Pascu, O. Farkas, V. Comanici, I. Stan; INSMC Alessandrescu-Rusescu, Bucharest, Romania

Background: FLNB-related disorders are a group of disorders characterized by skeletal abnormalities affecting the bones of the hands and feet, the bones of the spine (vertebrae), joint dislocations, and distinctive facial features. The FLNB disorders include a continuum spectrum of phenotypes ranging from mild to severe: spondylocarpotarsal synostosis (SCT) syndrome and Larsen syndrome, boomerang syndrome, atelosteogenesis type I (AOI) and type III (AOIII) and Piepkorn osteochondrodysplasia. While pathogenic variants in FLNB are typically inherited in an autosomal dominant fashion, autosomal recessive inheritance is also possible. Material and methods: Our study makes a comparative analysis among 3 probands with FLNB-mutated autosomal dominant and autosomal recessive spectrum, relating clinical phenotypes to molecular findings. All patients were screened for mutations by Whole Exome Sequencing (WES) technology because they exhibited variability of the clinical findings with a wide and unspecific range of clinical and radiological features involving the skeleton. Results: One patient was diagnosed with Larsen syndrome, an autosomal dominant skeletal dysplasia characterized by large joint dislocations and craniofacial dysmorphism and two patients were described with autosomal recessive spondylocarpotarsal synostosis syndrome, a complex phenotype characterized by disproportionate short stature, spine deformation, carpal and tarsal fusions, mildly dysmorphic features, one of them presented with a rare association between cricopharyngeal achalasia and laryngomalacia with spondylocarpotarsal synostosis syndrome. Our report expands the genetic spectrum of FLNB pathogenic variants with 3 novel mutations. Conclusion: Mutations in FLNB gene lead to various skeletal malformations, involving long bone, joints and vertebra. Clinical, skeletal findings, radiographic examination and molecular results could offer the final diagnosis for the patients. Milder or unusual manifestations of this group of diseases can result in misdiagnosis or no diagnosis at all. But the use of broader genetic testing, such as WES, is important to accurately diagnose individuals with FLNB-related conditions. Recognition of these syndromes also will have implications in genetic counselling for patients.

Session Title: Mendelian Phenotypes Poster Session III

PB4583 A consanguineous Yemeni Family with 7 sibs affected by either Severe Anterior chamber defects of eyes due to *PXDN* pathogenic variant *exon 17 C/T* and/or Intellectual disability of Unknown Etiology.

Authors:

M. Alwasiyah[1], C. Trujillo[2], R. Lebel[3], S. Scherer[4], M. Zarrei[4]; [1]Jeddah Regional Lab., Jeddah, 21523, Saudi Arabia, [2]Univ. CES, Medellin, Colombia, [3]SUNY Upstate Med Univ, Fayetteville, NY, [4]The Hosp. for Sick Children, Toronto, ON, Canada

A consanguineous Yemeni Family with 7 sibs affected by severe anterior chamber defects of the eyes and/or Intellectual disability. 6 brothers and one sister, parents are 1st cousins. Only one "normal male" 26 years old who came to our clinic for pre-conceptional counseling and testing. The eldest brother is 30 years old affected by severe intellectual disability, ataxic gait, and bilateral microphthalmia with microcornea and diffuse corneal opacities. The second 29 years old "normal" intellect microphthalmia with microcornea bilaterally, in the right eye had a previous surgery 20 years ago, has anterior polar cataract, adhesion with anterior posterior synechia of iris and lens. The third one is 27 years old male, "normal" eyes, severe intellectual disability. The fourth is the proband 26 years old, married, they are not relatives but their families come from the same village in Yemen. The fifth is a 20 years old female with mild myopia and "moderate" intellectual disability. The sixth is 17 years old male, normal intellect has bilateral microphthalmia and microcornea with diffuse corneal opacities, he had in his left eye pupil of slade shape, a corneal graft, and right eye cornea is clear, he can read at very close distance, the retina looks normal. Last one a 7 years old male, normal intellect, left eye, upper corneal opacity, microcornea, microphthalmia, anterior and posterior synechia. Right eye Bupthalmos, Glaucoma, diffuse central opacity and vascularization. We have completed WGS and targeted sequencing experiments, as well as the analysis of the data for this family. Based on the information we have available today, there were no SNVs observed that appear to be contributing to the ID phenotype.We do have a potentially, clinically relevant finding for the eye abnormality. We have detected the following *SNV:PXDN:NM_012293:exon17:c.G3488A:p.R1163Q Chr2: 1652064 C/T*

Session Title: Mendelian Phenotypes Poster Session I

PB4584 A *Drosophila* Model for MED12-associated Rare Genetic Diseases

Authors:

K. Bussey, N. Nazario-Yepiz, V. Shankar, R. R. H. Anholt, T. F. C. Mackay; Clemson Univ.-Ctr. for Human Genetics, Greenwood, SC

MED12 is a component of the Mediator Complex which regulates gene expression. Mutations in different domains of the MED12 protein are associated with five rare genetic disorders: Lujan Syndrome, FG Syndrome, X-Linked Ohdo Syndrome, Hardikar Syndrome, and Non-Specific Intellectual Disability. The clinical phenotypic spectrum of these disorders is broad, including intellectual disability, behavioral, and congenital defects. Individuals with identical genetic lesions in MED12 can present different severity of symptoms, indicating that genetic modifiers may act on MED12 to alter the manifestation and severity of these rare disorders. Due to the rarity and diversity of these disorders, identifying genetic modifiers is not possible in human populations. *Drosophila melanogaster* has a conserved functional ortholog of MED12, *khotalo* (*kto*). Knocking down *kto* expression using RNA-interference shows that *kto* affects startle response, a proxy for sensorimotor integration, and sleep and activity traits. We can use natural variation among the fully sequenced inbred lines of the *D. melanogaster* Genetic Reference Panel (DGRP) to identify genetic modifiers that alter the effects of *kto* mutations. We crossed a CRISPR/Cas9-generated null mutation of *kto* (kto^{Δ}), and a wild type *kto* allele (kto^{+}) in the same genetic background, to the DGRP lines, and quantified sleep and activity and startle-induced locomotion in the F_1 offspring across 200 different genetic backgrounds. We used analysis of variance to estimate the effects of *kto* genotype (*G*), DGRP line (*L*) and the genotype by line interaction (*G*×*L*) on these phenotypes. A significant *G*×*L* term indicates the presence of modifier loci, (*i.e.*, the effect of *kto* genotype is different across the DGRP lines). Variation in the phenotypic difference between F1 offspring of the wild-type and kto^{Δ} mutant for each DGRP line can be used as a parameter for genome-wide association analysis across the DGRP to identify candidate epistatic modifiers of *kto* with human orthologs. We will report the results of these analyses. Genetic modifiers identified in *Drosophila* will provide insight into how genetic modifiers may influence the manifestation and severity of *MED12*-related disorders that are potential therapeutic targets.

Session Title: Mendelian Phenotypes Poster Session II

PB4585 A Drosophila Model for Mucopolysaccharidosis IIIA

Authors:

R. Bishop, T. F. Mackay, R. R. Anholt, B. Simkhada, P. S. Freymuth, N. O. Nazario-Yapiz, V. Shankar, R. Lyman; Clemson Univ., Greenwood, SC

Sanfilippo syndrome IIIA is a rare lysosomal storage disorder due to a genetic defect within the N- *sulfoglucosamine sulfohydrolase* (*SGSH*) gene, which is a sulfamidase that cleaves a sulfamate bond from heparan sulfate and GM2 gangliosides during catabolic degradation within the lysosome. To assess the effect of genetic background modifiers on disease progression, we used a deletion mutant and nonsynonymous missense mutants of the Drosophila *Sgsh gene* which is orthologous to its human counterpart. Assessment of lysosome integrity in the brains of control and mutant flies for both sexes separately using the Lysotracker Green assay showed an age-dependent progressive increase in fluorescent punctae in the brains of mutants versus controls. Hyperactivity and fragmented sleep patterns of children between ages 5 -10 years-old are prominent hallmarks of the disease phenotype. To assess effects on locomotor behavior and sleep in the Drosophila model, we measured activity at weekly intervals up to 3 weeks of control and mutant flies for sexes separately using Drosophila activity monitors. This system monitors the number of times a single fly housed in a narrow tube crosses the beam of an infrared light. Both the Lysotracker system and the Drosophila activity monitor system can be used to identify genetic modifiers by crossing the *Sgsh* mutants and the control to flies of the Drosophila Genetic Reference Panel, a population of fully sequenced wild derived inbred lines, and analyze variation in differences between control crosses and the corresponding mutants in the resulting heterozygotes. Supported by a Cure Sanfilippo Foundation grant.

Session Title: Mendelian Phenotypes Poster Session III

PB4586 A Drosophila model for Mucopolysaccharidosis type IIIB (MPS IIIB)

Authors:

B. Simkhada, N. O. Nazario-Yepiz, R. R. H. Anholt, T. F. C. Mackay, P. S. Freymuth, R. A. Lyman, V. Shankar; Clemson Univ., Greenwood, SC

Mucopolysaccharidosis type IIIB (MPS IIIB) is a rare human lysosomal storage disorder caused by defects in the lysosomal enzyme alpha-*N*-acetylglucosaminidase (*NAGLU*). Affected individuals appear normal at birth, with clinical symptoms appearing only at around 3-5 years of age. Early symptoms include hyperactivity and lack of sleep, which manifests into neural degeneration by mid-teen years, and death before the age of 30. Mutations in *NAGLU* cause accumulation of partially degraded heparan sulfate in lysosomes, but how this relates to neural degeneration is unknown. We use the functional *Drosophila melanogaster* ortholog of *NAGLU*, *CG13397*, to characterize disease models induced by gene deletion, missense (Drosophila Y160C corresponding to human Y140C), and nonsense (Drosophila W422X corresponding to human W404X) mutations. Fluorescence microscopy on mutant fly brains using Lysotracker dye reveal a significant increase in acidic compartments, characteristic of lysosomal storage diseases. Using the Drosophila Activity Monitor (DAM) system to analyze activity and sleep patterns, we find hyperactivity and sleep defects in mutant flies. RNA sequencing of fly brains can reveal differentially expressed genes to provide mechanistic insights into MPS IIIB pathogenesis.

Session Title: Mendelian Phenotypes Poster Session I

PB4587 A dual diagnosis of Okur-Chung neurodevelopmental syndrome and Becker muscular dystrophy highlights the lower limits of neurodevelopmental functioning attributable to muscular dystrophy

Authors:

V. Liu[1], E. Hanson[1], J. Owens[2], A. Shillington[3]; [1]Univ. of Cincinnati Coll. of Med., Cincinnati, OH, [2]Cincinnati Children s Hosp., Cincinnati, OH, [3]Cincinnati Children s Hosp., Cincinnatti, OH

Muscular Dystrophies caused by pathogenic changes in the *DMD* gene present as a spectrum of X-linked recessive disorders. Disease severity correlates to the degree that pathogenic variants disrupt the dystrophin protein, and disruption of specific dystrophin isoforms in the brain can lead to more severe neurodevelopmental delays. Okur-Chung Neurodevelopmental Syndrome (OCNDS) is an autosomal dominant disorder characterized by developmental delay, intellectual disability, dysmorphic facial features, generalized hypotonia, and short stature due to pathogenic *CSNK2A1* variants. We report a 3-year-old male who presented with global developmental delay, postnatal growth failure, feeding difficulties, self-injurious behaviors, and recurrent infections. His facial features are notable for hypertelorism with epicanthic folds, a flat nasal bridge, mild cupid's bow lip, and mild generalized hypotonia. A SNP microarray revealed a maternally inherited 109 kb Xp21.1 interstitial duplication with out-of-frame breakpoints in exon 55 of *DMD*. His *DMD* variant and normal CPK / muscle MRI findings were suggestive of Becker Muscular dystrophy (BMD). Given his dysmorphic features and excessive global delays for BMD, he received whole exome sequencing revealing a de novo *CSNK2A1* c.593A>G (p.Lys198Arg) pathogenic variant consistent with a diagnosis of OCNDS. His clinical history emphasizes the limitations of neurodevelopmental functioning that can be attributed to BMD and Duchenne Muscular Dystrophy (DMD). Large cohorts predict a full scale IQ (FSIQ) of 88.3 ± 13.9 among all patients with BMD and 86.1 ± 15.0 among all patients with DMD, while pathogenic variants impacting the brain dystrophin isoform Dp140 are associated with FSIQ of 77.7 ± 10.8 in BMD and 78.8 ± 18.6 in DMD. We propose that all patients with muscular dystrophy with FSIQ 1 standard deviation below these expected ranges should be assessed for alternative causes for their neurodevelopmental delays, with specific consideration of broad-spectrum genetic testing. All patients with an FSIQ 2 standard deviations below these ranges should receive broad-spectrum genetic testing to screen for an additional genetic diagnosis.

Session Title: Mendelian Phenotypes Poster Session II

PB4588 A genetics-first approach to identify novel variants of the calcium sensing receptor associated with autosomal dominant hypocalcemia type 1

Authors:

J. B. Chang[1], M. M. Soliai[1], C. P. Barnhill[1], L. M. Stapleton Smith[1], A. S. Mathew[1], A. M. Apostolov[2], B. W. Dulken[1,3], A. Petukhov[1], A. V. Sridhar[1], J. Lasky-Su[4], C. Lange[5,4], R. Altman[6], S. H. Adler[1], M. Roberts[1], J. C. Fox[1], S-G. Ji[1]; [1]BridgeBio Pharma, Inc., Palo Alto, CA, [2]Dept. of Bioengineering, Stanford Univ., Stanford, CA, [3]Dept. of Pathology, Stanford Sch. of Med., Stanford, CA, [4]Channing Div. of Network Med., Brigham and Women's Hosp. and Harvard Med. Sch., Boston, MA, [5]Dept. of Biostatistics, Harvard T.H. Chan Sch. of Publ. Hlth., Boston, MA, [6]Dept.s of Bioengineering, Genetics, and Med., Stanford Univ., Stanford, CA

Autosomal dominant hypocalcemia type 1 (ADH1) is a rare genetic disorder caused by gain-of-function (GOF) variants in the calcium sensing receptor, *CASR*, which regulates calcium homeostasis by controlling parathyroid hormone secretion. Diagnosis of ADH1 can be challenging due to variability in clinical presentation and limited knowledge of which *CASR* variants result in GOF. To address these challenges, we integrated biobank and patient data to identify novel *CASR* variants associated with ADH1. We first determined that the prevalence of previously reported *CASR* GOF variants was 2.8-3.8 in 100,000 individuals in gnomAD, TOPMed, and the UK Biobank (UKB), similar to a previous report on the Geisinger cohort. Next, to identify novel GOF variants, we analyzed *CASR* variants in exome sequencing data from 500,000 participants in the UKB. We first confirmed the association of 3 out of 4 previously reported variants found in non-Hispanic White probands with reduced serum calcium level, a hallmark phenotype of ADH1 (Thr151Met: $p = 3.8\times10^{-6}$, Asn124Lys: $p = 2.4\times10^{-3}$, Val104Ile: $p = 8.8\times10^{-3}$, Thr888Met: $p = 0.13$; Bonferroni-adjusted individual significance threshold 1.3×10^{-2}). Val104Ile was also associated with reduced serum calcium level in 4 individuals of Indian ancestry ($p = 2x10^{-6}$). We applied SKAT-O to these 4 variants and 1,518 phenotypes and found that they were associated with hypoparathyroidism, another characteristic ADH1 phenotype ($p = 1\times10^{-56}$, Bonferroni-adjusted significance threshold 3.3×10^{-4}), and five additional phenotypes, suggesting these variants are pleiotropic. Next, we scored all 1,253 variants in the UKB within each ancestry based on regression modeling of ADH1 phenotypes, such as diagnosis code-based phenotypes. Our approach revealed novel associations of multiple variants with ADH1 phenotypes. Notably, one variant (Ser247Phe) scored highly across multiple ancestries. Finally, to search for novel variants beyond public datasets, we evaluated data from 169 patients participating in a sponsored genetic testing program for suspected genetic hypoparathyroidism. We identified a pathogenic/likely pathogenic variant or variant of unknown significance (VUS) of *CASR* in 21% (36/169) of patients. One of the VUS (Ala364Glu) was also found in the UKB and ranked highly by our scoring method. Currently, we are attempting to replicate our findings in two other biobanks (All of Us and Mass General Brigham) and are investigating how these novel ADH1-associated variants impact CASR function *in vitro*. Our study emphasizes the power of large, publicly available genomic datasets to clarify the genotype-phenotype relationship of rare diseases like ADH1.

Session Title: Mendelian Phenotypes Poster Session III

PB4589 A genotype-first approach identifies high incidence of *NF1* pathogenic variants with distinct disease associations

Authors:

T. Drivas[1], A. Safonov[2], T. Nomakuchi[3], E. Chao[4], V. Speare[4], C. Horton[4], J. Dolinksy[4], A. Yussuf[4], H. Laduca[4], Penn Medicine BioBank, Regeneron Genetics Center, Z. Bogus[1], M. Bonanni[1], A. Raper[1], M. Ritchie[1], S. Kallish[1], K. Nathanson[1]; [1]Univ. of Pennsylvania Perelman Sch. of Med., Philadelphia, PA, [2]Mem. Sloan Kettering Hosp., New York, NY, [3]Childrens Hosp. of Philadelphia, Philadelphia, PA, [4]Ambry Genetics, Irvine, CA

Loss of function variants in the *NF1* gene cause neurofibromatosis type 1 (NF1), an autosomal dominant genetic disorder classically characterized by complete penetrance, a prevalence of 1 in 3,000, characteristic physical exam findings, and a substantially increased risk for malignancy. However, our understanding of the disorder is entirely based on patients ascertained through phenotype-first approaches. We have recently been referred four patients with incidentally discovered pathogenic *NF1* variants, but with no features of the syndrome on exam or history. We hypothesized that the true population-level incidence of *NF1* pathogenic variants might be higher than reported, with reduced penetrance or a higher incidence of somatic mosaicism than is currently known.

To investigate this hypothesis, we evaluated two unique large patient cohorts from independent datasets that had undergone comprehensive sequencing of the *NF1* gene: the population-level Penn Medicine Biobank (PMBB, n = 43,731) and a database of patients clinically sequenced for cancer risk evaluation by Ambry Genetics (n = 118,768). We identified an unexpectedly high prevalence (1 in 450-750) of pathogenic variants in *NF1*, more than four times the rate expected given the reported prevalence of NF1. Half of these individuals lacked any evidence of syndromic NF1, and 15-30% of these individuals appeared to be post-zygotic mosaic for the *NF1* variant identified. The discovery of an incidental *NF1* pathogenic variant did not correlate with the presence of classic symptoms of NF1 but was associated with a significantly greater incidence of certain malignancies compared to a matched control population, including ovarian cancer (p=0.01), sarcoma (p=0.04), adrenal cancers (p=1.5e-11), CNS cancers (p=0.04), and hematologic malignancies (p=3.8e-04). Our findings suggest that *NF1* pathogenic variants are substantially more common than previously thought, often characterized by somatic mosaicism and reduced penetrance, and are important contributors to cancer risk in the general population.

Our experience with *NF1* led us to examine the incidence of somatic mosaicism on a larger, population-level scale. Within PMBB, there is clear evidence that nearly all individuals harbor multiple somatic-mosaic variants in various genes, with certain genes being significantly enriched for somatic mosaic variants, at least in peripheral blood. This identification of widespread mosaicism has major implications for future genetic testing and biobanking efforts, and the further investigation of this finding will be critical for accurate counseling of patients and families.

Session Title: Mendelian Phenotypes Poster Session I

PB4590 A heterozygous de novo *FEM1C* variant in a female patient presenting with developmental delay and ataxia.

Authors:

j. verheijen, L. Larson, A. Singla, R. Gavrilova, E. Klee; Mayo Clinic, Rochester, MN

FEM1C (Fem-1 Homolog C) located on chr5q22.3 encodes for a 617 amino-acid substrate recognition protein. FEM1C is involved in protein catabolism through it's role as subunit of the CUL2-RING E3 ubiquitin ligase complex. FEM1C is predicted to selectively interact with substrate proteins by binding to a carboxyl-terminus arginine degradation signal.

We describe a 21-year-old female patient presenting with developmental delay and ataxia. The patient was first evaluated for developmental delay at 13 months of age. The patient was diagnosed with static encephalopathy with hypotonia, ataxia, and global developmental delay at 3 years of age. She subsequently developed quadriplegic cerebral palsy, torsional nystagmus, and esotropia. Whole-genome trio sequencing identified heterozygous *de novo* missense variant c.376G>C p.Asp126His in *FEM1C* in the patient. No additional variants relevant to the patient's phenotype were identified. Previous chromosomal microarray, William's syndrome FISH analysis, and Prader Willi/Angelman analysis returned negative. The variant identified in *FEM1C* is located in residue p.Asp126, which is predicted to be directly involved in substrate binding. This residue is strongly conserved in evolution and present in all three members of the human *FEM1* gene family. Residue p.Asp126 is considered intolerant to substitution, and in-silico tools predict a deleterious effect for variant p.Asp126His. Protein modeling indicates a ΔG Gibbs free-energy change of -1.77.

FEM1C is currently not formally associated with any disorder and categorized as a gene of uncertain significance. However, recent reports have described a neurodevelopmental phenotype in some carriers of *de novo FEM1C* missense variants. Importantly, variant p.Asp126His and alternative substitution p.Asp126Val have recently been reported in pediatric neurodevelopmental disorder patients. In addition, a *de novo* variant in gene *FEM1B* has been recently reported in a pediatric global developmental delay patient, possibly indicating phenotypic overlap for disease variants in the *FEM1* family of genes.

Based on in-silico prediction, absence in healthy cohorts and *de novo* inheritance status, variant p.Asp126His in *FEM1C* is currently curated as VUS with the following criteria assigned: PM2_supporting, PS2_moderate, PS3_supporting. Follow-up analysis will include detailed assessment of variant impact on protein structure and function to further elucidate the pathomechanism for *FEM1C de novo* variants in neurodevelopment.

Session Title: Mendelian Phenotypes Poster Session II

PB4591 A heterozygous mutation in UBE2H in a patient with developmental delay leads to an aberrant brain development in zebrafish

Authors:

Y. Choi[1], U. Shin[1], H. Ko[2], K. Myung[1,2], S. Lee[1], C. Cheon[3], Y. Lee[2,4]; [1]Ulsan Natl. Inst. of Sci. and Technology, Ulsan, Korea, Republic of, [2]Inst. for Basic Sci. (IBS), Ulsan, Korea, Republic of, [3]Pusan Natl. Univ. Children's Hosp., Yangsan-si, Korea, Republic of, [4]Kyung Hee Univ., Seoul, Korea, Republic of

With the advancement of next-generation sequencing, there have been developments in techniques to identify causative variants in patients with rare diseases. However, due to cost and time limitations, panel sequencing is the primary method used. The importance of whole-genome sequencing (WGS) has been emphasized in identifying causative variants in patients who have not been diagnosed using conventional methods such as panel sequencing and microarrays. Additionally, a trio-based approach, in which the parents are also included, can be effective in discovering causative variants more reliably. Therefore, we went beyond conventional methods to identify causative variants in patients using trio-based WGS and experimentally validate them.We carried out trio-based WGS analysis of a patient with clinical phenotypes of developmental delay and intractable convulsion to identify causal mutations that could not be diagnosed using conventional fluorescence in situ hybridization or chromosome microarray tests. Point mutations of all family members (father, mother, and proband) were identified using HaplotypeCaller and only de novo variants, which were defined as those only present in the proband but not in the parents, were selected for further analysis. Rare de novo variants were identified after filtering based on allele frequency (< 0.01) from large-scale variome databases (1KGP, gnomAD, and Korea1K). Notably, the UBE2H (p.Thr150Met) variant was the only rare de novo mutation and was predicted to be deleterious simultaneously by SIFT, PolyPhen, and MutationTaster.Further characterization of the candidate gene, UBE2H, was performed using zebrafish through gene knockdown approaches. We found that Ube2h is required for normal brain development. Transcriptomic analysis using whole embryos of zebrafish knockdown morphants and additional functional studies identified downstream pathways of the Ube2h affecting neurogenesis. Moreover, depletion of ube2h led to the induction of apoptosis, specifically in the differentiated neural cells. Finally, we found that a missense mutation in zebrafish, ube2h (c.449C>T; p.Thr150Met), which mimics a variant identified in a patient with neurodevelopmental defects, causes aberrant Ube2h function in zebrafish embryos. Through our trio-based WGS study, we identified a rare de novo variant present in UBE2H, which may play a role in global developmental delay in pediatric patients. Therefore, we believe that our trio-based approach in analyzing WGS data is effective in discovering causative variants in rare disease patients.

Session Title: Mendelian Phenotypes Poster Session III

PB4592 † A homozygous rare missense variant in *YKT6* results in loss-of-function and is associated with infantile liver failure and developmental delays.

Authors:

M. Ganapathi[1], M. Ma[2], Y. Zheng[2], K-L. Tan[3], O. Kanca[2], N. Quintanilla[4], X. Pan[2], S. Lu[2], J. Rosenfeld[2], C. Thomas[5], C. Leduc[6], A. McPartland[1], J. Hagen[1], Y. Shen[1], P. Hertel[2], W. Chung[1], H. Bellen[2]; [1]Columbia Univ. Med. Ctr., New York, NY, [2]Baylor Coll. of Med., Houston, TX, [3]Baylor Coll. of Med.,, Houston, TX, [4]Texas Children's Hosp., Houston, TX, [5]Carver Coll. of Med., Iowa City, IA, [6]Columbia Univ, New York, NY

Hepatic cirrhosis in the first year of life is most commonly attributed to biliary atresia or genetic/ metabolic disorders, though in up to 15% of cases, the cause remains undetermined. We identified a rare, homozygous, recurrent, predicted likely pathogenic variant in *YKT6* (NM_006555.4:c.554A>G, p.Tyr185Cys) segregating with disease in two unrelated families of South Asian-Indian ancestry. The two affected infants presented with similar clinical features, which included normal prenatal course and birth, poor weight gain and developed micronodular cirrhosis progressing to liver failure. The infant from the first family died of liver failure at 9 months of age. He was also delayed in achieving his gross motor milestones. The second infant was subsequently diagnosed with hepatocellular carcinoma and underwent liver transplant at 1 year of age and is currently 5 years of age but developmentally delayed. These two individuals belong to the Syrian Christian community of Kerala, India, a group currently estimated to be comprised of about 5 million individuals worldwide, and haplotype analysis identified a shared chromosomal region carrying the variant, suggesting a common ancestor. *YKT6* encodes a unique and versatile SNARE (soluble N-ethylmaleimide-sensitive factor attachment protein receptor) that plays important roles in intracellular vesicle trafficking events. Currently, *YKT6* is not associated with any known Mendelian disorder. To investigate the impact of the *YKT6* variants, we performed functional assays in Drosophila. The fly ortholog *dYkt6* is an essential gene, and loss-of-function alleles cause lethality. Lethality was less efficiently rescued by the genomic rescue (GR) construct carrying the *dYkt6* p.Y186C variant (corresponding to human *YKT6* p.Y185C) compared to the wild type, and the rescued animals showed significantly reduced lifespan. A $dYkt6^{KG}$ (Kozak-GAL4) allele generated to assess the expression pattern of *dYkt6* detected expression in the fat body (analogous to vertebrate liver) and central nervous system. Loss of *dYkt6* in fat body cells caused autophagy flux defects, and the *dYkt6* p.Y186C GR construct was unable to fully rescue the autophagy phenotypes. Taken together, our findings suggest that the homozygous *YKT6* p.Tyr185Cys variant leads to partial loss-of-function and is associated with infantile liver failure with developmental delays.

Session Title: Mendelian Phenotypes Poster Session I

PB4593 A homozygous variant in triokinase and FMN cyclase abolishing triokinase activities is associated with isolated immunodeficiency.

Authors:

C. Tremblay-Laganiere[1,2], C. Michaud[1], R. Abourjaili-Bilodeau[1], A. Cabezas[3], J. Canales[3], M. Costas[3], J. Ribeiro[3], J. Leclerc-Blain[1,2], F. Touzot[1,2], E. Haddad[1,2], P. Teira[1,2], M. Duval[1,2], A. Onoufriadis[4], B. Meunier[5], J. Michaud[2,1], J. Cameselle[3], P. M. Campeau[2,1]; [1]Univ. de Montreal, Montreal, QC, Canada, [2]CHU Sainte-Justine, Montreal, QC, Canada, [3]Univ. de Extremadura, Badajoz, Spain, [4]Aristotle Univ. of Thessaloniki, Thessaloniki, Greece, [5]Inst. for Integrative Biology of the Cell (I2BC), Gif-sur-Yvette, France

Background: TKFC is a bifunctional enzyme involved in fructose metabolism. Triokinase catalyzes the phosphorylation of fructose-derived glyceraldehyde (GA) and exogenous dihydroxyacetone (DHA), while FMN cyclase generates cyclic FMN. Previously reported pathogenic variants in TKFC are associated with either a multisystemic disease including developmental delay, cataracts, cardiomyopathy, liver dysfunction, microcytic anemia, lactic acidosis, and cerebellar hypoplasia or isolated hypotrichosis with loose anagen hairs. Methods: Exome sequencing identified a homozygous novel variant in TKFC (c.1624G>A; p.Gly542Arg) in an individual with complex primary immunodeficiency. The variant was characterized using enzymatic assays and yeast studies. Results: The individual presented with recurrent otitis media, refractory rhinosinusitis, chronic active Epstein-Barr virus disease, chronic Herpes simplex virus gingivostomatitis, Varicella zoster virus infection, folliculitis, eosinophilic esophagitis, and food allergies. Clinical investigations revealed hypogammaglobulinemia, increased immunoglobulins A, near absent NK cells and decreased memory B cells. Enzymatic assays showed that this variant displayed inactive DHA and GA kinase activity while maintaining FMN cyclase activity. Conclusion: Our report suggests an important role of TKFC in immunological function. The pathological features of this variant are possibly linked with DHA/GA kinase inactivation with yet an unknown mechanism.

Session Title: Mendelian Phenotypes Poster Session II

PB4594 A late C-terminal frameshift variant in *GPKOW* is associated with a multisystemic X-linked disorder.

Authors:

K. Machol[1,2], L. Mackay[3], J-W. Mok[1,4], M. Balzo[5], E. Mizerik[1,2], F. Valle Sirias[6], M. Weisz-Hubshman[1,2], S. Ketkar[1], U. Polak[1], A. Tran[1], D. Kearney[7], N. Hanchard[1], O. Kanca[1,4], M. Wangler[1,4], H. Bellen[1,4], B. Lee[1,2], Baylor College of Medicine Center for Precision Medicine Models, S. Yamamoto[1,4]; [1]Baylor Coll. of Med., Houston, TX, [2]Texas Children's Hosp., Houston, TX, [3]UT Southwestern, Dallas, TX, [4]Jan and Dan Duncan Neurological Res. Inst., Houston, TX, [5]Baylor Scott and White, Temple, TX, [6]Rice Univ., Houston, TX, [7]Baylor Collage of Med., Houston, TX

GPKOW is an RNA binding nuclear protein with a role in post-transcriptional mRNA processing as a minor spliceosome subunit. *GPKOW* is located on Xp11.13 and encodes a protein with one G patch domain and two KOW motifs. GPKOW was shown to be important for embryonic development in *C. elegans* but its function has not been investigated in other model organisms. Only one published report associated a pathogenic variant in *GPKOW* to a human phenotype, which is characterized by male perinatal lethal microcephaly and intrauterine growth restriction due to a donor splice-site variant (c.331+5G>A).

We present two maternal half siblings with a maternally inherited hemizygous frameshift variant in *GPKOW* (c.1329dupG: p.S444EfsX28). This variant is expected to cause an exchange of the last 33 amino acids with 28 random amino acids and an early protein termination. Clinical presentation included brain and eye phenotype with microphthalmia, optic nerve hypoplasia, absent septum pellucidum, aqueductal stenosis and ventriculomegaly as well as congenital ichthyosis, skeletal abnormalities, and remote occlusion of portal vein branches. Unfortunately, both siblings died in the first year of life due to respiratory failure. mRNA sequencing in fibroblasts from one sibling confirmed that *GPKOW* mRNA escapes nonsense mediated decay with mRNA levels comparable to male controls. Interestingly, protein levels of GPKOW in patient's fibroblasts were overall reduced which may indicate protein instability.

In vivo studies in *Drosophila* showed that the *GPKOW* ortholog (*CG10324*) is essential and is expressed in a subset of neuronal and glial cells in the eyes and head of developing and adult fly among many other tissues. Knockdown of the fly gene in the developing eye resulted in severe small or no eye/head phenotype, and neuronal or glial knockdown caused lethality or functional defects, demonstrating the importance of this gene in eye development and nervous system function. The c.1329dupG variant human GPKOW is also unstable in flies, consistent with the observations in patient cells, but is unlikely to be a strong loss-of-function allele based on functional studies. Gain-of-function or a dominant-negative allele mechanisms are being evaluated in Drosophila. Of note, trio WES, CMA, and RNAseq studies suggested no other molecular etiology for the congenital ichthyosis phenotype. Our work contributes to the understanding of GPKOW's role in eye and brain development in human and flies and suggests a rare variant in *GPKOW* as a cause for a multisystemic X-linked syndrome with primarily brain and eye involvement with a possibility of congenital ichthyosis as part of the clinical presentation.

Session Title: Mendelian Phenotypes Poster Session III

PB4595 A mouse model of gyrate atrophy shows progressive retinal dysfunction and serves as a tool for investigating novel gene editing and gene replacement therapies.

Authors:

J. Johansson, M. Singh, D. Valle, J. J. Doyle; Johns Hopkins, Baltimore, MD

Gyrate Atrophy (GA) is an autosomal recessive retinal degeneration characterized by early onset high myopia and night blindness, onset of cataracts in the second decade, and progressive chorioretinal degeneration often resulting in vision loss by the fourth to fifth decade of life. GA results from biallelic loss of function mutations in *OAT* which encodes a mitochondrial matrix enzyme, OAT, that catalyzes the interconversion of ornithine and pyrroline-5-carboxylate (P5C), an intermediate in proline and glutamate metabolism. Patients with GA have < 5% fibroblast OAT activity and plasma ornithine levels >10 fold above normal. Treatment with an arginine-restricted diet (ARD) reduces plasma ornithine and slows progression of chorioretinal degeneration in mice and humans with GA, but it is a challenging diet to maintain. Hence, alternate mitigation strategies are needed. *Oat*$^{Rhg/Rhg}$ mice harbor a spontaneous missense mutation in *Oat* (p.G353C)(Bisaillon et al., 2014). We confirmed that *Oat*$^{Rhg/Rhg}$ mice display decreased Oat enzyme activity ($p<0.01$) and increased plasma ornithine levels (>20-fold WT littermates). We newly characterized retinal dysfunction in this mouse model using in-vivo electroretinography (ERG). Compared to WT littermates, *Oat*$^{Rhg/Rhg}$ mice show significantly impaired function of both rod photoreceptors (scotopic a-wave; $p<0.0001$) and downstream bipolar cells (scotopic b-wave amplitude; $p<0.01$) by 6 months of age. Both readouts progressively worsen by 12 months in *Oat*$^{Rhg/Rhg}$ mice ($p<0.01$, $p<0.01$, respectively), while WT responses remained stable. In contrast, we saw no sign of cone photoreceptor dysfunction within this time frame in *Oat*$^{Rhg/Rhg}$ mice. These findings are consistent with the early rod dysfunction and later cone involvement seen in humans with GA in the first few decades of life. Histopathological analysis demonstrates morphological alterations in both the photoreceptors and retinal pigment epithelium (RPE) of *Oat*$^{Rhg/Rhg}$ mice compared to WT controls at 12 months. Amino acid analysis confirmed a selective elevation in plasma ornithine in *Oat*$^{Rhg/Rhg}$ mice compared to WT controls ($p<0.00001$), with all other amino acids being equal between genotypes. Cumulatively these data demonstrate that *Oat*$^{Rhg/Rhg}$ mice recapitulate multiple biochemical, electrophysiological and histopathological features seen in humans with GA. Given the nature of the missense mutation in *Oat*$^{Rhg/Rhg}$ mice, the presence of a progressive retinal phenotype and a robust biochemical readout (plasma ornithine levels), the *Oat*$^{Rhg/Rhg}$ mouse model should serve as a valuable resource to interrogate both gene replacement and novel gene editing strategies.

Session Title: Mendelian Phenotypes Poster Session I

PB4596 A multi-omics approach to the characterization of a novel repeat expansion in FAM193B in a family with oculopharyngodistal myopathy

Authors:

C. Reuter[1], S. Fazal[2], T. Jensen[3], J. Gorzynski[4], S. Marwaha[5], D. Bonner[6], J. Kohler[7], H. Mendez[5], I. Voutos[1], S. Emami[1], E. Kravets[5], K. Smith[8], P. Goddard[9], R. Ungar[10], Undiagnosed Diseases Network, M. Danzi[11], J. Sampson[12], J. Bernstein[13], E. Ashley[14], S. Zuchner[15], S. Montgomery[14], M. Wheeler[16]; [1]Stanford, San Francisco, CA, [2]Univ. of Miami, Miami, FL, [3]Stanford, Mountain View, CA, [4]Stanford Univ., San Francisco, CA, [5]Stanford Univ., Palo Alto, CA, [6]Stanford Hlth.Care, Palo Alto, CA, [7]Stanford Med., Palo Alto, CA, [8]Stanford Univ, Stanford, CA, [9]Stanford Univ., Menlo Park, CA, [10]Stanford, Stanford, CA, [11]Univ. of Miami, Dakota Dunes, SD, [12]Stanford Univ., Hayward, CA, [13]Stanford Univ Sch Med, Stanford, CA, [14]Stanford Univ., Stanford, CA, [15]Univ. of Miami Miller Sch. of Med., Miami, FL, [16]Stanford Univ, Palo Alto, CA

Background Oculopharyngodistal myopathy (OPDM) is an autosomal dominant disorder characterized by facial and distal limb weakness, ptosis, and external ophthalmoplegia. Patients may have pharyngeal involvement or myopathic changes on skeletal muscle biopsy. Known molecular mechanisms of OPDMs include trinucleotide repeat expansions in LRP12, GIPC1, RILPL1, and NOTCH2NLC. The disease-causing motif is a 5' UTR CGG expansion. Still, roughly 30% of OPDM patients remain genetically undiagnosed.
Case report At the Stanford Center for Undiagnosed Diseases, we evaluated a 68 year old female with significant weakness of facial muscles, bilateral distal upper and lower extremity weakness, respiratory muscle weakness, osteoporosis, dysphagia, dysarthria, double vision, and hypothyroidism. The patient's sister is similarly affected. Clinical genetic workup including targeted testing for muscular dystrophies, exome sequencing, and short read genome sequencing (SR-GS) were nondiagnostic. Suspecting a short tandem repeat expansion, the patient was nominated for evaluation using an ExpansionHunter Denovo (EHDn) pipeline through the Undiagnosed Diseases Network. No expansions were detected in genes associated with OPDM - LRP12, GIPC1, RILPL1, or NOTCH2NLC. EHDn identified a rare, heterozygous CGG expansion in the 5' UTR of the FAM193B gene from SR-GS estimated at 72 [71-121] repeats. The expansion is present in both affected sisters and the unaffected mother, but absent in father and other unrelated controls. To better characterize the true size of the expansion, nanopore long read genome sequencing (LR-GS) was performed. The patient (245 repeats), sister (225 repeats), unaffected mother (185 repeats) were outliers compared to typical <50 repeats at this locus in a cohort of 100 unrelated individuals with LR-GS. Blood RNA sequencing revealed that FAM193B was over-expressed in both affected sisters compared to 282 unrelated controls. The mother's unaffected status despite carrying a smaller expansion may be explained if hers is a premutation allele.
Conclusions The CGG 5' UTR expansion in FAM193B may represent a novel molecular mechanism of OPDM. While SR-GS detected an expansion at this locus, LR-GS was superior in its ability to accurately characterize the expansion size. The integration of LR-GS and transcriptome analyses inform disease-causing potential of the FAM193B 5'UTR CGG repeat expansion. Further work is ongoing to characterize the functional impact of this expansion on RNA and protein and identify additional affected individuals.

Session Title: Mendelian Phenotypes Poster Session II

PB4597 A new gene discovery approach for structural birth defects employing an innovative machine learning algorithm

Authors:

D. Scott[1], T. M. Scott[2], A. Hardcastle[3], R. Belanger Deloge[4], M. R. Sy[1], T. Gordon[1], E. Huth[4], X. Zhao[4], S. Lalani[1], J. Rosenfeld[4], P. Luna[4], C. Shaw[1]; [1]Baylor Coll. Med., Houston, TX, [2]Univ. of Utah, Salt Lake City, UT, [3]Brigham Young Univ., Provo, UT, [4]Baylor Coll. of Med., Houston, TX

A molecular cause is not identified in most children with structural birth defects. This is due, in part, to an incomplete understanding of the genes that contribute to their development. Genes that are associated with low penetrance in humans may go unrecognized even though they play a critical role in development. Identifying these low-penetrance genes is a prerequisite to accurate genetic counseling for families affected by structural birth defects and enhances the ability of physicians to generate individualized medical plans for affected individuals. Deleterious coding variants in low-penetrance genes are also likely to represent very rare variants with modest to intermediate effect sizes, which are difficult to identify in traditional genetic studies. Such variants may represent important genetic modifiers of penetrance and may significantly contribute to the pathogenesis of structural birth defect cases that are associated with polygenic or multifactorial inheritance. To identify low-penetrance genes, we have developed an alternative gene discovery approach in which candidate genes are prioritized based on their similarity to genes already known to cause a phenotype. This prioritization is accomplished using an innovative machine learning algorithm that leverages data from large-scale genomic knowledge sources to generate phenotype-specific pathogenicity scores for all RefSeq genes. Additional evidence in support of a gene's role in the development of the phenotype is then identified from clinical, research, and public databases, expression profiles, animal models, and isolated case reports. Using this approach, we have implicated several human disease genes in the development of a wide variety of structural birth defects including congenital diaphragmatic hernia (*ALG12*, *ARID1A*, *ARID1B*, *ARID2*, *BRCA2*, *CREBBP*, *DPF2*, *EP300*, *FGFRL1*, *FOXP1*, *SMARCA4*, *SMARCA3*, *SMARCB1*, *SMARCC2*, *SMARCE1*, *UBA2*, *USP9X*), anorectal malformations (*ADNP*, *BBS1*, *CREBBP*, *EP300*, *FANCC*, *KDM6A*, *SETD2*, *SMARCA4*), esophageal atresia/tracheoesophageal fistula (*TCF4*, *NRXN1*, *FANCA*, *FANCB*, *FANCC*), orofacial clefting (*EBF3*, *HUWE1*) and anomalous pulmonary venous return (*EFTUD2*, *NAA15*, *NKX2-1*). These phenotype expansions provide compelling evidence that this approach can be used to identify genes for a wide variety of structural birth defects.

Session Title: Mendelian Phenotypes Poster Session III

PB4598 A new registry for genetic diagnosis of inborn errors of immunity within the military health system

Authors:

J. Villar[1,2,3], N. A. Boggs[4,5], J. A. Brunader[6], A. A. Rutt[6], B. J. Schornack[4], M. Leondaridis[7,8], G. Sukumar[1,9], X. Zhang[10,9], H. Lydia[1,2,7], A. L. Snow[11], C. L. Dalgard[1,12]; [1]Ctr. for Military Precision Hlth., Uniformed Services Univ., Bethesda, MD, [2]Dept. of Pediatrics, Uniformed Services Univ., Bethesda, MD, [3]Henry M Jackson Fundation for the advancement of Military Med., Bethesad, MD, [4]Allergy and Immunology Service, Walter Reed Natl. Military Med. Ctr., Bethesda, MD, [5]Dept. of medicine, Uniformed Services Univ. of Hlth. Sci.s, Bethesda, MD, [6]Defense Hlth.Agency Immunization Hlth.care Div., Falls Church, VA, [7]Henry M Jackson Fundation for the advancement of Military Med., Bethesda, MD, [8]Dept. of Pharmacology & Molecular Therapeutics, Uniformed Services Univ. of the Hlth.Sci., Bethesda, MD, [9]Henry M Jackson Fndn. for the Advancement of Military Med., Bethesda, MD, [10]Ctr. for Military Precision Hlth., Uniformed Services Univ. of the Hlth.Sci., Bethesda, MD, [11]Dept. of Pharmacology & Molecular Therapeutics, Uniformed Services Univ. of the Hlth.Sci.,, Bethesda, MD, [12]Dept. of Anatomy, Physiology & Genetics, Uniformed Services Univ. of the Hlth.Sci., Bethesda, MD

Genotypic definition of monogenic inborn errors of immunity (IEIs) continues to accelerate with broader access to next generation sequencing, underscoring this aggregated group of disorders as a major health burden impacting both civilian and military populations. At an estimated prevalence of 1 in 1200 individuals, IEIs affect ~8,000 patients within the Military Health System (MHS). Despite access to targeted gene/exome panels at military treatment facilities, most affected patients never receive a definitive genetic diagnosis that would significantly improve clinical care. To address this gap, we established the first registry of IEI patients within the MHS with the goal of identifying known and novel pathogenic genetic defects to increase diagnosis rates and enhance clinical care. Using the registry, a research protocol was opened in July 2022. Since July we have enrolled 75 IEI patients encompassing a breadth of phenotypes including severe and recurrent infections, bone marrow failure, autoimmunity/autoinflammation, atopic disease, and malignancy. Enrolled patients provide blood and bone marrow samples for whole genome, ultra-deep targeted panel and comprehensive transcriptome sequencing, plus cryopreservation of peripheral blood mononuclear cells for future functional studies. We are also implementing and developing analytical methods for identifying and interrogating non-coding and structural variants. Variants are classified according to the standards and guidelines for sequence variant interpretation of the American College of Medical Genetics and Genomics. These analyses subsequently inform *in vitro* experiments to validate causative mutations using cell reporter systems and primary patient cells. Clinical variant validation and return of genetic results are planned with genetic counseling provided. As a proof of principle, this integrated genetic evaluation pipeline revealed a novel hemizygous*TLR7* nonsense variant that deletes the last 24 amino acids in two adolescent brothers who both endured critical COVID-19 pneumonia, requiring mechanical ventilation and extracorporeal membrane oxygenation. *In vitro* functional studies demonstrate that this nonsense variant affects protein function. Our protocol is therefore poised to greatly enrich clinical genetics resources available in the MHS for IEI patients, contributing to better diagnosis rates, informed family counseling, and targeted treatments that collectively improve the health and readiness of the military community. Moreover, our efforts should yield new mechanistic insights on immune pathogenesis for a broad variety of known and novel IEIs.

Session Title: Mendelian Phenotypes Poster Session I

PB4599 A Novel Autosomal Dominant Childhood-Onset Disorder Associated with Pathogenic Variants in*VCP*

Authors:

A. Mah-Som[1,2], J. Daw[2], D. Huynh[3], M. Wu[3], B. C. Creekmore[4], W. Burns[5], S. Skinner[5], Ø. L. Holla[6], M. F. Smeland[7], M. Planes[8], K. Uguen[8], S. Redon[8], T. Bierhals[9], T. Scholz[9], J. Denecke[9], M. A. Mensah[10], H. Sczakiel[10], H. Tichy[11], E. Schreiner[11], S. Verheyen[11], J. Blatterer[11], J. Thies[12], C. Lam[13], C. Spaeth[14], L. Pena[14], K. Ramsey[15], V. Narayanan[16], L. Seaver[17], D. Rodriguez[18], A. Afenjar[18], L. Burglen[19], T-F. Chou[3], C. C. Weihl[2], M. S. Shinawi[20]; [1]Havard Med. Genetics Training Program, Boston, MA, [2]Washington Univ. in St. Louis, St. Louis, MO, [3]California Inst. of Technology, Pasadena, CA, [4]Univ. of Pennsylvania Perelman Sch. of Med., Philadelphia, PA, [5]Greenwood Genetic Ctr., Greenwood, SC, [6]Telemark Hosp., Skien, Norway, [7]Univ. of Norway, Tromsø, Norway, [8]CHU de Brest, Brest, France, [9]Univ. Med. Ctr. Hamburg-Eppendorf, Hamburg, Germany, [10]Charité - Univ.smedizin Berlin, Berlin, Germany, [11]Med. Univ. of Graz, Graz, Austria, [12]Seattle Children's Hosp., Seattle, WA, [13]Univ. of Washington, Seattle, WA, [14]Cincinnati Children's Hosp., Cincinnati, OH, [15]TGen, Phoenix, AZ, [16]Translational Genomics Res. Inst., Phoenix, AZ, [17]Corewell Hlth.Helen Devos Children's Hosp., Grand Rapids, MI, [18]APHP-Sorbonne Univ., Paris, France, [19]APHP-Hosp. Trousseau, Paris, France, [20]Washington Univ. in St. Louis, Saint Louis, MO

Valosin-containing protein (VCP) is an AAA+ ATPase that plays critical roles in multiple ubiquitin-dependent cellular processes. Dominant pathogenic variants in *VCP* are associated with adult-onset multisystem proteinopathy (MSP) that presents with myopathy, bone disease, dementia, and/or motor neuron disease. Through GeneMatcher, we identified 13 unrelated individuals who carry novel heterozygous *VCP* variants (12 *de novo*, 1 inherited) associated with a childhood-onset disorder characterized by developmental delay, intellectual disability, hypotonia, and macrocephaly. Trio exome sequencing or multigene panel identified nine missense variants, two in-frame deletions, one frameshift, and one splicing variant. We performed *in vitro* functional studies and *in silico* modelling to investigate the impact of these variants on protein function. In contrast to MSP variants, most missense variants had decreased ATPase activity, and one caused hyperactivation. Other variants were predicted to cause haploinsufficiency, suggesting a loss-of-function mechanism. This is the first description of *VCP*-related neurodevelopmental disease presenting in childhood.

Session Title: Mendelian Phenotypes Poster Session II

PB4600 A Novel Compound Heterozygous Variant in ATAD3A Associated with Neurodegeneration, Mitochondrial DNA Depletion and Epilepsia Partialis Continua

Authors:

M. Muriello, N. D. Patel, N. Mehta, T. Gonzalez, C. Kielczewski, K. L. Jones; Med. Coll. of Wisconsin, Milwaukee, WI

Progressive neurodegenerative disorders are one of the most urgent situations faced by clinical geneticists. While the optimal diagnostic approach is not always clear, evidence is amassing for exome or genome analysis to become the standard of care. Here we present a 3-year-old girl with a history of congenital cataracts, hypotonia, and global developmental delay with a severe neurologic course. Initial genetic evaluation at age 14 months included exome sequencing with copy-number-variant evaluation that was non-diagnostic, and further evaluation was deferred due to adequate developmental progress. At age 3 years she developed left arm jerking and was found to have epilepsia partialis continua (EPC). Brain MRI showed mild cerebellar encephalomalacia. Due to the new clinical concern, exome reanalysis was ordered. While this was pending, the family noted regression of skills with worsening tone, abnormal movements, and loss of self-care skills. She was admitted and found to have choreoathetosis and ataxia and was in refractory status epilepticus that required multiple attempts at burst-suppression. Brain MRI spectroscopy demonstrated interval volume loss, and a small lactate peak in the putamen. Biochemical testing was non-diagnostic, including global metabolomics of plasma and urine. After preliminary exome reanalysis results were non-diagnostic, both rapid whole genome sequencing (rWGS) and a muscle biopsy were performed due to concern for mitochondrial disease, although the EPC and neurodegenerative course were the only red flags at that point. Muscle biopsy found type 2 atrophy and decreased mitochondrial DNA content (45% of control), with negative mtDNA sequencing. rWGS identified biallelic variants in the *ATAD3A* gene; a 35.56 kb deletion which included exons 1-4 and a missense variant c.412C>T, p.Arg138Trp located in exon 4. *ATAD3A* has been associated with a neurodevelopmental disorder, Harel-Yoon syndrome, mainly characterized by developmental delay, congenital cataracts, hypotonia, cerebellar atrophy, and more recently mitochondrial phenotypes. To date, only one patient with *ATAD3A*-related disease has been reported with EPC in the literature. Additionally, we present this patient as the first case with evidence of mtDNA depletion. This patient's diagnostic course demonstrates the utility of rapid genome sequencing, in a patient with negative exome sequencing results, as an earlier result would have saved healthcare costs and improved patient care.

Session Title: Mendelian Phenotypes Poster Session III

PB4601 A novel deep intronic *PHEX* variant associated with X-linked hypophosphatemia in a two-generation Finnish family

Authors:

L. Koljonen[1,2], M. Pekkinen[3,2,4], M. Muurinen[1,2,4], S. Rusanen[2], S. Hussain[1], P. Nevalainen[5], O. Mäkitie[1,2,4,6]; [1]Univ. of Helsinki, Helsinki, Finland, [2]Folkhälsan Res. Ctr., Helsinki, Finland, [3]Univ. of Helsinki, University of Helsinki, Finland, [4]Children's Hosp., Helsinki, Finland, [5]Tampere Univ. Hosp., Tampere, Finland, [6]Dept. of Molecular Med. and Surgery, Karolinska Inst.t, and Clinical Genetics, Stockholm, Sweden

Hypophosphatemic rickets is a rare bone disease characterized by short stature, bone abnormalities, and dental problems. Several genetic forms have been recognized, presenting with hypophosphatemia and hyperphosphaturia that can be dependent or independent of excessive FGF23. Most commonly, hypophosphatemic rickets is due to pathological variants in the X-chromosomal *PHEX* gene, but also autosomal dominant and autosomal recessive forms exist. We investigated a Finnish family in which the son (index, 29 years) and mother (57 years) had hypophosphatemia since childhood. Both had low plasma phosphate (index: 0.56 mmol/L; reference:0.71-1.53 mmol/L), high FGF23 considering hypophosphatemia (index:107 kRU/L, reference:26-110 kRU/L), and high hemoglobin (index:202 g/L; reference:134-167 g/L). Moreover, both had short stature (162 cm and 145 cm, respectively) and curvature of the legs. The index also had dental abscesses and a pathological femur fracture with delayed healing. Both subjects were treated according to conventional XLH protocols. To determine the genetic cause, a gene panel for hypophosphatemic rickets (*OCRL, ALPL, CLCN5, CYP27B1, CYP2R1, DMP1, ENPP1, FAH, FGF23, KL, PHEX, SLC34A1, SLC34A3*, and *VDR*) was performed in two commercial laboratories (Blueprint Genetics, Finland and Center for Genomics and Transcriptomics, Germany). No pathogenic variants were detected. Molecular karyotyping analysis was also normal. No pathogenic copy number variants were found in known disease genes. Whole exome sequencing identified a rare heterozygous missense variant (NM_002646: c.875G>A, p.Arg292His) in the *PIK3C2B* in both affected individuals. Since the gene had not previously been associated with phosphate metabolism, we next performed whole genome sequencing. It revealed a novel deep intronic *PHEX* variant (NM_000444: c.2147+1197A>G) in both subjects. With the current evidence, the variant is classified as a variant of uncertain significance by ACMG classification. According to predictions by SpliceAI and NetGene2, this intronic *PHEX* variant affects splicing. The variant was also predicted to be disease coding by MutationTaster. The effects of the variant on splicing will be evaluated by RT-PCR and sequencing. Based on the clinical characteristics and genetic findings, it is likely that the patients have *PHEX*-mediated X-linked hypophosphatemia. Our findings emphasize the importance of considering intronic variants in XLH pathogenesis.

Session Title: Mendelian Phenotypes Poster Session I

PB4602 A novel FAM92A gene variant further confirms its association with non-syndromic postaxial polydactyly type A9

Authors:

M. Umair, M. Alfadhel; KAIMRC, Riyadh, Saudi Arabia

A novel *FAM92A* gene variant further confirms its association with non-syndromic postaxial polydactyly type A9

ABSTRACT Background: Polydactyly is a very common digit anomaly, having extra digits in hands and/or toes. Non-syndromic polydactyly is caused by disease-causing mutations in several genes, including *GLI3*, *ZNF141*, *IQCE*, *GLI1*, *FAM92A*, *STKLD1*, *KIAA0825*, *MIPOL1*, *DACH1*, and *PITX1*. It can be inherited as an autosomal dominant or autosomal recessive disorder. **Methods:** Whole-exome sequencing (WES) was conducted for the single affected member (II-1) to reveal the disease causative variant/gene. Segregation was performed using bi-directional Sanger sequencing. Homology protein modeling was carried out to determine the effect of mutation on the protein structure, and advanced microscopy imaging approaches were used to reveal the localization of the FAM92A protein at the base of primary cilia. **Results:** A novel homozygous missense variant (p.Ala158Pro) was identified in exon 6 of *the FAM92A* gene. The variant segregated perfectly with the disease phenotype using Sanger sequencing. Furthermore, Insilco analysis revealed that the variant significantly changes the protein secondary structure, which might result in improper or loss-of-function of FAM92A protein. Moreover, siRNA-mediated depletion of FAM92 showed a key role of Fam92 in ciliogenesis and cilia function. **Conclusions:** We report the second disease-causing mutation with further phenotypes in the *FAM92A*, causing autosomal recessive non-syndromic PAPA type 9. This confirms the contribution of *FAM92A* in limb development and patterning.

Session Title: Mendelian Phenotypes Poster Session II

PB4603 A novel *FBN1* intron variant causing isolated ectopia lentis via an in-frame exon skipping

Authors:

N. Shimizu[1,2], Y. Mashimo[3], H. Yokouchi[1], Y. Nishio[4,5], S. Sawai[6,7], T. Ichikawa[7,8], T. Ogi[5,9,10,11], T. Baba[1], Y. Onouchi[3,7]; [1]Dept. of Ophthalmology and Visual Sci., Chiba Univ. Graduate Sch. of Med., Chiba, Japan, [2]Dept. of Ophthalmology, Japanese Red Cross Narita Hosp., Narita, Japan, [3]Dept. of Publ. Hlth., Chiba Univ. Graduate Sch. of Med., Chiba, Japan, [4]Dept. of Pediatrics, Nagoya Univ. Graduate Sch. of Med., Nagoya, Japan, [5]Dept. of Genetics, Res. Inst. of Environmental Med., Nagoya Univ., Nagoya, Japan, [6]Dept. of Neurology, Chiba Aoba Municipal Hosp., Chiba, Japan, [7]Div. of Clinical Genetics, Chiba Univ. Hosp., Chiba, Japan, [8]Dept. of Urology, Chiba Univ. Graduate Sch. of Med., Chiba, Japan, [9]Dept. of Human Genetics and Molecular Biology, Graduate Sch. of Med., Nagoya Univ., Nagoya, Japan, [10]Ctr. for One Med. Innovative Translational Res. (COMIT), Nagoya Univ. Inst. for Advanced Study, Nagoya, Japan, [11]Div. of Molecular Physiology and Dynamics, Inst. for Glyco-core Res. (iGCORE), Tokai Natl. Higher Ed. and Res. System, Nagoya, Japan

Ectopia lentis (EL) is the dislocation or displacement of the eye lens, and cases with no phenotype in organs other than the eyes are termed isolated EL (IEL). EL can occur after trauma or as part of a systemic manifestation of pseudoexfoliation syndrome or as the result of a genetic disorder. *FBN1* gene mutations, known to cause a variety of EL-related systemic diseases such as Marfan syndrome (MFS) and Weill-Marchesani syndrome, have also been reported as a cause of familial IEL. Mutations in *FBN1* cause a range of clinical phenotypes, but the genotype-phenotype relationships underlying these effects remain unclarified. In this study, we performed a genetic study for a Japanese three-generation family of IEL. A whole-exome sequencing on trio, including an affected mother-daughter pair identified no coding variants of genes previously associated with hereditary lens luxation. By contrast, we found a novel intronic variant within intron 11 of *FBN1* that was shared by the two affected individuals. The SNV, c.1327+3A>C was predicted to have high deleteriousness, with a Combined Annotation-Dependent Depletion (CADD) score of 26.1, and to influence the function of the splicing donor, with a delta score of 0.93 for donor loss in SpliceAI. We next recruited two additional affected relatives and confirmed that they also had the variant. In RT-PCR of mRNAs from peripheral blood mononuclear cells, aberrant transcripts lacking exon 11 of *FBN1* were observed in all four affected individuals. Skipping of exon 11 were also observed in transcripts from *FBN1* minigene models corresponding to either the c.1327+3A>C or c.1327+1G>A, which is a variant previously identified variant in two unrelated EL families without MFS manifestations. The exon 11 encodes 60 amino acids, which correspond to the proline-rich region of FBN1 protein and skipping of this exon may results in an in-frame deletion. These results strongly support to consider c.1327+3A>C of FBN1 as the cause of IEL in this family. Further elucidation of the impact of the deletion of the proline-rich region on FBN1 function is expected to provide insights into the pathogenesis of EL in this family as well as genotype-phenotype correlations of *FBN1*-related diseases.

Session Title: Mendelian Phenotypes Poster Session III

PB4604 A novel *FLCN*-related syndrome leading to intellectual disability, immunodeficiency, and leukemia predisposition.

Authors:

C. Maya Gonzalez[1], L. Boeckemeier[2], T. Campbell[3], B. Khoshnood[4], J. Bobeck[5], M. Nordenskjold[1,6], Y. Bryceson[3], A. Lindqvist[2], A. Nordgren[1,6,7,8], F. Taylan[1,6]; [1]Dept. of Molecular Med. and Surgery, Ctr. for Molecular Med., Karolinska Inst., Stockholm, Sweden, [2]Dept. of Cell and Molecular Biology, Karolinska Inst., Stockholm, Sweden, [3]Dept. of Med., Ctr. for Hematology and Regenerative Med., Karolinska Univ. Hosp. Huddinge, Stockholm, Sweden, [4]Dept. of Neurobiology, Div. for Neurogeriatrics, Ctr. for Alzheimer Res., Care Sci. and Society, Karolinska Inst., Stockholm, Sweden, [5]Dept. of Pediatrics, County Hosp., Kalmar, Sweden, [6]Dept. of Clinical Genetics, Karolinska Univ. Hosp., Stockholm, Sweden, [7]Dept. of Lab. Med., Inst. of Biomedicine, Univ. of Gothenburg, Gothenburg, Sweden, [8]Dept. of Clinical Genetics and Genomics, Sahlgrenska Univ. Hosp., Gothenburg, Sweden

Background: *FLCN*, the disease-causing gene for autosomal dominant Birt-Hogg-Dubé syndrome, encodes a protein with multiple cellular roles, including the regulation of early embryonic development and metabolic modulation through the mTORC1 pathway and transcription factors TFE3/TFEB. Recently, an interactor of FLCN, *FNIP1* has been associated with an autosomal recessive immunodeficiency and cardiomyopathy syndrome, while mosaic variants in *TFE3* lead to a disorder characterized by intellectual disability and growth retardation. Here we present a 14-year-old boy with intellectual disability, short stature, coarse facial features, and immunodeficiency, who developed acute lymphoblastic leukemia at 1 year of age. In the patient, we detected a germline homozygous *FLCN* variant and aimed to characterize the genetic and molecular mechanisms leading to this novel syndrome. **Methods:** We carried out 30X Whole Genome Sequencing (WGS), Sanger sequencing, and flow cytometry from blood samples to confirm the variant's inheritance and evaluate the patient's immunological parameters. Additionally, to investigate the effect of the variant in FLCN-mediated metabolic modulation, we used droplet-digital PCR (ddPCR), western blot (WB) and immunofluorescence (IF) staining on skin fibroblasts. **Results:** WGS analysis in the patient identified a homozygous ultra-rare, and likely damaging missense variant in *FLCN* (p.G15S). The variant was validated by Sanger sequencing, which showed parental inheritance. Flow cytometry results confirmed hypogammaglobinemia (IgM, IgG), and low B- and NK-cells' counts. We did not observe changes in the activation of the mTORC1 pathway in patient's fibroblasts. However, IF staining revealed a significant increase in TFE3 nuclear translocation in the patient, resulting in the transcriptional activation of genes related with mitochondrial biogenesis, glycolysis, lysosomal biogenesis, and nucleic acid metabolism, as previously observed in *FLCN*-deficient animal models. **Conclusion:** Our results suggest that homozygous *FLCN* p.G15S affects TFE3-mediated transcriptional activation of genes involved in cell metabolism. We hypothesize that the variant is hypomorphic, leading to a novel autosomal recessive *FLCN*-related syndrome, characterized by leukemia predisposition, immunodeficiency, and intellectual disability. Interestingly, the phenotype in this syndrome overlaps with *FLCN*-deficient animal models, mosaic *TFE3* intellectual disability syndrome, and autosomal recessive *FNIP1* immunodeficiency.

Session Title: Mendelian Phenotypes Poster Session I

PB4605 A novel hemizygous variant in *FGD1* in a male patient with X-linked Aarskog-Scott syndrome.

Authors:

I. Showpnil[1], S. Ramadesikan[1], M. Marhabaie[1], A. Daley[1], M. Pastore[2], J. Hunter[1], K. McBride[2,3,4], K. Truxal[2,3], D. Koboldt[1,3], R. Wilson[1,3], R. Stottmann[1,3]; [1]The Steve and Cindy Rasmussen Inst. for Genomic Med., Nationwide Children's Hosp., Columbus, OH, [2]Div. of Genetic and Genomic Med., Nationwide Children's Hosp., Columbus, OH, [3]Dept. of Pediatrics, The Ohio State Univ. Coll. of Med., Columbus, OH, [4]Univ. of Calgary, Calgary, AB, Canada

Craniofacial anomalies are common congenital malformations that can occur as part of a syndromic condition or in isolation. Next-generation sequencing methodologies paired with basic research studies using animal models offer a powerful interdisciplinary approach to identify new genetic causes of craniofacial conditions. In this study, we describe a 6-year-old male who was small for gestational age and presented with short stature (<1 percentile), dysmorphic features that included cleft lip/palate, growth and developmental delays, gastroesophageal reflux disease, chronic otitis media, and eustachian tube dysfunction. Trio whole genome sequencing of the proband and his unaffected parents revealed a maternally inherited hemizygous variant in faciogenital dysplasia 1 (*FGD1*) (c.2581G>T: p.Asp861Tyr).

FGD1 encodes a guanine nucleotide exchange factor that specifically binds and activates the GTPase CDC42 by catalyzing the exchange of inactive GDP for active GTP. This leads to activation of signaling pathways that are critical for development. Pathogenic variants in *FGD1* are associated with Aarskog-Scott syndrome (MIM #305400), an X-linked recessive disorder characterized by short stature, dysmorphic facies, cleft lip/palate, shawl scrotum, and brachydactyly. The disorder exhibits wide phenotypic variability, even within families. The variant identified in our patient is absent from public databases and maps to the first nucleotide of exon 18 causing a missense change at an evolutionarily conserved residue (p.Asp861Tyr). This change is predicted to be damaging by 15 out of 17 in silico pathogenicity prediction tools and has a CADD score of 33.0. Gene constraint metrics show that the *FGD1* transcript is highly intolerant to both missense (Z=3.52) and loss of function (pLI=1.0) variants (gnomAD). In retrospect, our patient exhibits many of the classical features of Aarskog-Scott syndrome, which was considered in the clinical differential.

Pathogenic *FGD1* variants were recently identified in patients with Robinow syndrome (MIM#180700). In fact, the same study showed a high degree of phenotypic similarity between Aarskog-Scott and Robinow Syndromes, suggesting that FGD1 may play a role in the non-canonical WNT/ planar cell polarity signaling pathway, perturbations of which are associated with Robinow syndrome. Our finding therefore provides a long-sought molecular diagnosis for this patient and adds to this expanding phenotypic spectrum of *FGD1* associated anomalies.

Session Title: Mendelian Phenotypes Poster Session II

PB4606 A novel likely pathogenic *FBN1* variant c.2686T>A, p.(Cys896Ser) associated with Marfan syndrome.

Authors:

M. Ben Amor; CHUDUMONT, Dieppe, NB, Canada

Background. Fibrillin-1 (FBN1) is a major structural component of the extracellular matrix, providing strength and stability to tissues. Pathogenic variants lead to the development of *FBN1*-associated syndromes which comprise a broad host of phenotypes, and more commonly, Marfan syndrome (MFS). MFS is typically diagnosed in patients presenting with ectopia lentis, thoracic or aortic disease, and skeletal features, which may prompt genetic testing. A 21-year-old Caucasian female was evaluated at the New Brunswick medical genetics clinic upon referral from an ophthalmologist regarding ocular anomalies. More specifically, bilateral superior ectopia lentis for which the patient was treated with lensectomy. **Objective**. The objective was to identify the genetic underpinning for the clinical presentation as well as the mode of inheritance. **Methods**. Blood samples were collected from the family with written informed consent. DNA extraction, next generation sequencing, and data analysis were performed in a targeted panel to identify the causal gene and mode of inheritance. **Results**. Herein, we describe the reclassification of a newly identified heterozygous *FBN1* variant, c.2686T>A, p.(Cys896Ser), to likely-pathogenic in a female patient presenting with abnormal superior bilateral ectopia lentis, and mild thoracolumbar scoliosis. Identification of this variant led to cascade testing in the patient's 49-year-old mother. The latter has a history of isolated congenital ectopia lentis and was found to be a carrier of the same *FBN1* variant, thus prompting standard treatment and management as MFS which identified an aortic dilatation. The segregation of the phenotype in both patient and mother, the family member testing, the abnormal aorta in the mother, the variant's absence in control populations, and all in silico tools predicting pathogenicity led to the reclassification of this *FBN1* variant to likely-pathogenic. **Discussion**. The reclassification will contribute to timely actionable interventions earlier in the natural history of MFS and other *FBN1*-associated syndromes.

Session Title: Mendelian Phenotypes Poster Session III

PB4607 A novel missense variant in *CLPP* gene causing Perrault syndrome in a large Sudanese family

Authors:

A. Babai[1], L. Elsayed[2], I. Mohammed[3], A. Hamed[3], M. Elseed[1]; [1]Univ. of Khartoum, Khartoum, Sudan, [2]Princess Nourah bint Abdulrahman Univ., Riyadh, Saudi Arabia, [3]Faculty of Med., Univ. of Khartoum, Khartoum, Sudan

Perrault syndrome is a rare autosomal recessive condition characterised by sensory neural hearing loss, ovarian dysfunction and neurological features like cerebellar ataxia, sensory motor neuropathy and learning difficulty.
In this study, we report a family with five siblings (four boys and one girl) from a large consanguineous Sudanese family. All of the five siblings presented with sensory neural hearing loss, ataxia, spasticity and learning disability, all within variable stages of severity. The brain MRI showed features of white matter demyelinating changes and brain atrophy. We used whole exome sequencing to investigate the underlying culprit genetic cause of their condition. We were able to identify a novel homozygous missense variant in *CLPP* gene (c.1634C>T) which is predicted to be pathogenic following the ACMG classification and using in silico bioinformatics tools and confirmatory family segregation studies.
This is the first Sudanese family to be diagnosed with Perrault syndrome. We were able to add to the clinical phenotype spectrum of this rare condition and enrich the genetic pool for the involved genetic variants. Our case report also highlights the importance of exploring the African population genomics to reach a better understanding of the rare monogenic neurogenetic diseases.

Session Title: Mendelian Phenotypes Poster Session I

PB4608 A novel SHANK3 missense mutation in a child with autism spectrum disorder, motor delay, and hypotonia.

Authors:

P. Gordon[1], S. R. Katz[2], B. Smith-Packard[1]; [1]Penn State Hlth.Children's Hosp., Hershey, PA, [2]Dept. of Pediatrics, Cincinnati Children's Hosp. Med. Ctr., Cincinnati, OH

Deficiency of the neuronal synapse scaffolding protein *SHANK3* has been implicated in neurodevelopmental disorders including autism spectrum disorder and *SHANK3*-related Phelan-McDermid spectrum disorder (OMIM #606232). Although the majority of cases of Phelan-McDermid Syndrome involve deletions of all or part of *SHANK3*, other pathogenic variants within *SHANK3* have been found to cause both Phelan-McDermid Syndrome and autism spectrum disorder. We report the case of a 4.5-year-old phenotypic female with autism spectrum disorder, generalized hypotonia, and motor-specific developmental delay. Genetic testing utilizing the GeneDX Autism/ID Xpanded panel revealed a heterozygous single nucleotide change of c.5170 G>A in the *SHANK3* gene (NM_033517.1). This change causes a missense mutation in the *SHANK3* gene product of an Ala1799Thr substitution in the sterile alpha motif (SAM) domain of the *SHANK3* protein, a region reportedly critical for *SHANK3* polymerization and localization to the postsynaptic density. Thus, in silico analyses of the patient's mutation understandably predict disruption of the protein function and likely pathogenicity. Although the patient shares some features with Phelan-McDermid Syndrome, including autism spectrum disorder, hypotonia, motor delays, and motor planning deficits, she does not have the expected most striking features of absent to severely delayed speech and moderate to profound intellectual disability. She has no signs of intellectual disability and exhibits above-average for age expressive language skills. This case serves to further our understanding of the spectrum of expression of pathogenic *SHANK3* mutations, particularly point mutations not predicted to cause protein truncation, and contributes to our understanding of the range of presentation and phenotypic variability for this rare disorder.

Session Title: Mendelian Phenotypes Poster Session II

PB4609 A novel type of Benign Adult Familial Myoclonic Epilepsy (BAFME) in a large Malian family

Authors:

A. Maiga[1], P. Yeetong[2], M. Dembélé[1], M. Pongpanich[3,4], L. Cissé[5], C. Srichomthong[6,7], K. Dembélé[8], A. Assawapitaksakul[6,7], S. Bamba[1], A. Yalcouyé[1], S. Diarra[1,9,10], S. Mefoung[1], S. Rakwongkhachon[6,7], O. Traoré[1], S. Tongkobpetch[6,7], K. Fischbeck[10], W. Gahl[11], C. Guinto[1,5], V. Shotelersuk[6,7], G. Landoure[1,5]; [1]Univ. of Sci., Techniques and Technologies of Bamako, Bamako, Mali, [2]Div. of Human Genetics, Dept. of Botany, Faculty of Sci., Chulalongkorn Univ., Bangkok, Thailand, [3]Dept. of Mathematics and Computer Sci., Faculty of Sci., Chulalongkorn Univ., Bangkok, Thailand, [4]Omics Sci. and Bioinformatics Ctr., Faculty of Sci., Chulalongkorn Univ., Bangkok, Thailand, [5]Service de Neurologie, Ctr. Hosp.ier Univ.ire du Point G, Bamako, Mali, [6]Ctr. of Excellence for Med. Genomics, Dept. of Pediatrics, Faculty of Med., Chulalongkorn Univ., Bangkok, Thailand, [7]Excellence Ctr. for Genomics and Precision Med., King Chulalongkorn Mem. Hosp., the Thai Red Cross Society, Bangkok, Thailand, [8]Hôpital du District de la Commune IV, Bamako, Mali, [9]Yale Univ., Pediatric Genomics Discovery Program, Dept. of Pediatrics, New Haven, CT, [10]Neurogenetics Branch, NINDS, NIH, Bethesda, MD, [11]Natl. Human Genome Res. Inst., NIH, Bethesda, MD

Introduction: Benign adult familial myoclonus epilepsy (BAFME), also known by numerous other names including autosomal dominant cortical myoclonus and epilepsy (ADCME), familial adult myoclonic epilepsy (FAME), and familial cortical myoclonic tremor with epilepsy (FCMTE), is an autosomal dominant disorder characterized by cortical tremors, myoclonus, and infrequent seizures, which progress slowly. Six types of BAFME, all caused by pentanucleotide repeat expansions, have been previously reported. However, their pathomechanism remains unclear. Here, we report a large Malian family with 10 members affected by BAFME with a novel pattern of repeat expansion and insertion.
Aim: To characterize both clinically and genetically a large Malian family with late-onset myoclonic tremor.
Methods: 27 individuals from a large Bambara family were enrolled in our protocol after giving full and informed consent. In addition to a thorough neurological examination, some laboratory works have been performed in available patients including brain imaging, blood chemistry, EEG and Nerve Conduction Study (NCS) in two siblings with neuropathic features. For genetic testing, genomic DNA and RNA were extracted for Long-read sequencing (LRS) and Expression studies. Whole genome sequencing using PacBio's long-read HIFI technique was performed in the 10 affected individuals. In addition, 200 ethnically matched samples were used as controls.
Result: In total,10 individuals were found to have symptoms consistent with BAFME. In addition, two sisters from consanguineous parents also had clinical and electrical features of peripheral Neuropathy. LRS identified TTTTA repeat expansions and TTTCA repeat insertions in intron 4 of the *RAI1* gene that co-segregated with the disease status. The two siblings with neuropathy had a homozygous single nucleotide substitution (c.1A>G) in *PEX10*, leading to initiator codon loss (p.Met1?).
Conclusion: We established a new type of BAFME, BAFME8, in the first African family. Interestingly, this is the only BAFME type where the repeats are in a gene previously associated with a single gene disorder, Smith-Magenis syndrome, which is caused by RAI1 haploinsufficiency. Moreover, Biallelic pathogenic variants in *PEX10* are known to cause the peroxisome biogenesis disorders, suggesting that further studies might be needed in the branch of the family with Neuropathic features.

Session Title: Mendelian Phenotypes Poster Session III

PB4610 A patient with *ATOH7* variants with Delayed Sleep-Wake Phase Disorder and Optic Nerve Hypoplasia.

Authors:

A. Kaden, J. Brzezynski, C. Johnson, A. Goldberg, C. Miller, S. Smieszek, C. Xiao, C. Polymeropoulos, G. Birznieks, M. Polymeropoulos; Vanda Pharmaceuticals Inc., Washington, DC

We are conducting a double-blind, randomized study to evaluate the effects of tasimelteon versus placebo in Delayed Sleep-Wake Phase Disorder (DSWPD) patients with extensive clinical phenotyping. We report our first completed patient from the 11-month Open-Label Extension (OLE), who has a history of Optic Nerve Hypoplasia (ONH). The study consists of screening and treatment phases, followed by an OLE to explore the long-term safety and efficacy of daily dosing with tasimelteon over 11 months. During the OLE, patients answer daily sleep diaries and are instructed to take one dose of tasimelteon 1 hour before their desired bedtime. Patients determine their desired bedtime as the time they would need to go to bed the night before a commitment in order to feel fully rested in the morning. As this study is currently ongoing, conclusions cannot be made about the blinded treatment phase. We present a case of a 24-year-old female diagnosed with DSWPD and ONH, with a confirmed delayed Dim Light Melatonin Onset, who reports the inability to fall asleep at their desired bedtime and the ability to have a full night's sleep when not required to be up at a specific time for societal requirements. After completing the blinded treatment phase of the study, they opted into the OLE. The patient's average sleep onset was 01:27 during screening and 00:42 during the OLE, an improvement of 45 minutes. At screening, the patient reported their symptoms as moderate on the Patient Global Impression of Severity (PGI-S). On average during the OLE, the patient reported their symptoms as mild on the PGI-S and as much improved on the Patient Global Impression of Change. The patient reported a history of ONH, which could indicate lack of proper optic response to light. Abnormal rest-activity rhythmicity patterns are present in 30% of children with ONH. Further research is necessary to characterize the relationship between light response and sleep in this individual. This patient carries two variants in the Atonal BHLH Transcription Factor 7 (*ATOH7*) gene, rs61854782 and rs7916697, known to be associated with ONH. They have a variable number of tandem repeat (VNTR) $PER3^{5/5}$ genotype, associated with normal sleep patterns, and no identified predicted loss-of-function mutations within other circadian genes. This case illustrates the general positive effect of tasimelteon on a patient diagnosed with DSWPD, based on an earlier sleep onset shift and overall improvement in PGI-S responses during OLE compared to screening. It also provides the opportunity for research into ONH and genetics, as well as the relationship with DSWPD and potentially sighted Non-24-Hour Sleep-Wake Disorder.

Session Title: Mendelian Phenotypes Poster Session I

PB4611 † A phenome-wide association study of methylated GC-rich repeats identifies a GCC repeat expansion in *AFF3* as a significant cause of intellectual disability.

Authors:

B. Jadhav[1], P. Garg[1], J. J. F. A. van Vugt[2], K. Ibanez[3], D. Gagliardi[4], W. Lee[1], A. Martin Trujillo[1], M. Barbosa[1], M. Jain[5], H. Houlden[6], B. Paten[5], Genomics England Research Consortium, Project MinE ALS Sequencing Consortium, J. Veldink[2], A. Tucci[3], A. Sharp[1]; [1]Icahn Sch. of Med. at Mount Sinai, New York, NY, [2]Utrecht Univ., Utrecht, Netherlands, [3]Queen Mary Univ. of London, London, United Kingdom, [4]Queen Mary Univ. of London, Univ. Coll. London, London, United Kingdom, [5]Univ. of California, Santa Cruz, CA, [6]Univ. Coll. London, London, United Kingdom

GC-rich tandem repeat expansions (TREs) are often associated with DNA hypermethylation, gene silencing and folate-sensitive fragile sites and underlie several congenital and late-onset disorders, including fragile X and ALS. Through a combination of DNA methylation profiling, TR genotyping and long read sequencing, we identified 24 loci where GC-rich TREs cause local DNA hypermethylation. We hypothesized that these TREs represent strong candidates for human disease. To investigate this, we generated TR genotypes and performed phenome-wide association studies in 168,641 individuals from the UK Biobank, identifying 156 significant TRE:trait associations (10% FDR) involving 17 different TREs. In line with its causal role in ALS, all of the traits associated with *C9orf72* expansions were consistent with the known symptoms and treatments of ALS, such as reduced brain volume, increased incidences of dysphagia, neuroticism, brain imaging studies and death attributed to motor neuron disease. However, we also identified 16 other TREs that showed significant associations with diverse phenotypes, including blood cell traits, alcohol consumption, altered microstructure of brain white matter and reductions in cerebellar volume, lung function, height, and overall health. Notably, a GCC repeat expansion in the promoter of *AFF3* was linked with a 2.4-fold reduced probability of completing secondary education ($p=9.4x10^{-11}$), an effect size comparable to several recurrent microdeletion disorders. Given its strong negative influence on educational attainment, we hypothesized that TREs of *AFF3* might represent the pathogenic mutation in some patients with neurodevelopmental disorders. In support of this, in a cohort of 6,371 probands with neurodevelopmental problems of unknown etiology from the 100,000 Genomes Project, we observed a significant enrichment (p=0.002) for expanded *AFF3* alleles compared to controls. With an estimated population prevalence of ~1 per 850, which is 5- to 10-fold higher than the TRE that causes fragile X syndrome and 3- to 8-fold higher than recurrent microdeletions of 1q21.1, 15q13.3 and 16p11.2, *AFF3* expansions therefore represent a common and clinically relevant cause of intellectual disability. Finally, we identified SNVs located within the *AFF3* locus that associate with presence of the TRE and observed that many of these are the same SNVs at the *AFF3* locus that have been reported in prior GWAS for educational attainment. Our observations therefore indicate that a subset of GWAS signals in the genome are likely driven by underlying TREs that preferentially occur on specific founder haplotypes.

Session Title: Mendelian Phenotypes Poster Session II

PB4612 A proposal of mechanism of non-syndromic deafness associated with a novel *BCAP31* variant: mitochondrial dysfunction and high sensitivity to cisplatin

Authors:

B. Kim[1], Y. Kim[2], Y. Kim[3], S-H. Yu[3], J. Han[2], M. Kim[2], L. Seoeun[3], Y. Kang[3], C-H. Kim[3], B. Choi[2]; [1]Chungnam Natl. Univ. Coll. of Med., Sejong, Korea, Republic of, [2]Seoul Natl. Univ. Bundang Hosp., Seongnam, Korea, Republic of, [3]Paean Biotechnology, Seoul, Korea, Republic of

B-cell receptor-associated protein 31 (BAP31 or BCAP31) is an integral ER membrane protein, which involves the transport and quality control of transmembrane proteins. BCAP31 is also important for the cross-talk of apoptotic signals between the ER and mitochondria. Due to its critical role in cellular physiology, the BCAP31 dysfunction has been associated with numerous human diseases including deafness, dystonia, and central hypomyelination (DDCH) syndrome, cancer, metabolic syndrome, cystic fibrosis, and neurodegenerative diseases. Recently, we have found a novel in-frame insertion variant in the *BCAP31* gene from a family segregating only non-syndromic hearing loss in an X-linked, recessive fashion. It is not known how this variant contributes to the hearing loss, which is an important issue to understand the molecular pathogenesis. To address it, we compared the mitochondrial function between the patient-derived lymphoblastoid cell lines (LCLs) and normal LCLs. The patient-derived LCLs showed the elevation in ROS, and the decrease in ATP and membrane potential intracellularly compared to normal LCLs. Surprisingly, the administration of mitochondria (PN-101) isolated from umbilical cord mesenchymal stem cells (UC-MSC) was able to rescue the mitochondrial dysfunction in the patient-derived LCLs. Furthermore, patient-derived LCLs demonstrated more pronounced cisplatin-induced cell death than did normal LCLs by confirming the increase in the expression of pro-apoptotic genes. Taken together, the novel *BCAP31* variant may contribute to the pathogenesis of the impaired hearing due to the mitochondrial dysfunction.

Session Title: Mendelian Phenotypes Poster Session III

PB4613 A rare form of myopathy, with extrapyramidal signs, in a 14-year-old female: case report.

Authors:

S. Ahmed[1], J. Black[2], R. Lebel[2], N. Dosa[2], K. Scott[3]; [1]Dept. of Pediatrics, SUNY Upstate Med. Univ., Syracuse, NY, [2]Ctr. for Dev., Behavior and Genetics, SUNY Upstate Med. Univ., Syracuse, NY, [3]Dept. of Pediatric Hematology/Oncology, SUNY Upstate Med. Univ., Syracuse, NY

Myopathy with extrapyramidal signs (MPXPS) is a disorder characterized by developmental delay (motor and speech) and learning disability, which varies in severity. It is characterized by early childhood onset of non-progressive proximal muscle weakness. Laboratory findings include a significantly elevated CK level, elevated transaminases, and myopathic features on muscle biopsy. Most patients develop extrapyramidal signs which include chorea, tremor, dystonic posturing, and orofacial dyskinesia. These signs are often progressive, and combined with muscle weakness may result in disabling immobility. Other features may include ataxia, ophthalmoplegia, optic atrophy, peripheral axonal neuropathy, ptosis, and seizures. Less frequently reported features include poor growth/short stature, ichthyosis, atopic dermatitis, hepatomegaly, ventricular septal defect, amblyopia, sleep disorder, and hemolytic anemia. This disorder is caused by pathogenic variants in the *MICU1* gene (OMIM #605084), following an autosomal recessive pattern. To our knowledge, there are 30 cases in the literature, many associated with consanguineous parentage.

We report a 14-year-old female born by vaginal delivery at 38 weeks gestation to a 32-year-old primigravid European mother, and father of unknown age. Union is not known to be consanguineous. This patient presented in adolescence with developmental and intellectual disability, speech apraxia, dysmenorrhea, dysphagia, symptomatic erythrocytosis, atopic dermatitis, spasticity, dystonia, and dysmorphic features. She was diagnosed with MPXPS when whole exome sequencing revealed compound heterozygous pathogenic variants in the *MICU1* gene, likely in trans: c.161+1G>A, maternally inherited, and c.386G>C(p.R129P), of uncertain origin (since father did not participate).

The literature reports one other female with this exact genotype, who presented at age 12 with developmental delay, myopathy, ataxia, amblyopia, hyperreflexia, clonus, seizures, and dysmorphic features. Given the rarity of this condition, and the fact that there is another patient with an identical genotype reported, we offer this patient's unique feature (erythrocytosis), as an extension of the phenotypic spectrum.

Session Title: Mendelian Phenotypes Poster Session I

PB4614 A *RING1* contribution to neurogenesis and schizophrenia

Authors:

C. Ryan; Univ. of Michigan, Ann Arbor, MI

Intellectual disability affects 2-3% of the population in industrialized nations, with genetic causes accounting for approximately half of these cases. Whole genome sequencing has led to the discovery of several novel pathogenic variants underlying neurodevelopmental disorders. Among these variants, proteins involved in chromatin post-translational modifications are highly represented. Variants in the polycomb repressive complex 1 (PRC1), the primary catalyst of histone 2A monoubiquitination (H2AUb1), illustrate this trend. RING1 and RNF2 are paralogues that serve as the constituent E3 ubiquitin ligase of PRC1, and missense variants in these genes that blunt H2AUb1 catalysis are associated with a neurodevelopmental disorder characterized by microcephaly, intellectual disability, and early onset schizophrenia. To explore the role of PRC1- dependent H2AUb1 in the context of human corticogenesis, we have generated isogenic human embryonic stem cell (hESC) lines harboring pathogenic variants in *RING1* and subjected them to neural differentiation to generate 3D cerebral organoids. Previous studies suggest that PRC1-dependent H2AUb1 has roles in transcriptional repression and double strand break repair. Consistent with these previous reports, we demonstrate that the $RING1^{G284A/G284A}$ variant diminishes global H2AUb1 levels, leads to increased expression of DNA damage repair genes, and disrupts double strand break repair in neural progenitor cells (NPCs). These defects lead to stalled cell cycle progression due to a prolonged S phase. This work implicates RING1-dependent H2AUb1 in the maintenance of genome integrity during human corticogenesis.

Session Title: Mendelian Phenotypes Poster Session II

PB4615 A study of the natural history of *RARB*-related disorders.

Authors:

V. Chu, A. Tremblay, J. L. Michaud; CHU Sainte-Justine Res. Ctr., Montreal, QC, Canada

Gain-of-function and dominant-negative variants in the Retinoic Acid Receptor Beta (*RARB*) gene cause a syndrome known as MCOPS12 that is characterized by the presence of developmental eye defects, other congenital anomalies, and global developmental delay with dystonia. Although the majority of affected individuals display microphthalmia as well as severe developmental delay and motor impairment, MCOPS12 is variable as cardinal features of the syndrome such as eye anomalies and neurodevelopmental involvement can be absent or minimal in some affected individuals. Moreover, it is not possible to predict at an early stage the clinical course of affected individuals. In order to gain insight into the natural history of MCOPS12, we initiated a prospective study of its course in partnership with patient organizations. Individuals carrying pathogenic or likely pathogenic variants in *RARB* are eligible to participate in the study. In addition to the baseline phenotype of the participants, developmental milestones (Vineland-3, ASQ-3), behavioral profiles (SRS-2, BASC, Sensory Profile 2) and severity of dystonia (Burke-Fahn-Marsden Dystonia Rating Scale, the Hypertonia Assessment Tool scoring chart and the Global Dystonia Severity Rating Scale) are documented on a yearly basis using standardized tools. Caregivers are invited to provide videos of the participants using a standardized protocol for the study of movement disorders. Brain MRI images generated on a clinical basis are also collected. The study framework includes a physician-driven arm and a participant-powered arm, the latter of which involves interviews with the research team to complete some of the assessment tools. For both arms, data are collected via a REDCap-based platform. Consent to participate in the study is also obtained using this platform. We will use *in vitro* transcriptional assays to determine whether novel variants induce gain-of-function or dominant-negative effects. The study is being primed by a series of 43 participants who were previously reported by our group (Srour et al., 2013, Srour et al., 2016, Caron et al., 2023). Patient organisations Cure MCOPS12 and A Cure for Sienna have participated in the development of the study and will provide feedback as it rolls out. As therapeutic options for MCOPS12 are being actively investigated, it is imperative to document the natural history of individuals with this condition. Our study will lead to the establishment of a framework for the development of clinical trials examining MCOPS12.

Session Title: Mendelian Phenotypes Poster Session III

PB4616 A targeted NGS panel with a 20% diagnostic yield in Short Stature reveals the contribution of digenic/oligogenic inheritance in growth disorders disclosing novel potential gene interactions

Authors:

M. Giordano[1], D. Vurchio[1], S. Mellone[2]; [1]Univ. of Eastern Piedmont, Novara, Italy, [2]Ospedale Maggiore della Carità, Novara, Italy

Short stature is one of the most common endocrinological conditions of childhood that occur isolated or in presence of complex phenotypes. To date monogenic causes have been identified in more than 200 mendelian syndromes and skeletal dysplasia. The diagnostic yield of a 106 gene NGS panel was evaluated in a cohort of 302 patients with idiopathic and syndromic short stature. Pathogenic/Likely pathogenic (P/LP) variants were identified in 61 patients representing the 20% of cases. Twenty-nine genes were mutated in at least one patient and 12 genes in more than one, namely *ACAN, KISSIR, GNAS, NF1, SLC26A2, GLI2, PTPN11, GNRHR, SOS, HEXS1, NPR2, COMP*. Interestingly 9 patients carried a VUS in addition to the P/LP variant and three patients carried 2 pathogenic variants suggesting digenic inheritance involving the tested genes. Among these, patient #1006 carried variants in *ACAN* and *FBN1*. From the IPA Ingenuity database, we extracted a predictive indirect gene interaction *ACAN-FBN1* mediated by FN1 a dimeric glycoprotein of the cartilage extracellular matrix. Patient # 2553 carried variants in *SOS1* and *RAF1*, both belonging to the RAS-MAPK pathway. Patient # 1698 carried pathogenic variants in *SHOX* and *COMP* suggestive of interaction between these two chondrogenesis proteins. In conclusion digenic/oligogenic inheritance should be taken into account in deciphering the genetic basis of some forms of short stature, also considering those VUS, that in presence of a P/LP variants are usually underestimated. Besides the relevance for the diagnosis this might lead to highlight novel interactions involved in the complex process of growth.

Session Title: Mendelian Phenotypes Poster Session I

PB4617 A trial of resveratrol therapy in a patient with *CMIP* haploinsufficiency and autism spectrum disorder.

Authors:

C. Hartlage[1], M. K. Diamreyan[1], J. Owens[2], A. Shillington[2]; [1]Univ. of Cincinnati, Cincinnati, OH, [2]Cincinnati Children's Hosp., Cincinnati, OH

Autism spectrum disorder (ASD) is a neurodevelopmental disorder characterized by difficulties with social communication and restricted behaviors. It can be caused by environmental or genetic factors, including the gene *CMIP*. *CMIP* is a scaffolding protein involved in brain development and negative regulation of T cells. Previously reported cases have attributed ASD to *CMIP* loss of function, theorizing a connection to neuroinflammation. Resveratrol, a polyphenol with anti-inflammatory properties, has demonstrated some success as a treatment for ASD symptoms. This report presents a 15-year-old male with ASD and a de-novo *CMIP* variant who showed clinical improvement with resveratrol supplementation.

The patient has a history of ASD, developmental delays, speech and language deficits, generalized anxiety disorder, ADHD, intermittent explosive disorder, constipation, and pancreatitis. He was diagnosed with ASD at age 4. Simultaneous developmental testing assessed his receptive and expressive language to be further delayed than visual perception and fine motor skills. He has a history of recurrent psychiatric admissions due to overactivity and impulsivity with concerns for self-injury. While admitted to inpatient psychiatry at age 12, an ASD-focused gene panel revealed a c.1483G>A (p.Glu495Lys) *CMIP* variant interpreted as contributory. The patient was initiated on resveratrol after his *CMIP* diagnosis at age 13. After 18 months, the patient avoided further admissions and showed improved impulsivity and aggression according to reports from parents. Objective behavioral assessments showed reduced scores compared to before initiation of therapy.

By comparing our patient to reported cases, we demonstrate *CMIP* haploinsufficiency is characterized by ASD, developmental delay, speech and language deficits, hypotonia, gastroesophageal reflux, and feeding intolerance. Our patient’s expressive and receptive language deficits are greater than his gross motor delays; this is supportive of linkage studies that identified *CMIP* SNPs as associated with specific language impairment and general reading ability. Furthermore, the patient responded positively to resveratrol treatment with decreased impulsivity and non-compliant behaviors. While likely multifactorial, we hypothesize that resveratrol's anti-inflammatory properties reduced neuroinflammation, leading to an improvement in ASD and ADHD symptoms. The patient's case helps establish the symptoms associated with *CMIP* haploinsufficiency and demonstrates that resveratrol may help impulsive behaviors in *CMIP*-associated ASD.

Session Title: Mendelian Phenotypes Poster Session II

PB4618 A variant in mitochondrial protein import complex leads to microcephalic osteodysplastic dwarfism with moyamoya disease

Authors:

C. Li[1], P-X. Lin[1], Y-C. Wang[1], L-W. Chen[2], M-H. Shih[3], M-C. Tsai[2], Y-Y. Chou[2], P-C. Chen[1]; [1]Inst. of Clinical Med., Coll. of Med., Natl. Cheng Kung Univ., Tainan, Taiwan, [2]Dept. of Pediatrics, Natl. Cheng Kung Univ. Hosp., Coll. of Med., Natl. Cheng Kung Univ., Tainan, Taiwan, [3]Dept. of Ophthalmology, Natl. Cheng Kung Univ. Hosp., Coll. of Med., Natl. Cheng Kung Univ., Tainan, Taiwan

Mitochondrial diseases caused by impaired mitochondrial import machinery are phenotypically heterogeneous. Through trio-based exome sequencing, we identified a recurrent homozygous missense variant in a small translocase of outer mitochondrial membrane (TOM) in three unrelated probands with microcephaly, short stature, developmental delay, bird-headed faces, atrophic macular scarring, and moyamoya disease. Their parents are all heterozygotes, and the unaffected sibling has the common allele. Translocase of outer mitochondrial membrane that is essential in maintaining proper mitochondrial protein import. We established induced pluripotent cells (iPSCs) with patient-specific variant through CRISPR/Cas9 genome editing and differentiated the iPSCs into endothelial cells to examine the biogenesis of the TOM complex, mitochondrial bioenergetics, quality control, proteomics, and transcriptome. We identified a differential abundance of proteins involved in ATP production, TCA cycle, and gluconeogenesis in homozygous cells. Mitochondrial respiration and ATP production were decreased in homozygous and knock-out cells. Furthermore, transcriptomic studies revealed upregulation of genes involved in glycolysis and hypoxia in homozygous and knockout cells. Finally, CRISPR/Cas9-mediated zebrafish model showed facial dysmorphia, small brain size, and cerebrovascular defects that recapitulated patients' phenotypes. Our results revealed the molecular and cellular mechanisms through which mitochondrial protein import defects lead to a rare autosomal recessive microcephalic osteodysplastic dwarfism with moyamoya disease.

Session Title: Mendelian Phenotypes Poster Session III

PB4619 A Variant in Sperm-Specific Glycolytic Enzyme Enolase 4 *ENO4*Causes Human Male Infertility.

Authors:

S. Nawaz[1], S. Hussain[2], M. Bilal[3], N. Syed[1], W. Ahmad[4], K. Fakhro[5]; [1]Sidra Med., Doha, Qatar, [2]Clinical And Molecular Metabolism Res. Program, Faculty of Med., Univ. of Helsinki, Finland, Helsinki, Finland, [3]Dept. of Pathology and Lab. Med., Agha Khan Univ., Karachi, Pakistan, [4]Dept. of Biochemistry, Doha, Qatar, [5]Sidra Med. and Res. Ctr., Doha, Qatar, Qatar

Though abnormalities in sperm morphology and physiology lead to male infertility, in many instances, the exact disruption of molecular pathways in a given patient is not known. In the present study, we characterize a family with asthenozoospermia and abnormal sperm morphology at clinical and genetic levels. Further, we use computer-assisted semen analysis (CASA), papanicolaou smear staining (Pap smear), and scanning electron microscopy (SEM) to examine sperm motility and morphology. Analysis by whole exome sequencing revealed a homozygous variant [c.293A>G, p.(Lys98Arg)] in the Enolase 4 (*ENO4)* that segregated with infertility in the consanguineous family, shared by affected but not controls. In view of the association of asthenozoospermia and abnormal sperm morphology in *Eno4* knockout mice, we believe this is the first report describing the involvement of *ENO4* gene in human male infertility. We also explore the possible involvement of another variant in explaining other phenotypic features in this family.

Session Title: Mendelian Phenotypes Poster Session I

PB4620 *ABCA1*missense mutations are associated with Lewy body dementia

Authors:

P. Reho[1], A. Ray[1], G. Logroscino[2], S. Barmada[3], O. A. Ross[4], S. W. Scholz[1,5]; [1]Neurodegenerative Diseases Res. Unit, Natl. Inst. of Neurological Disorders and Stroke, Bethesda, MD, [2]Dept. of Basic Med. NeuroSci. and Sense Organs, Univ. of Bari, Bari, Italy, [3]Dept. of Neurology, Univ. of Michigan, Ann Arbor, MI, [4]Dept. of NeuroSci., Mayo Clinic, Jacksonville, FL, [5]Dept. of Neurology, Johns Hopkins Univ. Med. Ctr., Baltimore, MD

Lewy body dementia (LBD) is the second most common dementia after Alzheimer's disease (AD), affecting more than 1.4 million people in the United States. Increasing evidence shows that the genetic architecture of LBD intersects with AD. A recent study identified rare mutations in *ABCA1* and *ATP8B4* genes in Alzheimer's disease patients. Herein, we decided to investigate the association between rare variants in these genes with risk of developing LBD.We studied rare mutations in whole-genome sequence data from a cohort of 2,591 LBD cases and 4,032 controls of European ancestry. All participants have been previously sequenced at 35x coverage on an Illumina 10x platform using PCR-free library preparations. The sequence data were aligned to hg38 and jointly called, followed by stringent sample-level and variant-level quality control checks. We extracted and annotated missense and loss-of-function (LOF) mutations in *ABCA1* (NM_005502.4) and *ATP8B4* (NM_024837.4) genes, setting a minor allele frequency of <0.01 in the reference population. We tested the association of coding mutations with risk for developing LBD by applying the optimized sequence kernel association test (SKAT-O). Single-variant association analysis was performed by Fisher's exact test followed by Bonferroni correction for multiple testing.We identified a significant enrichment of *ABCA1* missense mutations with LBD (SKAT-O p = 2.68E-02), while no associations of *ABCA1* and *ATP8B4* LOF or *ATP8B4* missense mutations were detected. Interestingly, two *ABCA1*missense variants (*c.5398A>C*; p.N1800H and *c.3544G>A*; p.A1182T) showed significant variant-level association with LBD (Fisher p = 1.16E-04, p = 4.29E-02), and the p.N1800H substitution also survived the Bonferroni multiple testing correction (adjusted p = 1.14E-02). Specifically, this substitution was found in twelve LBD patients and one control (odds ratio = 18.48, 95%CI = 2.4-142.8). It is predicted to be pathogenic by the American College of Medical Genetics and has previously been associated with an increased risk for AD. Of note, seven *ABCA1* p.N1800H carriers were pathologically diagnosed cases, among which six patients carried an *APOE* ε4 risk allele and three carried one *GBA* pathogenic mutation (p.N409 or p.S146L). No deleterious mutations in *APP*, *PSEN1*, and *PSEN2* were identified among the *ABCA1* p.N1800H carriers. Our study identifies a significant enrichment of *ABCA1* missense mutations in LBD patients compared to healthy controls, emphasizing the pathogenic role of the p.N1800H substitution and the genomic complexity surrounding LBD.

Session Title: Mendelian Phenotypes Poster Session II

PB4621 ABCC6 three-state structural model enhances functional interpretation of genetic variants.

Authors:

J. Wagenknecht[1], N. Haque[1], S. Jorge[1], B. Ratnasinghe[1], R. Urrutia[1], W. Gahl[2], S. Ziegler[3], M. Zimmermann[1]; [1]Med. Coll. of Wisconsin, Milwaukee, WI, [2]NHGRI (NIH), Bethesda, MD, [3]Johns Hopkins Univ Sch. of Med., Baltimore, MD

Genetic variation within ATP Binding Cassette Subfamily C Member 6 (*ABCC6*) can cause either Pseudoxanthoma Elasticum (PXE) or the more severe condition, Generalized Arterial Calcification of Infancy (GACI). However, the molecular underpinnings are unknown for how variants alter ABCC6 protein function and if those differences contribute to why certain patients develop PXE versus GACI. These two recessive diseases are characterized by calcification of body tissues via hydroxyapatite deposition, yet have drastic differences in phenotype, progression, and prognosis. Furthermore, the same variant in one family can be associated with each condition in different patients. Functional interpretation of genetic variants within the protein is limited without an experimentally solved full-length protein structure. Thus, to better interpret the genetic variation within ABCC6, we used multiple homologous protein structures to generate three structural models that capture the main conformations that ABCC6 takes, i.e., unbound, substrate-bound, and ATP-bound. These models were used in structural calculations of variants from nationwide databases (ClinVar, gnomAD, and HGMD) to analyze 1,215 distinct missense genetic variants. Our use of multiple models is efficient for understanding where each variant acts in the functional mechanism, producing much deeper information than any singular molecular structure. This analysis was performed sequentially and structurally, focusing on the destabilizing effects on each state and the variants' proximity (both sequentially and spatially) to sites of interest (having a motif, post-translational modification, or known functional site). From this, not only were we able to predict a mechanism of damaging effect for 423 of the 991 variants of unknown significance, but we were able to analyze demographic and phenotypic data to look for patterns in inheritance across population groups and biologic sex, along with discovering 3D hotspots for mutations in PXE and GACI, and more. Our results aid in the mechanistic understanding of PXE and GACI and reveal a novel high throughput bioinformatic protein variation analysis method through multi-conformational models informed by functional data.

Session Title: Mendelian Phenotypes Poster Session III

PB4622 Abnormalities of TBX1 result in broad overlapping features of 22q11.2 deletion syndrome

Authors:

D. McDonald-McGinn[1,2,3], V. Giunta[4], T. Crowley[5], D. McGinn[6], B. Emanuel[4], E. Zackai[7], R. Smith[8], E. Moran[9], M. Geremek[10], B. Nowakowska[11]; [1]Children`s Hosp. of Philadelphia, Philadelphia, PA, [2]Perelman Sch. of Med. of the Univ. of Pennsylvania, Philadelphia, PA, [3]Sapienza Univ., Rome, Italy, [4]Children's Hosp. of Philadelphia, Philadelphia, PA, [5]Children s Hosp. of Philadelphia, Philadelphia, PA, [6]Children s Hosp. of Philadelphia and UPenn, Philadelphia, PA, [7]Children S Hosp. of Philadelphia, Philadelphia, PA, [8]Maine Med Partners, Pediatric Specialty Care, Portland, ME, [9]NYU-Hosp Joint Diseases, New York, NY, [10]Inst. of Mother and Child, Warsaw, Poland, [11]Inst Mother & Child, Warsaw, Poland

Background: Tbx1 is a member of the Tbox family of binding domain transcription factors. Mice haploinsufficient for Tbx1 have features associated with 22q11.2 deletion syndrome (22q11.2DS). Tbx1-homozygous null mutant mouse embryos die at birth with cleft palate, absent thymus and parathyroid glands, truncus arteriosus and ventricular septal defects. Despite this association with structural anomalies, and more recently autism, clinicians may be unaware of the importance of considering variants in TBX1, located within the 22q11.2 DiGeorge critical region, as a potential etiology in individuals with overlapping features of 22q11.2DS but without a deletion. Here we report findings in 8 patients with typical features of 22q11.2DS without 22q11.2DS, or another identifiable cause such as diabetic or retinoic acid embryopathy or CHD7 mutations, but with variants in TBX1, providing additional human evidence to support such clinical investigations. **Methods:** We identified 8 patients from 5 families with TBX1 variants under an IRB-approved investigation into 22q11.2DS phenocopies. **Results:** 5 probands, 1 half-sibling, and 2 parents had TBX1 variants (1 nonsense, 1 frameshift, and 1 familial missense mutation, 1 deletion, and 1 familial duplication). 5/8 were female and 5/8 white. Findings included: congenital heart disease (4) including Interrupted Aortic Arch (IAA) (2), tetralogy of Fallot, and vascular ring, significant developmental delay (4), hypocalcemia (2), immunodeficiency (3), FTT (2), lymphedema, and chorioretinal coloboma. The index case with IAA, bicuspid aortic arch, hypocalcemia, thymic hypoplasia, FTT, and GERD, was negative for 22q11.2DS in infancy. She was referred again for 22q11.2 testing at 5 years of age due to her history, as well as velopharyngeal dysfunction, chronic otitis media, obstructive sleep apnea, thrombocytopenia, cervical spine anomalies, winged scapula, hypotonia, developmental and speech delay. This led to investigation of additional patients with 22q11.2DS phenocopies, where the prevalence of associated features in those with TBX1 variants revealed few differences. **Conclusions:** 8 patients with TBX1 variants had overlapping features of 22q11.2DS, adding to a handful of previously reported cases in the literature (Yagi 2003, Zweier 2007, Pan 2015). These findings serve to further confirm the phenotypic overlap between patients with TBX1 variants and 22q11.2DS emphasizing the need to test all patients with 22q11.2DS phenocopies for TBX1 variants immediately when a deletion is not identified in order to provide appropriate medical management, genetic counseling, and to obviate a protracted diagnostic odyssey.

Session Title: Mendelian Phenotypes Poster Session I

PB4623 Adulthood vs. Childhood Onset MELAS: Clinical insights and prognostic factors explored.

Authors:

H. Bae[1], S. Hwang[1], G-H. Kim[2], J. Yoon[1], D. Kim[1], H. Shin[3], M. Kang[3], I. Choi[4], B. Lee[1,2]; [1]Dept. of Pediatrics, Asan Med. Ctr. Children's Hosp., Univ. of Ulsan Coll. of Med., Seoul, Korea, Republic of, [2]Med. Genetics Ctr., Asan Med. Ctr., Univ. of Ulsan Coll. of Med., Seoul, Korea, Republic of, [3]Asan Inst. for Life Sci., Asan Med. Ctr., Univ. of Ulsan Coll. of Med., Seoul, Korea, Republic of, [4]Dept. of Genetic Counseling, Univ. of Ulsan Coll. of Med., Seoul, Korea, Republic of

Introduction: MELAS (Mitochondrial Encephalopathy, Lactic Acidosis, and Stroke-like episodes) generally presents before age 20. However, MELAS develops in some patients after their twenties. Heteroplasmy, the coexistence of mutated and normal mitochondrial DNA within cells, significantly impacts phenotypic expression. This study aims to explore the clinical and molecular genetic characteristics of MELAS patients. **Method:** The medical records of MELAS patients diagnosed at Asan Medical Center, Seoul, Korea, from January 1998 to May 2023, were reviewed, and compared between childhood-onset and adult-onset MELAS. **Results:** A total of 55 patients from 41 families were enrolled. Fifty-one patients had m.3243 A>G heteroplasmy, 2 patients had m.9957 T>C homoplasmy, 1 patient had m.13513 G>A heteroplasmy, and 1 patient had m.1644 G>A homoplasmy. Eight asymptomatic patients were identified through family testing. Symptomatic patients were divided into two groups based on age of diagnosis: childhood-onset (diagnosed before age 20) and adult-onset. The childhood-onset group exhibited significantly higher mitochondrial heteroplasmy levels compared to the adult-onset group (52.7% [37.4-66.6] vs 29.9% [18.8-36.8], $p<0.001$). The earliest clinical manifestations were different between the two groups: seizures (39.1%), developmental delay (26.1%), and stroke (21.7%) were common in the childhood-onset group, whereas stroke (33.3%), diabetes mellitus (25%), and hearing loss (20.8%) were common in the adult-onset group ($p=0.003$). Median age of death was lower in the childhood-onset group (median, 17.8 [interquartile range, 13.7-27.2] years vs. 41.4 [35.4-53.4] years, $p=0.003$). Poor prognosis was associated with cardiomyopathy, T2/FLAIR hyperintensity on brain MRI, stroke episodes, speech disturbance, gait disturbance, hemiplegia, muscle weakness, short stature at diagnosis, and progression of chronic kidney disease. **Conclusion:** Significant differences were observed in the clinical manifestations and courses between childhood-onset and adult-onset MELAS the young and adult-onset groups of MELAS, associated with heteroplasmy levels. Further study is required to identify other factors contributing to these phenotypic differences.

Session Title: Mendelian Phenotypes Poster Session II

PB4624 † Advancing insights in Xia-Gibbs Syndrome: A universal registry platform for phenotypic analysis and longitudinal patient engagement

Authors:

A. Hansen[1], K. Ward[1], D. Belliston[1], T. R. Greer[1], T. Gerber[2], C. G. Hansen[1], M. Murugan[2], J. Hu[2], R. A. Gibbs[2]; [1]Geneial, Houston, TX, [2]Baylor Coll. of Med., Houston, TX

The pace of discovery of Mendelian disorders and rare diseases has far outweighed the growth of the clinical genetics workforce. Moreover, elucidating the phenotypic spectrum and natural history of these conditions is a labor-intensive process dependent on curation of high-quality phenotypic data. To reduce clinician burden - and to capture a more comprehensive picture including quality of life and social determinants of health - collecting survey data directly from patients or their caregivers is an essential element. Thus, there is a significant need for infrastructure to streamline the process of establishing and managing high-quality registries capable of integrating electronic patient-reported outcomes (ePROs) alongside molecular diagnostic and other clinical data.
To address these needs, we have developed a universal Mendelian disorder registry platform in partnership with the Xia-Gibbs Society that accelerates research and facilitates connectivity with external partners while maintaining data ownership, control, and privacy. We used the platform to rapidly generate an updated variant map for *AHDC1* - the gene underlying Xia-Gibbs Syndrome - and to compare phenotypes of patients joining the registry before 8/9/2021 (publication date of the most recent XGS phenotypic study) (N=94) vs. those joining between 8/9/2021 and 6/7/2023 (N=59). Sixteen HPO-mapped phenotypes were compared against the cohorts, seven of which revealed a statistically significant difference. Notably, all phenotypes with a change in proportion over time were less prevalent in the new cohort, which may suggest more patients are joining the registry with either a milder phenotype or benign *AHDC1* variant over time. For example, delayed speech proportion decreased from 91.49% to 65.52% (p=0.000185, chi-square); motor delay decreased from 90.43% to 63.79% (p=0.000187).
For patients/caregivers, the HIPAA-compliant iOS and Android app currently supports e-Consent and completion of standardized survey data. Through interviews with numerous registry participants and rare disease advocacy group leaders, we have identified an opportunity to return immediate value to participants through expansion of the app to support medical history organization, community connectivity, and dissemination of information. By doing so, we hypothesize that researchers will receive higher quality longitudinal data and an increased ability to reconnect with individuals who opt-in for clinical trial recruitment and other use cases.

Session Title: Mendelian Phenotypes Poster Session III

PB4625 Advancing Tay Sachs Disease Carrier Screening: Insights from Combined Enzyme and Molecular Approaches

Authors:

Q. Sun[1,2], K. V. Schulze[1,2], N. Liu[1,2], L. Meng[1,2], R. Sutton[1,2,3]; [1]Baylor Coll. of Med., Houston, TX, [2]Baylor Genetics, Houston, TX, [3]Texas Children's Hosp., Houston, TX

Tay Sachs disease carrier screening in a diverse, pan-ethnic population often employs both beta-hexosaminidase A (Hex A) enzyme and *HEXA* gene assays; the DNA assay may be either full-exon or targeted variant sequencing. To assess the efficacy of enzyme carrier testing, we conducted a retrospective analysis of carrier sequencing results from our laboratory database.

A total of 30 cases with positive Hex A enzyme analysis were selected. None of the enzyme activities in these cases were in the indeterminate range. These samples also underwent molecular carrier testing as part of a customizable, comprehensive panel of up to >400 genes, including the *HEXA* gene. While full gene sequencing was performed, the DNA carrier panel reported only selected pathogenic/likely pathogenic or pseudodeficient variants in the *HEXA* gene. Comparative analysis revealed consistent findings in 20 cases, 19 of which were pathogenic/likely pathogenic alleles and 1 pseudodeficient allele in the *HEXA* gene.

Two patients chose a customized carrier panel that initially excluded the HEXA gene. However, on further review in light of the positive enzyme test, both patients were found to carry pathogenic alleles in the *HEXA* gene. Notably, one of these patients was of North European descent. The remaining 8 cases were reported negative for molecular findings in *HEXA.* Nevertheless, further scrutiny of the sequencing results identified several variants of interest. Among these findings, three variants of uncertain significance (VUS) exhibited characteristics suggesting potential pathogenicity: c.1288G>A (p.D430N), c.1061_1063del (p.F354del), and c.590A>C (p.K197T). Another VUS, while likely benign, displayed moderate splicing predictions for an acceptor gain, necessitating further functional studies (c.673-13T>C). No changes were found in the last four cases.

Our results highlight the potential value of including molecular testing for *HEXA* in carrier screening and underscore the impact of novel VUS in Tay Sachs carrier sequencing. Furthermore, we demonstrate that incorporating functional screening of Hex A enzyme facilitates the curation and potential reclassification of VUS, thereby enhancing the sensitivity of Tay Sachs carrier screening. Done in combination, Hex A enzyme and *HEXA* sequencing assays were more effective in identifying true carriers of Tay Sachs disease than either methodology alone.

Session Title: Mendelian Phenotypes Poster Session I

PB4626 Altered lipid storage and metabolism in Smith-Magenis syndrome:*Rai1*haploinsufficiency causes increased storage of cholesterol esters, diacylglycerols, and triacylglycerols in a mouse model of SMS.

Authors:

J. Hauck[1], M. Sun[1,2], S. Hari[1], T. A. Wilson[1], S. H. Elsea[1]; [1]Baylor Coll. of Med., Houston, TX, [2]Rice Univ., Houston, TX

Smith-Magenis syndrome (SMS, OMIM #182290) is a complex neurodevelopmental disorder characterized by a distinct behavioral phenotype, intellectual disability, circadian rhythm disorder, obesity, and craniofacial and skeletal anomalies. Most features of SMS arise due to haploinsufficiency of *RAI1* due to either intragenic variation or 17p11.2 deletion. Previous work has shown potential links between SMS, *RAI1*, obesity, and lipid metabolism abnormalities, including hypercholesterolemia. To better characterize obesity and lipid metabolism alterations observed in SMS, baseline metabolic phenotyping of *Rai1*$^{+/-}$ mice, which display hyperphagia, early onset obesity, and altered adiposity, was performed to include activity, food consumption, growth, and metabolite and lipid analyses. Prior to onset of obesity, 6-week-old singly-housed mice displayed reduced cage activity, lack of wheel running, modestly shifted circadian activity, increased food intake, and rapid weight gain compared to normal littermates ($p<0.01$). Untargeted metabolomic profiling of >1000 metabolites extracted from livers of 5-month-old *Rai1*$^{+/-}$ and wildtype littermates showed altered niacin, biopterin, ascorbate, and glutathione metabolism, as well as decreased levels of plasmalogens and lysoplasmalogens, and evidence of oxidative stress. When compared to wildtype, lipidomic profiling of 883 unique, targeted lipid species in *Rai1*$^{+/-}$livers showed marked elevations of triacylglycerols (4.7-fold), diacylglycerols (1.9-fold), and cholesterol esters (2.9-fold), with decreased monoacylglycerol esters (0.19-fold), and phosphoethanolamine lipid species (0.74-fold) (Welch's two-sample t-test, $P<0.05$). C16, C18, and C20-containing lipids were most consistently altered, with reduction of larger lipid species, suggesting that processing of dietary fats and downstream fatty acid synthesis may be impaired in SMS, supportive of impaired lipid transport and/or processing in the liver. These results support previous claims that haploinsufficiency of *RAI1* contributes to lipid metabolism and autophagy defects in SMS and suggest that individuals with SMS may have an underlying lipid storage disorder. Targeted treatment of lipid metabolism could prove beneficial to persons with SMS, including lifestyle and diet modifications to moderate and manage intake of specific exogenous lipids toward normalizing hepatic lipid metabolism.

Session Title: Mendelian Phenotypes Poster Session II

PB4627 Aminoacylase 1 deficiency: case report on three affected siblings

Authors:

J. Srovnal[1], V. Smolka[2], D. Friedecky[2], O. Tkacik[2], J. Kolarova[2], H. Foltenova[2], V. Bekarek[2], P. Vrtel[2], M. Prochazka[2]; [1]Faculty of Med. and Dentistry, Palacky Univ., Olomouc, Czech Republic, [2]Univ. Hosp. Olomouc, Olomouc, Czech Republic

Background: Aminoacylase 1 (*ACY1*, EC 3.5.1.14) deficiency is a very rare inherited metabolic disease (IMD) with autosomal recessive inheritance (OMIM #609924). It is diagnosed by detecting acetylated amino acids among the patient's urine organic acids by gas chromatography-mass spectrometry. Its clinical manifestations are highly variable, ranging from severe neurological symptoms to being asymptomatic. **Case description:** We present a 14-year-old boy with mild intellectual disability who exhibited increased urinary excretion of acetylated amino acids during testing for inherited metabolic disorders. A suspected *ACY1* deficiency was confirmed by molecular genetic analysis, which revealed the presence of a homozygous pathogenic missense mutation in the *ACY1* gene, c.1057C>T (p.Arg353Cys). The proband underwent speech education with good outcome. The same homozygous mutation was found in the boy's two brothers, who exhibited slightly uneven intellectual ability profile distributions. Follow-up examinations of the siblings revealed no deterioration in their mental skills. **Conclusion:** These results suggest that uneven mental abilities in pediatric patients may be sufficient grounds to warrant metabolic testing for *ACY1* deficiency.**Funding:** The study was supported by Ministry of Education, Youth and Sport of the Czech Republic (NCMG - LM2023067, EATRIS-CZ - LM2018133), Palacky University Olomouc (LF 2023_006) and the project National Institute for Cancer Research (Programme EXCELES, ID Project No. LX22NPO5102) - Funded by the European Union - Next Generation EU.

Session Title: Mendelian Phenotypes Poster Session III

PB4628 An adult male with *UPF1*-related neurodevelopmental disorder

Authors:

M. Chung[1,2], K. Baker[2], M. Wiznitzer[3], J. Chenbhanich[1,2]; [1]Ctr. for Human Genetics, Univ. Hosp. Cleveland Med. Ctr., Cleveland, OH, [2]Dept. of Genetics and Genomic Sci., Case Western Reserve Univ., Cleveland, OH, [3]Dept. of Neurology, Case Western Reserve Univ., Cleveland, OH

The *UPF1* gene encodes for an ATP-driven RNA helicase involved in mRNA nuclear export and mRNA surveillance in the form of initiating mRNA nonsense-mediated decay (NMD). This process is responsible for recognition and destruction of aberrant mRNA containing premature stop codons. To the best of our knowledge, only two patients with *UPF1*-related neurodevelopmental disorder with clinical descriptions have been reported to date.

We describe a 38-year-old male with a history of infantile hypotonia, global developmental delay, intellectual disability, autism, scoliosis, myoclonic epilepsy, bilateral high-frequency hearing loss, and delayed pubertal development. The patient exhibited overgrowth features, as indicated by measurements of head circumference, height, and weight consistently exceeding two standard deviations since birth. Physical exam was remarkable for brachycephaly, tall and narrow face with broad forehead, baldness, midface hypoplasia, coarse facial features, deep-set eyes, downslanting palpebral fissures, high arched palate, large hands and feet, mild camptodacyly, pes planus, hindfoot deformity, and joint laxity. Echocardiogram was unremarkable. Chromosome microarray and fragile X testing were negative. Trio exome sequencing revealed a *de novo* c. 1846 A>G (p.Thr616Ala) variant in exon 14 of the *UPF1* gene. The variant was classified as Likely Pathogenic according to the American College of Medical Genetics and Genomics guidelines.

Further investigation is needed to elucidate the specific molecular mechanisms of *UPF1*-related neurodevelopmental disorder, particularly with respect to the disruption of the NMD pathway. Such knowledge is crucial as it may open up therapeutic possibilities, as evidenced by preclinical studies involving NMD inhibitors in other conditions. Notably, overgrowth has never been described in association with mutations in *UPF1*. Identification of other cases with variants in *UPF1* will allow for a more in-depth characterization of the phenotypic spectrum of this rare disorder.

Session Title: Mendelian Phenotypes Poster Session I

PB4629 An ultra-rare variant of GPX4 reveals the structural basis for lipid membrane attachment

Authors:

A. Wahida; Helmholtz Ctr. Munich, Munich, Germany

Since its coining in 2012, ferroptosis has rapidly evolved into a well-recognized cell death modality with wide implications in various disease entities. Hence, cellular protection mechanisms against ferroptosis constitute critical features of homeostasis. To date, only a few such systems have been characterized, with the enzymatic function of glutathione peroxidase 4 (GPX4) probably being the most important. However, little is known about the structural basis for its anti-ferroptotic function at membranous compartments. Sedaghatian-type Spondylometaphyseal dysplasia (SSMD) is a lethal and ultra-rare inborn disorder affecting newborns and has been linked to genetic lesions in GPX4. Here, we report on a newly discovered missense variant in GPX4 (R152H) which, unlike other truncating mutations, does not lead to loss of protein expression. Surprisingly, although its enzymatic activity remains unaffected by the mutation, the R152H variant invariably failed to rescue from ferroptosis induced by genetic deletion of the wildtype enzyme. A detailed structural analysis of GPX4_R152H using NMR showed that WT and GPX4_R152H were overall well-folded, with some chemical shift perturbations being limited and mapping primarily to a region centered around R152H. This region constitutes a fin-like loop, extruding from the 3D structure in the wildtype enzyme, which partially collapsed in the mutant protein as shown by X-ray structure analysis. Accordingly, NMR spectroscopy analysis of GPX4 wildtype and mutant enzymes with bicelles revealed a marked loss of interaction of the R152H variant with lipid bicelles in contrast to wildtype GPX4. From these studies, we concluded that the R152 is positioned in such a way that it stabilizes a loop essential for GPX4 to be partially immersed into lipid bilayers.

Session Title: Mendelian Phenotypes Poster Session II

PB4630 Analysis of cardiopulmonary function in patients with Gaucher disease

Authors:

G. Lopez[1], J. Lichtenberg[1], R. Seid[1], N. Tayebi[2], A. Steward[3], N. Crooks[1], E. Ryan[4], E. Sidransky[5]; [1]NHGRI (NIH), Bethesda, MD, [2]NIH/NHGRI, Bethesda, MD, [3]NIH, Bethesda, MD, [4]NIH, Bethesda, MD, [5]MGB/NHGRI/NIH, Bethesda, MD

Gaucher disease (GD) is an autosomal recessive lysosomal storage disorder (LSD) characterized by decreased function of the lysosomal enzyme glucocerebrosidase and a subsequent overabundance of the glycosphingolipid glucocerebroside. Glucocerebroside-engorged cells infiltrate various organs, leading to a range of clinical manifestations including enlarged spleen and liver, as well as skeletal and hematological abnormalities. GD is not typically associated with cardiac complications, except for genotype p.D448H/p.D448H, associated with cardiac valvular calcifications. However, few comprehensive studies of cardiac involvement in patients with GD have been reported. This study analyzed the echocardiogram data from 110 patients with either type 1 or type 3 GD, collected between 2006 and 2019. Using normal ranges for echocardiographic measurements established in the 2015 American Society of Echocardiography guidelines, our results show a substantial number of patients with abnormal aortic root and ascending aorta diameter. We also observed individuals in the cohort with parameters indicative of mild to moderate left ventricular hypertrophy. Mixed effects regressions with Subject ID as a random effect suggest a correlation between genotypes except those with an L483P allele and the extent of the enlargement of ascending aorta diameter and the left ventricular posterior wall diastolic thickness. In addition to genotype, regression analysis determined significant correlations between age and sex with different cardiac parameters. This exploratory analysis suggests that mild cardiac abnormalities may be more prevalent in patients with GD and potentially correlated with genotype. Longitudinal evaluation of patients with GD demonstrating abnormal echocardiogram measures is needed to determine whether these patients develop cardiac symptoms, or whether they remain abnormal but stable over time.

Session Title: Mendelian Phenotypes Poster Session III

PB4631 Anophthalmia/ Microphthalmia Spectrum: Clinical, cytogenetics and molecular results of 45 Egyptian Patients

Authors:

G. El-Makkawy[1], A. Nabil[1], F. Harms[2], D. Schorderet[3], V. Ruiz-Perez[4], M. Ali[5], D. Maskill[6], N. Elshakankiri[7], E. Abdalla[1]; [1]Dept. of Human Genetics, Med. Res. Inst., Alexandria Univ., Egypt, Alexandria, Egypt, [2]Inst. of Human Genetics, Univ. Med. Ctr. Hamburg-Eppendorf, Hamburg, Germany, Hamburg, Germany, [3]Faculty of Biology and Med. - Jules Gonin Eye Hosp. (HOP) -, Lausanne, France, [4]Inst. de Investigaciones Biomedicas de Madrid, madrid, Spain, [5]Leeds Inst. of Med. Res., Univ. of Leeds, Leeds, United Kingdom, [6]Leeds Inst. of Med. Res., Univ. of Leeds, leeds, United Kingdom, [7]Dept. of Ophthalmology, Univ. of Alexandria, Egypt, Alexandria, Egypt

Anophthalmia and microphthalmia (A/M) constitutes a spectrum of developmental ocular disorders that can be caused by chromosomal aberrations, copy number variations, and single gene mutations, besides environmental factors. A/M may present as an isolated anomaly or associated with other extraocular anomalies. Syndromic A/M represents the greater proportion of cases and is also known to have higher identifiable genetic causes and higher diagnostic rates. Identifying the underlying genetic cause is the key to provide patients with appropriate care by a multidisciplinary team and to offer their families a proper genetic counseling. The wide genetic heterogeneity of A/M, together with variation in expression, incomplete penetrance, and overlapping phenotypes, make diagnosis and genetic counseling very challenging. Aiming to gain insight into genetic etiologies of A/M in the Egyptian population, we studied a cohort of 45 patients derived from 39 unrelated Egyptian families with A/M spectrum. All patients were subjected to full clinical examination and ophthalmological assessment. Environmental factors were identified in 3 patients who were, thus, excluded from further genetic testing, and 4 patients were diagnosed as trisomy 13 and confirmed by karyotyping. For the remaining patients, molecular testing was conducted in 23 families; either in the form of direct sequencing of candidate genes or whole exome sequencing. The causative mutations were identified in 11 families by WES; 7 pathogenic, 2 likely pathogenic variants, and 2 variants of unknown significance in *OTX2, CHD7, HMX1, PRR12, ATOH7, STRA6, ZBTB11, B3GALNT2, GCNT2*, 8 of which are novel mutations, in addition to a pathogenic variant in *DPH1* gene in a female patient with bilateral severe microphthalmia and this is a novel phenotype associated with mutations in this gene. However, 9 families remained unsolved. Furthermore, *SOX2* Sanger sequencing was performed to 3 patients with features suggestive for anophthalmia-esophageal-genital syndrome and all of them turned out negative. We also clinically diagnosed other A/M syndromes as Oculo-dento-digital syndrome, Fryns syndrome, and Oculo-auriculo-vertebral spectrum. Collectively, the diagnostic yield was higher in the syndromic group than in the isolated A/M as potentially causal mutation was identified in the majority of the cohort with syndromic A/M 60% versus 40% in isolated A/M. Our data confirm that genetic factors constitute a predominant cause of both syndromic and non-syndromic A/M and, expand the mutational spectrum of A/M by reporting novel mutations in our population, also expand the phenotypic spectrum of specific syndromes.

Session Title: Mendelian Phenotypes Poster Session I

PB4632 Apert Syndrome: A New Variant With Uncertain Significance

Authors:

M. Arteaga Pichardo, F. Bernate, L. C. Rosas, A. V. Mora, M. C. Vega, D. D. Mora, A. Vargas, V. A. Peña, M. Ramirez, M. Saenz, S. Atuesta Escobar, L. G. Celis Regalado; Univ. de la Sabana, Cundinamarca, Colombia

Introduction: Apert syndrome is an autosomal disorder of low prevalence is a genetic disorder caused by the mutation of the gene that encodes the Fibroblast Growth Factor Receptor 2 *(FGFR2)* and is characterized by the presence of craniosynostosis accompanied in most cases by symmetric syndactyly. Its diagnosis is generally made prenatally during the first trimester of pregnancy, given that most craniofacial anomalies can manifest and be visible at week 19 of pregnancy.
Methods: The case of a 31 year old female G2P0A1,with normal prenatal controls displayed an ultrasound at 30.5 weeks gestation reporting cerebral ventricular asymmetry with ventriculomegaly of the left fetal ventricle. The patient is born at 39.2 weeks of gestation, XX genotype with adequate neonatal adaptation, weight and length. During the first months of life, the patient develops relevant phenotypic features: strabismus and proptosis of the right eye, hyperostosis of the right coronal suture with craniofacial deformity due to reduced diameter of the right orbit, right frontal planning with left frontal compensation and bilateral increase in hallux size. At 5 months of age the patient is intervened surgically for orbital correction and remodeling. A conventional cytogenetics and molecular panel of 38 genes for craniosynostosis is done.
Results: The study identifies a pathogenic heterozygous variant of the FGFR2 gene, consistent with a change of cytosine for guanine at position 755 of DNAc, in exon 7; level in which the protein produces a missense mutation of a serine for a tryptophan in the 252 amino acid. This variant is reported in databases as pathogenic for Apert Syndrome with an allelic frequency in the control population of 0.000004 and the responsible of 70% of cases in Apert syndrome. Another variant of important clinical significance in the same gene generated a change in cytosine for thymine in the 532 position in the DNAc, exon 5, at which the protein produces a missense mutation of arginine for a cysteine in the 178 amino acid moderately conserved. This variant has not been reported in gene databases and has an allelic frequency of 0.00001, classified as a deleterious allelic variant of uncertain significance.
Conclusions: A new variant was identified not yet reported in gen databases and it is classified as a deleterious allelic variant of uncertain significance for which further investigation in impact is merited. Additionally, there is a necessity to find a consensus with clearer, more specific clinical criteria to help diagnose this syndrome, as well as to determine the age of surgical intervention, personalized treatment and genetic counseling for each case.

Session Title: Mendelian Phenotypes Poster Session II

PB4633 Argininosuccinate lyase deficiency is associated with blood-brain barrier disruption via nitric oxide-mediated dysregulation of claudin expression.

Authors:

J. Kho[1], U. Polak[1], M-M. Jiang[1], J. Odom[1], J. Hunter[1], S. Ali[1], L. Burrage[1], S. Nagamani[1], R. Pautler[1], H. P. Thompson[2], A. Urayama[2], Z. Jin[1], B. Lee[1]; [1]Baylor Coll. of Med., Houston, TX, [2]Univ. of Texas Hlth.Sci. Ctr., Houston, TX

Argininosuccinic aciduria is the second most common urea cycle disorder (UCD) caused by deficiency of argininosuccinate lyase (ASL), a metabolic enzyme that catalyzes the fourth reaction in the urea cycle. Humans with argininosuccinate lyase deficiency (ASLD) are at increased risk for developing hyperammonemia due to a block in ureagenesis. However, natural history studies have shown that individuals with ASLD can present with more complex clinical phenotypes including neurocognitive deficits that can be independent of hyperammonemia. Thus, there's a significant gap in understanding the pathogenesis of neurocognitive deficits in ASLD. Blood-brain barrier (BBB) breakdown has recently been shown as an early biomarker of cognitive impairment in humans. We have previously discovered that beyond its role in ureagenesis in the liver, ASL is also required for the systemic synthesis of nitric oxide (NO), a critical signaling molecule that regulates BBB permeability and has been linked to pathogenesis of many neurocognitive diseases. Here, using ASLD as a model of cell-autonomous, nitric oxide synthase (NOS)-dependent NO deficiency, we investigated the effects of NO on brain endothelial cells *in vitro* and the BBB *in vivo*. *ASL* knockdown in human brain microvascular endothelial cells (HBMECs) led to decreased NO production and transendothelial electrical resistance (TEER), indicative of increased cell permeability. Loss of ASL also caused dysregulation of BBB-associated genes, marked by a decreased expression of CLDN5, a key regulator of BBB, and an increased expression of CLDN1, which is normally present in low concentration in the BBB. We found that treatment with an external source of NO or inhibition of *Claudin-1* improved barrier integrity in ASL-deficient HBMECs. To further evaluate the effects of NO deficiency on BBB *in vivo*, we independently assessed BBB leakage in a hypomorphic mouse model of ASLD by performing DCE-MRI (after injection with a gadolinium based-contrast agent) and Evans blue assay. We found a significant increase of both gadolinium enhancement and Evans blue dye extravasation in mutant mice, suggesting an evidence of BBB breakdown. Consistent with our *in vitro* data, we also discovered that BBB leakage could be partially rescued by NO supplementation. In conclusion, our results suggest that ASL-mediated NO synthesis is required for proper maintenance of brain microvascular endothelial cell functions as well as BBB integrity.

Session Title: Mendelian Phenotypes Poster Session III

PB4634 *ARHGAP1* identified as a candidate gene for a novel autosomal dominant syndromic neurodevelopmental disorder.

Authors:

H. Mendez[1,2], Y. Huang[3,4], D. Bonner[1,2], O. Kanca[3,4], R. Levy[5,2], E. Kravets[1,2], J. Kohler[1,2], C. Reuter[1,2], S. Marwaha[1,2], M. Wangler[3,4], S. Yamamoto[3,4], J. A. Bernstein[1,2], H. Bellen[3,4], M. T. Wheeler[1,2], The Undiagnosed Diseases Network; [1]Stanford Ctr. for Undiagnosed Diseases, Stanford, CA, [2]Stanford Univ., Stanford, CA, [3]Dept. of Molecular and Human Genetics, Baylor Coll. of Med., Houston, TX, [4]Jan and Dan Duncan Neurological Res. Inst., Houston, TX, [5]Stanford Hlth.Care, Stanford, CA

Background: GTPase-activating proteins play a crucial role in a wide spectrum of signal transduction pathways regulated by GTP-binding proteins. In this study, we provide initial clinical and functional evidence supporting a role for a Rho GAP gene, *ARHGAP1*, in a patient presenting with a syndromic neurodevelopmental disorder. **Case report:** We report a 6-year-old female presenting with microcephaly, global developmental delay, and dysmorphic features. She is the first child of a non-consanguineous couple with no relevant family history. Prenatal history of controlled gestational diabetes and decreased fetal movement was reported. She was born at full term but required admission to the NICU due to inadequate feeding. She started sitting at 7 months, walking at over 23 months and spoke her first words at 23 months. Physical examination revealed a depressed nasal root, broad bridge, and bilateral short fifth fingers. X-ray of the hands revealed short fifth metacarpals, angel-shaped fifth middle phalanges, and small distal phalanges. Additionally, a brain MRI displayed a few small cystic changes in the posterior periventricular white matter. Trio genome sequencing identified a de novo heterozygous missense variant in *ARHGAP1* (NM_004308.5): c.309G>T (p.Lys103Asn). In silico prediction tools support a deleterious effect (CAAD Score: 23.3, PolyPhen and SIFT: damaging), and the variant is absent in population databases (gnomAD and Bravo TopMED). **Model organism:** To investigate the functional consequences of the *ARHGAP1* variant, we performed experiments in Drosophila. First, we generated a versatile T2-GAL4 gene trap line in the fly *ARHGAP1* (RhoGAP68F) and found that this is not an essential gene. Second, we found that RhoGAP68F is expressed in a subset of neurons and glia in the nervous system. Third, while overexpression of the variant human *ARHGAP1* causes semi-lethality accompanied by shortened lifespan and developmental defects in appendages such as the wing and legs, the reference protein expressed under the same condition causes much milder or no phenotype. Since similar defects are found when the fly RhoGAP68F is overexpressed, the human *ARHGAP1* p.Lys103Asn variant is likely to be a gain-of-function allele. **Conclusions:** These data provide initial evidence that *ARHGAP1* gain-of-function variants might be the cause of a novel syndromic neurological disease involving severe neurological impairment with distal limb anomalies and dysmorphism. This association would be strengthened by the identification of other patients with *ARHGAP1* variants.

Session Title: Mendelian Phenotypes Poster Session I

PB4635 Assessing the genetic and phenotypic spectrum of *DNM1*-related disorders in the landscape of neurodevelopmental disorders

Authors:

S. Parthasarathy[1], A. Harrison[1], K. Thalwitzer[1], M. Pendziwiat[2], Y. Weber[2], S. Ruggiero[1], I. Helbig[1]; [1]Children's Hosp. of Philadelphia, Philadelphia, PA, [2]Univ.sklinikum Schleswig Holstein Campus, Kiel, Germany

Disease-causing variants in *DNM1* are an established cause of a severe neurodevelopmental disorder (NDD) including hypotonia, severe to profound developmental delay, and West syndrome, a severe epilepsy characterized by infantile spasms and hypsarrhythmia. Known pathogenic variants are predominantly missense variants causing dominant-negative impairment of vesicle fission by dynamin-1 homooligomers. Since the initial descriptions of *DNM1*-related disorder, a more complex genetic landscape has begun to emerge, including alternative exons, splicing variants, and biallelic loss-of-function; moreover, several clinical reports in the literature are expanding the known phenotypic spectrum. Accordingly, we harmonized clinical data across 73 individuals with *DNM1*-related disorder, including 62 previously reported and 11 unreported individuals. We used the Human Phenotype Ontology to capture 3,508 phenotypic annotations and compared this cohort to data from previous studies on two other severe genetic epilepsies, *SCN2A*-related (n = 413) and *STXBP1*-related (n = 534) disorders. Using an HPO-based semantic similarity approach, we show that *DNM1*-related disorder represents a homogeneous disease entity within the context of other neurodevelopmental disorders ($p < 0.01$). Furthermore, individuals with variants in the alternatively spliced exon 10a, including the recurrent p.I398_R399insCR variant, had a highly similar phenotypic picture within the *DNM1* spectrum (n = 13, $p < 0.01$); this similarity was driven by higher frequency of cortical visual impairment (HP:0100704, $p < 0.001$, OR = 22.8 [95% CI = 4.55-144]) and more profound developmental delay (HP:0012736, OR = 112 [95% CI = 11.6-5,600]), as well as lower frequency of speech impairment (HP:0002167, $p = 0.02$, OR = 0.211 [95% CI = 0.0419-0.872]). By using a new semantic similarity method for comparison between subgroups, we demonstrate that both this exon 10a-related phenotype ($p < 0.01$) and that of individuals with biallelic loss-of-function variants (n=3, $p < 0.01$) fall in the overall *DNM1* spectrum within a larger NDD context, illustrating that this phenotype is caused by reduction of a majority of dynamin-1 protein function. In summary, through a computational phenotyping approach, we define the *DNM1*-related phenotype as a homogeneous disease entity, establishing a more comprehensive baseline for granular anticipatory guidance and development of therapies targeted towards largely recurrent dominant-negative variants.

Session Title: Mendelian Phenotypes Poster Session II

PB4636 Assessment of potential loss of function variants in children with DS-associated congenital heart defects as possible modifying factors.

Authors:

E. Feldman[1], D. Cutler[1], T. Rosser[1], S. Wechsler[1], L. Sanclemente[2], A. Rachubinski[3], K. Rabin[2], M. Wagner[4], B. Gelb[5], J. Espinosa[3], P. Lupo[2], S. Sherman[1], E. Leslie[1]; [1]Emory Univ., Atlanta, GA, [2]Baylor Coll. of Med., Houston, TX, [3]Univ. of Colorado, Aurora, CO, [4]Cincinnati Children's Hosp. Med. Ctr., Cincinnati, OH, [5]Icahn Sch. of Med. at Mount Sinai, New York, NY

Congenital heart defects (CHDs) are the most common structural birth defect and are present in 40-50% of children with Down syndrome (DS). In DS, the most common types of CHD are atrial septal defects (ASD), ventricular septal defects (VSD), and atrioventricular septal defects (AVSD). While increased dosage of genes on chromosome 21 contributes to this risk, about 50% of infants with DS have structurally normal hearts. Thus, additional modifying genetic or environmental factors may exist. In order to characterize the genetic architecture of various DS-associated CHDs, we sequenced genomes of a multiethnic group of children with DS and a CHD (DS+CHD, n=886: AVSD, n=438; ASD, n=122; VSD, n=170; Other CHD, n = 156) or DS and a structurally normal heart (DS+NH, n=572) as part of the Gabriella Miller Kids First Pediatric Research Program and the INCLUDE (INvestigation of Co-occurring conditions across the Lifespan to Understand Down syndromE) Project. We analyzed rare variants (allele frequency < 0.0001 in gnomAD) that are predicted to be loss of function (pLoF). Exome-wide, rare pLoF variants were not enriched in DS+CHD vs. DS+NH ($p = 0.96$) nor were they enriched when restricting to genes constrained to loss-of-function variation. We then prioritized rare pLoF variants in 611 genes included on clinical genetic testing panels for CHD. We found 105 such variants in 78 genes but these were also not enriched among individuals with DS+CHD ($p = 1$). A total of 64 DS+CHD cases and 38 DS+NH controls carried at least one rare pLOF variant in the gene list. The majority of individuals carried a single pLoF allele in a gene associated with autosomal recessive (AR) disease. Interestingly, two individuals, both with NH+DS, had two heterozygous variants (and are suspected to be compound heterozygous) in *IFT172* and *DNAAF1*, both associated with autosomal recessive syndromes that typically include CHD. 9 individuals with DS+CHD were heterozygous for variants in genes associated with autosomal dominant CHD or heart dysfunction including *EP300* (DS+VSD), *EVC2* (DS+ASD), *KCNA5* (DS+AVSD), *LAMA4* (DS+AVSD), *MIB1* (DS+AVSD), *SCN2B* (DS+AVSD, DS+ASD+VSD), *SMC3* (DS+ASD+VSD), and *ZMIZ1* (DS+AVSD). Overall, we find that approximately 7% of individuals with DS+CHD and those with DS+NH harbor rare, potentially diagnostic pLoF variants in CHD-associated genes. Most of the observed variants were in genes unconstrained to LoF variation and associated with recessive disease. Here, the absence of an enrichment in DS+CHD cases suggests that being a carrier for AR-associated CHD genes does not interact with trisomy 21 to lead to CHD.

Session Title: Mendelian Phenotypes Poster Session III

PB4637 Association between rs1045642, rs2032582 and rs1128503 polymorphism of ABCB1 gene in Asian population and resistance to AEDs, a meta-analysis.

Authors:

Z. Agha, m. salman, i. shaheen; COMSATS Univ. ISLAMABAD, Islamabad, Pakistan

Epilepsy is a neurological disorder characterized by prolonged and recurrent seizures. It is considered the most common intellectual disorder relate to the nervous system. Despite the increased use of anti-epileptic drugs (AEDs) for treatment, 30% of patients remain unresponsive to them. Drug-resistant epilepsy may cause by pharmacokinetic and pharmacogenetic factors and cause severe disability. Several studies reported the association of ATP-binding cassette superfamily members such as ABCB1 gene polymorphisms C3435T, G2677T, and C1236T with drug resistance to AEDs. In the present study, we performed a meta-analysis to assess the association between common polymorphism in *ABCB1* and the risk of drug resistance in epileptic patients. We collected data from multiple online libraries before November 2021, and 12 studies were included selected for our meta-analysis. To assess the strength of association, we calculated pooled odds ratios with 95% confidence intervals. The Fisher's Z test suggests there is no publication bias in selected studies and results of statistical analysis showed that common SNPs rs1045642 (C3435T), rs1128503 (C1236T), and rs2032582 (G2677T) of ABCB1 are associated with drug resistance in an Asian population with (OR= 2.0606, CI95% [1.5827; 2.6828]; P= 7.85e-08). A sub-group analysis showed C3435T and C1236T variants are associated with drug resistance in Caucasians while G2677T is presented in the Arab ethnic group. Conclusion: We conclude that the polymorphisms of *ABCB1* (rs2032582, rs1045642, rs1128503) are associated with the risk of drug-resistant epilepsy in general epileptic patients. Meta-analysis indicates the association of *ABCB1* polymorphism with drug-resistant epilepsy is approximately independent of the ethnic group.

Session Title: Mendelian Phenotypes Poster Session I

PB4638 *ASXL3*: Connecting chromatin biology to neurodevelopmental disorders.

Authors:

E. Peirent, Y-C. Tsan, S. Regan, C. Ryan, S. Bielas; Univ. of Michigan, Ann Arbor, MI

Human genetics studies of neurodevelopmental disorders highlight chromatin's importance in corticogenesis, with pathogenic variants enriched in networks linked to chromatin regulation. Dynamic regulation of histone modifications is especially critical for the transcriptional plasticity required during this cellular differentiation. One such modification is mono-ubiquitination of histone H2A (H2AUb1), a conserved, traditionally repressive histone mark that is reversed by the polycomb repressive deubiquitinase (PR-DUB) complex. We identified *de novo* dominant truncating variants in *ASXL3*, a key component of PR-DUB, as the genetic basis of both Bainbridge Ropers Syndrome (BRS) and autism spectrum disorder (ASD), characterized by failure to thrive, global developmental delay, feeding problems, hypotonia, profound speech deficits, and intellectual disability. We identified dysregulation of H2AUb1 as a key molecular pathology in primary cells derived from individuals with BRS. To investigate this *ASXL3*-dependent neuropathology in early corticogenesis, we generated human neural progenitor cells and 3D cerebral organoids that recapitulate context-dependent features of *in vivo* cortical development in a reproducible manner. In CRISPR-edited and patient-derived iPSC lines, we observe *ASXL3*-dependent defects in the differentiation of pluripotent cells to the full spectrum of mature cortical neuron subtypes. We then utilized transcriptomic and epigenomic techniques including RNA-sequencing and CUT&RUN to probe the role of *ASXL3*-dependent H2AUb1 deubiquitination in regulating transcriptional profiles critical to NPC fate decisions during corticogenesis. Finally, we examined how such changes in dynamic H2AUb1 exchange impact DNA damage repair. Together, our functional investigation of BRS- and ASD-associated genetic variants provides molecular insights towards uncovering the elusive role of H2AUb1 in neural development.

Session Title: Mendelian Phenotypes Poster Session II

PB4639 Atypical clinical spectrum in *NDUFS2*-related mitochondrial complex I deficiency.

Authors:

P. Jacob[1], A. Vasanth[2], S. J. Patil[3]; [1]Dept. of Med. Genetics, JSS Med. Coll. and JSS Hosp., Mysuru, India, [2]Pediatric Nephrology, Dept. of Pediatrics, Mazumdar Shaw Med. Ctr., Narayana Hrudayalaya Hosp., Bangalore, India, [3]Dept. of Med. Genetics, Mazumdar Shaw Med. Ctr., Narayana Hrudayalaya Hosp., Bangalore, India

IntroductionMitochondrial complex I deficiency, nuclear type 6 (MIM#618228) is a rare subtype of oxidative phosphorylation (OXPHOS) disease caused by biallelic variants in *NDUFS2*. This early-onset, multi-systemic disorder is characterized by respiratory abnormalities, neurologic regression, muscular atrophy, cardiac anomalies, and lactic acidosis, often leading to a lethal outcome before the age of ten years.
MethodsOne affected individual born of a consanguineously married couple was ascertained for this study. Detailed clinical features, family history, along with radiological, and biochemical findings were recorded. Systemic investigations including MRI, USG (abdomen), ECHO, GCMS (urine), and TMS (blood) were performed. Exome sequencing (ES) was performed for the proband, data analysis, and interpretation were performed using standard procedures.
ResultsThe proband initially presented with failure to thrive, developmental delay, polyuria, and polydipsia at the age of one year. At 1 year 8 months of age, he had subtle facial dysmorphism, distal renal tubular acidosis, and increased serum lactate levels (3.8 mmol/L). Other systemic examinations were unremarkable. ES helped in the identification of a likely pathogenic variant, c.685C>A p.(Pro229Thr) in homozygous state in *NDUFS2* which causes Mitochondrial complex I deficiency, nuclear type 6. A novel missense change at the same amino acid residue p.(Pro229Gln) has been reported previously.
ConclusionOur study provides the first evidence of a patient with mild clinical spectrum of *NDUFS2*-related mitochondrial complex I deficiency. Early molecular diagnosis helped us to initiate the clinical management of the patient with Riboflavin. Interestingly, the reported patient with the variant at the same amino acid codon has severe clinical manifestations of the disease. Further functional studies would help to determine the rationale for the clinical heterogeneity of this rare mitochondrial disease.

Session Title: Mendelian Phenotypes Poster Session III

PB4640 Atypical plantar hyperkeratosis, multiple pigmented nevi, facial dymorphism, skeletal abnormalities, and mental retardation.

Authors:

A. Rincón[1], S. A. Ramírez-García[2], R. Hernández-Ramírez[3], V. C. Ledezma-Rodríguez[4], V. Y. Jimenez-Rodríguez[3], N. O. Dávalos-Rodríguez[5,6]; [1]IBMMTG, Depto. Molecular Biology and Genomic, Univ. of Guadalajara,, Guadalajara, Jalisco., Mexico, [2]MePiel Dermatological Ctr., Guadalajara, Jalisco., Mexico, [3]Med. degree, CUCS Univ. of Guadalajara,, Guadalajara, Jalisco, Mexico, [4]Dept. of Plastic and Reconstructive Surgery of the Hosp. de Especialidades, CMNO IMSS, Guadalajara, Mexico, [5]Inst. of Human Genetics, Dept. of Molecular Biology and Genomics, CUCS Univ. of Guadalajara,, Guadalajara, Jalisco,, Mexico, [6]Genetics, Regional Hosp. "Dr. Valentín Gómez Farías" ISSSTE., Zapopán, Jalisco, Mexico

Pigmentary lesions and palmoplantar hyperkeratosis may be related to different syndromes. However, the combination of facial dysmorphia, skeletal abnormalities, and palmoplantar keratoderma does not fit into any known syndromes. **Case report.** A 24-year-old female with facial dysmorphism, multiple nevi, plantar hyperkeratosis, mental retardation, and skeletal abnormalities is described. The proposita was the 3rd pregnancy of non-consanguineous parents, with normal karyotype and irrelevant hereditary family history. The arm span was 171 cm revealing a disproportion in upper limbs. She had facial dysmorphism, brachycephaly, arched and sparse eyebrows, down-slanting palpebral fissures, deep-set eyes, epicanthal folds, a bulbous nasal tip, and smooth philtrum, thick upper lip, malpositioned teeth, and mandibular prognathism. Bilateral cubitus valgus and thoracic scoliosis. Large hands according to the body, bilateral clinodactyly of the fifth fingers, and distal camptodactyly of the 2nd and 3rd fingers bilaterally. The nails were convex, and she had deep creases on the palms. Feet revealed hallux valgus bilateral, partial proximal cutaneous syndactyly between the 1st and 2nd toe and overlapping of the second over the first toe. The skin showed multiple pigmented nevi over the body. Radiological studies revealed temporomandibular sclerosis, wide sella turcica, and hypertrophic nasal conchae. Thorax with wide ribs, thoracolumbar scoliosis, and lumbar hyperlordosis; irregular vertebral bodies with a big right apophysis transverse on C7, intervertebral collapse at the dorsal level, sclerosis, and osteophytes grade II-IV at the lumbar level and interphase arthrosis; L5 has incomplete dysraphism and big apophyses transverses. Extremities upper: the radius bone showed a not a well-defined limit in the middle's union with the proximal third segment. Lower: *genu valgus* with medial arthrosis and *pes valgus* secondary to a tarsal coalition. Large bones still present growth nucleus rest. **Discussion**: Multiple nevi can occur in families (MIM#162900), punctate hereditary palmoplantar keratoderma (MIM#148600), and one case was described by Öktenli et al., 2003 with facial dysmorphism, multiple pigmented nevi, osteoporosis, brachydactyly, and skeletal anomalies. Although the present case showed clinical findings similar to those previously mentioned, our patient differs in facial dysmorphism, skeletal abnormalities, multiple pigmented nevus, plantar hyperkeratosis, and mental retardation. We conclude that the patient presented several findings not included in any known syndrome reported that could represent a new entity.

Session Title: Mendelian Phenotypes Poster Session I

PB4641 Autonomic nervous symptom may be a novel clinical manifestation of stress-induced childhood-onset neurodegeneration with variable ataxia and seizures syndrome caused by compound heterozygous variants in *ADPRHL2*.

Authors:

E. Sato[1], M. Fujimoto[1], T. Iwaki[1], Y. Nakamura[1], D. Ieda[1], Y. Negishi[1], A. Hattori[1], Y. Kubota[2], M. Asai[2], S. Saitoh[1]; [1]Nagoya City Univ. Graduate Sch. of Med. Sci., Nagoya, Japan, [2]Daido Hosp., Nagoya, Japan

[Background]*ADPRHL2* encodes ADP-ribosylhydrolase3(ARH3), which removes poly(ADP-ribose) polymers after ADP-ribosylation process under the cellular stress. Stress-induced childhood-onset neurodegeneration with variable ataxia and seizures (CONDSIAS) is an autosomal recessive disease and caused by loss of function in *ADPRHL2*. Other clinical features are developmental delay or regression and axonal sensorimotor neuropathy. Although all previous cases had been homozygous variants in *ADPRHL2* since first reported in 2018, we describe an additional patient with a novel compound heterozygous variant in *ADPRHL2* following two patients of compound heterozygous variants reported last year.[Case report]The patient is a 7-year-old male born with a weight of 2870g at 41 weeks of an uneventful pregnancy. His early development was normal. At 1 year 2 months of age, he was able to walk independently. Since 1 year 6 months of age, he had had recurrent episodes of ataxia and walking difficulty with fever. After those episodes he had recovered gradually in several weeks. At the age of 7 years, he showed walking difficulty, limb weakness and urinary retention in association with fever. The amplitude and conduction velocity of the motor nerve decreased. His head and spinal cords MRI did not show obvious pathological findings. We performed whole exome sequencing analysis and identified compound heterozygous variants in *ADPRHL2* gene (NM_017825:c.580C>T:p.Gln194Ter, NM_017825:c.712G>T:p.Glu238Ter). Sanger sequencing was applied to validate the patient and his parents. Those variants c.580C>T and c.712G>T inherited from his father and mother, respectively. c.712G>T was a novel variant while c.580C>T was reported previously in a patient with CONDSIAS. According to the ACMG guideline, both of the variants were classified as pathogenic.[Conclusion]As a previous report of a patient with compound heterozygous variants in *ADPRHL2*, he showed autonomic nervous dysfunction such as urinary retention, which was not mentioned in patients with homozygous variants. We consider that autonomic nervous dysfunction could be added to classical manifestations of CONDSIAS.

Session Title: Mendelian Phenotypes Poster Session II

PB4642 *AUTS2* syndrome: a recognizable clinical entity.

Authors:

A. Pinto[1], L. Loberti[2], A. Loredaria[2], E. Antolini[2], G. Casamassima[2], A. Destrèè[3], N. Brunetti-Pierri[4], R. genesio[5], D. Genevieve[6], P. Christophe[7], C. Coubes[6], H. Van Esch[8], T. Herget[9], A. Möllring[9], M. Zenker[10], J. Levy[11], L. Perrin[12], A. Maruani[12], A. Sorlin[13], D. Stieber[14], L. Herissant[15], K. Dahan[15], L. Sinibaldi[16], R. Capolino[16], M. Dentici[16], B. Dallapiccola[17], A. Novelli[18], L. Garavelli[19], O. Zuffardi[20], S. Caraffi[19], I. Valenzuela[21], M. Digilio[16], R. Caumes[22], C. Knopp[23], K. Chwiałkowska[24], S. Grosso[25], R. Canitano[26], R. Tita[1], M. Bruttini[27], I. Longo[28], M. Baldassarri[2], M. Mencarelli[28], A. Renieri[2]; [1]Med. Gentics, Azienda Ospedaliera Univ.ria Senese, Siena, Italy, [2]Med. Genetics, Univ. of Siena, Siena, Italy, [3]Inst. for Pathology and Genetics, Gosselies, Belgium, [4]Telethon Inst Genetics & Med., Pozzuoli, Italy, [5]Univ. of Naples Federico II, Naples, Italy, [6]Département de génétique médicale, Hôpital Arnaud de Villeneuve, Montpellier, France, [7]Unité Fonctionnelle Innovation en Diagnostic Génomique des Maladies Rares, Dijon, France, [8]Ctr. for Human Genetics, Univ. Hosp. Leuven, Leuven, Belgium, [9]Inst. of Human Genetics, Univ. Med. Ctr. Hamburg-Eppendorf, Hamburg, Germany, [10]Univ. Hosp. Magdeburg, Magdeburg, Germany, [11]AP-HP, PARIS, France, [12]Dept. of Genetics, Robert Debré Hosp., AP-HP, Paris, France, [13]Laboratoire Natl. de Santé, Dudelange, Luxembourg, [14]Laboratoire Natl. de Santé, Dudelange, Luxembourg, [15]Laboratoire Natl. de Santé, Dudelange, Luxembourg, [16]Ospedale Pediatrico Bambino Gesù, Roma, Italy, [17]Inst CSS-Mendel, Rome, Italy, [18]Lab. of Med. genetics translational cytogenomics research unit, Roma, Italy, [19]Genetica medica diartimento materno infantile, IRCSS Arcispedale S. Maria Nuova, Reggio Emilia, Italy, [20]Univ Pavia, Pavia Pv, Italy, [21]Area of Clinical and Molecular Genetics, Vall d'Hebron Univ. Hosp., Barcellona, Spain, [22]Service de Génétique clinique, Hosp. jeanne de flandre, lille, France, [23]Inst. for Human Genetics and Genomic Med., Med. Faculty, Rheinisch-Westfälische Technische Hochschule Aachen Univ., 52074 Aachen, Germany, Aachen, Germany, [24]Imagene.ME SA, Tychy, Poland, [25]Pediatric Departmen, Univ. of Siena, Siena, Italy, [26]Div. of Child and Adolescent Neuropsychiatry, Univ. of Siena, Siena, Italy, [27]Med. Gentics, Univ. of siena, Siena, Italy, [28]Med. Genetics, Azienda Ospedaliera Univ.ria Senese, Siena, Italy

AUTS2 syndrome is a fairly newly identified condition characterized by autism spectrum disorder, developmental delay and intellectual disability, hyperactivity, short stature, distinctive facial features, feeding difficulties, seizures. AUTS2 gene transcribes a distinct long and short isoforms from alternative promoters. In the context of a European collaborative research study, using array CGH combined with next generation sequencing techniques, we collected a large cohort of 37 patients harboring genomic rearrangements in the AUTS2 region (13/37) and point mutations (24/37). The identified pathogenic variants were mostly loss of function and some of them, like the p. R316* in exon 7 or the p.His535_Thr542del in exon 9, were recurrently found in patients from different countries suggesting the existence of mutational hotspots. The genomic rearrangements consisted mostly of interstitial deletions (6/13) spanning from exon 3 to exon 6 with two deletions in the 5' part of the gene (2/13),one of which affecting neighboring genes and one C terminal part deletion (1/13). A duplication of the 5' region was detected in three patients and a duplication extended from exon 7 to the 3'UTR was detected in one patient. A common clinical phenotype was observed independently from the underlying molecular mechanisms; it consists of autistic behavior, hyperactivity, learning difficulties and speech delay. A statistically significant more recognizable phenotype, with a typical facial gestalt consisting of microcephaly, stereotypic movements, hypertelorism, short and downslanting palpebral fissures, broad nasal bridge, anteverted nares and prominent nasal tip, was mostly observed in patients harboring loss of function variants compared to genomic rearrangements Neurological signs such as arthrogryposis and joint contractures were only noticed in patients harboring point mutations in the first exons (2-5) suggesting a role of the longer isoform in these aspects.Our data broaden the clinical and molecular spectrum associated with AUTS2-related disorder. They allow concluding that while AUTS2 rearrangements are mostly characterized by behavioral features and intellectual disability, point mutations are distinguished by a recognizable facial gestalt which can resemble Rubinstein-Taybi syndrome.

Session Title: Mendelian Phenotypes Poster Session III

PB4643 BCM-GREGoR: A rare disease program to solve the unsolved with novel methods and analytical approaches.

Authors:

S. Jhangiani[1], F. Sedlazeck[1,2], H. Du[1], R. Duan[1], D. Calame[1], V. R. Sutton[1,3], M. Dawood[1], M. Mahmoud[1], S. Behera[1], A. K. Saad[1], A. Rivera[1], A. Jolly[1], A. Rai[4], C. Munderloh[1], C. M. Grochowski[1], D. Marafi[1], D. Pehlivan[1], I. K. Chinn[1,3], J. Fatih[1], N. Gogate[1], P. Kaur[1], P. Liu[1], R. M. Belanger Deloge[1], R. Lewis[1,3], S. Barish[1], S. T. Beheshti[1], S. Morris[1,3], T. Bozkurt[4], Z. Dardas[1], Z. Coban Akdemir[4], Z. Khan[1], K. Walker[1], H. Doddapaneni[1], J. R. Lupski[1,3], D. M. Muzny[1], R. A. Gibbs[1], J. E. Posey[1]; [1]Baylor Coll. of Med., Houston, TX, [2]Rice Univ., Houston, TX, [3]Texas Children's Hosp., Houston, TX, [4]Univ. of Texas Hlth.and Sci. Ctr., Houston, TX

Nearly 30 million people in the United States are affected by rare diseases, although many families lack a molecular diagnosis. Since its inception in 2021, the GREGoR Consortium (Genomics Research to Elucidate the Genetics of Rare diseases) continues to optimize benchside methods, utilize diverse sequencing approaches, and develop novel analytical tools toward the aim of 'solving the unsolved'. BCM-GREGoR has enrolled more than 1,400 individuals (340 families), and the Human Genome Sequencing Center has generated sequence data for 1,065 study participants and performed reanalysis of extant data for more than 200 samples utilizing a stepwise approach to apply exomes, short and long read genome sequencing technologies, and RNA sequencing. Additionally, we are investigating the diagnostic utility of short and long read RNA sequencing in 8 families. For a cost-effective long read approach, we developed a custom panel and a sequencing protocol that covers 389 genes inaccessible with short reads alone and yields on average, 94% of targeted bases at 8x coverage or greater. Application of this panel to several families is in progress. Novel analytical approaches leverage this rare disease dataset, and we are currently investigating the role of reference/pangenomes in identifying molecular diagnoses. One study utilizes a modified GRCh38 version that accounts for reference errors involving 33 protein coding genes, of which 12 are medically relevant, and will assess the difference between the diagnostic yield of short read data to those data from long reads with this improved reference. Analyses of these approaches is challenging but is aided by tools such as VizCNV, an in-house developed analytical tool for detection, visualization, and interpretation of copy number variations and other products of structural variation mutagenesis. Major phenotypes studied include neurodevelopmental disorders, structural brain abnormalities, and intellectual disability. Thirty-five percent of this dataset includes a unique cohort from the Middle East and Northern Africa with a molecular diagnostic rate of at least 41%, including both novel and known genes. Cumulative analysis of the BCM-GREGoR dataset has yielded 22 manuscripts detailing discoveries in 230 genes of which 38% are novel. Data and methods are rapidly disseminated through AnVIL, GeneMatcher, VariantMatcher, ClinVar, and GitHub. Each discovery provides insight into the genomics of rare diseases and insights into human disease biology - bringing us a step closer to equitable personalized medicine.

Session Title: Mendelian Phenotypes Poster Session I

PB4644 Bi-allelic *ACBD6* variants lead to a neurodevelopmental syndrome with progressive complex movement disorders

Authors:

S-J. Lin[1], R. Kaiyrzhanov[2], A. Rad[3,4], A. Bertoli-Avella[5], W. W. Kallemeijn[6,7], A. Godwin[8], ACBD6 International Study Group, M. Guille[9], E. W. Tate[6,7], H. Houlden[2], R. Maroofian[2], G. K. Varshney[1]; [1]Oklahoma Med. Res. Fndn., Oklahoma City, OK, [2]UCL Inst. of Neurology, London, United Kingdom, [3]Eberhard Karls Univ., Tübingen, Germany, [4]Sabzevar Univ. of Med. Sci., Sabzevar, Iran, Islamic Republic of, [5]Centogene AG, Rostock, Germany, [6]Imperial Coll. London, Molecular Sci. Res. Hub, London, United Kingdom, [7]The Francis Crick Inst., London, United Kingdom, [8]Univ. of Portsmouth, European Xenopus Resource Ctr., Portsmouth, United Kingdom, [9]Univ. of Portsmouth, Portsmouth, United Kingdom

The acyl-CoA-binding domain-containing protein 6 (ACBD6) is involved in lipid and protein acylation and regulates protein *N*-myristoylation through *N*-myristoyltransferase enzymes (NMTs). However, the precise function of ACBD6 in cells and its impact on human pathophysiology remain unclear. To shed light on this, we studied 43 affected individuals from 27 families with bi-allelic pathogenic variants in *ACBD6*, using exome sequencing and international collaborations. Most variants were loss-of-function mutations (18/20). We generated acbd6 knockouts in zebrafish and Xenopus using CRISPR/Cas9 and investigated ACBD6's role in protein *N*-myristoylation in these models and human cells. We also examined ACBD6's peroxisomal localization in human cells. The affected individuals (21 males, 22 females) exhibited a complex and progressive disease with global developmental delay/intellectual disability (100%), expressive language impairment (97%), facial dysmorphism (94%), movement disorders (94%), mild cerebellar ataxia (85%), gait impairment (94%), limb spasticity/hypertonia (74%), oculomotor abnormalities (68%), behavioral abnormalities (63%), weight gain (59%), microcephaly (38%), and epilepsy (38%). Dystonia was the predominant movement disorder (90%), causing early-onset postural deformities (93%), limb dystonia (40%), and cervical dystonia (25%). Other common movement disorders included jerky tremors in upper limbs (61%), mild head tremor (56%), parkinsonism/hypokinesia with age (31%), and simple motor/vocal tics. Neuroimaging revealed midline brain malformations: corpus callosum abnormalities (67%), anterior commissure hypoplasia/agenesis (63%), short midbrain and small inferior cerebellar vermis (41% each), and clava hypertrophy (19%). Acbd6-deficient zebrafish and Xenopus models recapitulated patient phenotypes, including movement disorders, progressive neuromotor impairment, seizures, microcephaly, craniofacial dysmorphism, and midbrain defects with developmental delay and increased mortality. Unlike ACBD5, ACBD6 deficiency did not affect peroxisomal parameters in patient fibroblasts. Patient fibroblasts showed significant differences in labeling of 68 co- and 18 post-translationally *N*-myristoylated proteins using YnMyr-labelling. *N*-myristoylation was similarly affected in Acbd6-deficient zebrafish and Xenopus models, involving proteins linked to neurological diseases (FUS, MARCKS, CHCHD). This study establishes that bi-allelic pathogenic variants in *ACBD6* cause a distinct neurodevelopmental syndrome characterized by complex and progressive cognitive and movement disorders.

Session Title: Mendelian Phenotypes Poster Session II

PB4645 Biallelic *FLVCR1* variants cause a disease spectrum from adult neurodegeneration to severe neurodevelopmental disorders through disrupted choline transport

Authors:

D. Calame[1], V. Sukumar[2], D. T. Nguyen[2], H. Ha[2], J. Wong[2], P. Panda[2], S. Balamurugan[2], J. Chua[2], T. Q. Nguyen[2], A. Cazenave-Gassiot[2], J. Fatih[1], Z. Dardas[1], H. Du[3], Z. Coban Akdemir[4], F. Ammouri[5], B. Rezich[5], I. Herman[6], V. M. Srinivasan[7], V. K. Gowda[7], L. Rodan[8], L. AlAbdi[9], S. Isikay[10], D. Pehlivan[1], J. Posey[1], R. Gibbs[11], R. Maroofian[12], F. Alkuraya[9], J. R. Lupski[1], L. N. Nguyen[2]; [1]Baylor Coll. of Med., Houston, TX, [2]Natl. Univ. of Singapore, Singapore, Singapore, [3]BCM, Houston, TX, [4]UTHlth.Sch. of Publ. Hlth., Houston, TX, [5]Univ. of Nebraska Med. Ctr., Omaha, NE, [6]Boys Town Natl. Res. Hosp., Boys Town, NE, [7]Near NIMHANS, Bengaluru, India, [8]Boston Children's Hosp., Bostn, MA, [9]King Faisal Specialist Hosp. & Res. Ctr., Riyadh, Saudi Arabia, [10]Univ. of Gaziantep, Gaziantep, Turkey, [11]Baylor Coll. Med., Houston, TX, [12]Univ. Coll. London, London, United Kingdom

Background: FLVCR1 is a solute transporter implicated in cellular heme export and detoxification. *Flvcr1* knockout mice resemble Diamond-Blackfan anemia (DBA) with macrocytic anemia, craniofacial and digital deformities, and embryonic lethality. *FLVCR1* encodes mitochondrial and plasma membrane isoforms; both are required for murine viability. In contrast, biallelic *FLVCR1* variants cause autosomal recessive neurodegenerative phenotypes in humans: posterior column ataxia with retinitis pigmentosa and hereditary sensory and autonomic neuropathy (PCARP-HSAN). Recent evidence implicates FLVCR1 in the Kennedy pathway - a pathway responsible for *de novo* synthesis of the most abundant phospholipids in mammalian cells: phosphatidylcholine and phosphatidylethanolamine. The physiological significance of FLVCR1 in heme detoxification has been questioned.
Methods: To investigate apparent discrepancies between human and mouse *FLVCR1* phenotypes, we analyzed genomic data from >30,000 individuals with rare diseases for biallelic rare damaging *FLVCR1* variants. We performed molecular modeling of *FLVCR1* variants and developed a transport assay using FLVCR1-transfected HEK293 cells incubated with radiolabeled Kennedy pathway substrates. We examined the impact of *FLVCR1* missense variants on choline transport using the transport assay.
Results: We identified 20 individuals from 14 unrelated families with biallelic *FLVCR1* variants and severe developmental disorders (DD) distinct from PCARP-HSAN. Common findings included microcephaly, brain atrophy, profound developmental delay, epilepsy, and premature death. 8 individuals including 2 human knockouts exhibited additional findings seen in *Flvcr1* knockout mice and DBA [megaloblastic anemia, craniofacial malformations, and digital anomalies]. Severe DD associated missense alleles cluster in the protein transmembrane domains 9-11 and impact mitochondrial and plasma membrane isoforms. PCARP-HSAN associated missense variants are widely distributed and often only impact the plasma membrane isoform. We confirm FLVCR1 is a selective and highly conserved choline transporter through overexpression and knockdown studies. Furthermore, we show nearly all disease associated *FLVCR1* missense variants reduce choline transport.
Conclusions: The *FLVCR1* allelic series i) reconciles observed differences between mouse and human phenotypes, ii) demonstrates a broad human disease spectrum from adult-onset neurodegeneration to severe DD, and iii) suggests genotype-phenotype correlations. Ongoing research efforts will explore disease mechanisms and possible treatments for this disease.

Session Title: Mendelian Phenotypes Poster Session III

PB4646 Biallelic genetic variants in the translational GTPases GTPBP1 and GTPBP2 cause a distinct identical neurodevelopmental syndrome.

Authors:

V. Salpietro[1], R. Maroofian[1], M. S. Zaki[2], J. Wangen[3], A. Ciolfi[4], S. Barresi[4], F. Al Mutairi[5], A. Lamaze[1], S. Efthymiou[1], A. Rad[6], C. Rocca[1], E. Calì[1], A. Accogli[7], F. Zara[8], P. Striano[8], B. Dallapiccola[4], E. Faqeih[9], K. K. Tallur[10], M. Alfadhel[11], E. Alobeid[12], S. Maddirevula[13], K. Mankad[14], S. Banka[15], E. Ghayoor-Karimiani[16], M. Tartaglia[17], W. Chung[18], R. Green[3], F. Alkuraya[19], J. Jepson[1], H. Houlden[1]; [1]UCL Inst. of Neurology, London, United Kingdom, [2]Dept. of Clinical Genetics, Human Genetics and Genome Res. Inst., Natl. Res. Ctr., Cairo, Egypt, [3]Howard Hughes Med. Inst., Dept. of Molecular Biology and Genetics, Johns Hopkins Univ. Sch. of Med., Baltimore, MD, [4]Genetics and Rare Diseases Res. Div., Ospedale Pediatrico Bambino Gesù, Rome, Italy, [5]King Abdulaziz Med. City Riyadh, Riyadh, Saudi Arabia, [6]Arcensus GmbH, Rostock, Germany, [7]Div. of Med. Genetics, Dept. of Pediatrics, McGill Univ., Montreal, QC, Canada, [8]Istituto Giannina Gaslini, Genova, Italy, [9]King Fahad Med City, Children Specialist Hosp., Riyadh, Saudi Arabia, [10]Royal Hosp. for Sick Children, Edinburgh, United Kingdom, [11]King Abdullah Specialized Children Hosp., Riyadh, Saudi Arabia, [12]Dept. of Genetics, King Faisal Specialist Hosp. and Res. Ctr., Riyadh, Saudi Arabia, [13]KFHSRC, Riyadh, Saudi Arabia, [14]Great Ormond Street Hosp., London, United Kingdom, [15]Univ. of Manchester, Manchester, United Kingdom, [16]Genetics Res. Ctr., Molecular and Clinical Sci. Inst., St George's Univ. of London, London, United Kingdom, [17]Ospedale Pediatrico Bambino Gesù, Rome, Italy, [18]Columbia Univ., New York, NY, [19]King Faisal Specialist Hosp. & Res. Ctr., Riyadh, Saudi Arabia

Background: The homologous *GTPBP1* and *GTPBP2* genes encode GTP-binding proteins 1 and 2 that are involved in ribosomal homeostasis. Pathogenic variants in *GTPBP2* were recently shown to be an ultra-rare cause of neurodegenerative or neurodevelopmental disorders (NDDs). Until now, no human phenotype has been linked to *GTPBP1*. Here, we describe the first patients carrying biallelic *GTPBP1* variants, that display an identical phenotype with *GTPBP2,* and characterize the overall spectrum of GTP-binding protein (1/2)-related disorders. **Methods:** Twenty patients from 16 families with distinct NDDs and syndromic facial features were investigated by whole exome (WES) or genome (WGS) sequencing. To assess the functional impact of the identified genetic variants, semi-quantitative PCR, western blot and ribosome profiling assays were performed in patient-derived fibroblasts. *Drosophila melanogaster* lines harbouring hypomorphic mutations in *cg2017*, the fruit fly homolog of human *GTPBP1/2*, were also investigated. **Results:** Patients with biallelic *GTPBP1* or *GTPBP2* variants presented with microcephaly, pathognomonic craniofacial features and ectodermal defects. Neurodevelopmental impairment was profound in all cases. Abnormal vision and/or hearing, progressive spasticity, choreoathetoid movements, refractory epilepsy, and brain and cerebellar atrophy were part of the core phenotype of this syndrome. Cell lines studies identified a loss-of-function (LoF) impact of the disease-associated variants on both mRNA and protein levels, but no significant abnormalities on ribosome profiling assays. Reduced expression of *cg2017* isoforms was associated with locomotor impairment in *Drosophila*. **Conclusions**: Biallelic *GTPBP1* and *GTPBP2* LoF variants cause an identical, distinct and severe neurodevelopmental syndrome. Based on our observations, we propose to define this condition as **G**tpbp1/2 **R**elated **E**ctodermal **N**euro**D**evelopmental (GREND) syndrome. Our findings add insights into the consequences of biallelic LoF genetic defects in both GTPBP1 and GTPBP2, highlighting that 'GREND syndrome' can result from the loss of either homologous translational GTPase in humans. Neurological impairment of mutant flies highlight the conserved role for GTP-binding proteins in nervous system development across species.

Session Title: Mendelian Phenotypes Poster Session I

PB4647 Biallelic inactivating variants in *DMAP1* underlie a syndromic neurodevelopmental disorder

Authors:

D. Li[1], Q. Wang[1], C. Tirrito[1], L. Matsuoka[1], C. Applegate[2], F. Alkuraya[3], E. Bhoj[1], N. Couque[4], N. Giesbertz[5], A. Hamosh[2], T. Hjortshøj[6], M. Mori[7], M. Leung[7], K. Platzer[8], S. Redler[9], R. Sulaiman[3], A. Verloes[10], L. van der Laan[11], J. Kerkhof[11], H. McConkey[12], B. Sadikovic[11], A. Sobering[13], Y. Song[1], H. Hakonarson[1]; [1]Children's Hosp. of Philadelphia, Philadelphia, PA, [2]Johns Hopkins Univ., Baltimore, MD, [3]King Faisal Specialist Hosp. & Res. Ctr., Riyadh, Saudi Arabia, [4]CHU Robert debré, Reims, France, [5]Univ. Med. Ctr. Utrecht, Utrecht, Netherlands, [6]Univ. Hosp. of Copenhagen, Copenhagen, Denmark, [7]Nationwide Children's Hosp., Columbus, OH, [8]Inst. of Human Genetics, Univ. of Leipzig, Leipzig, Germany, [9]Univ. of Leipzig Med. Ctr., Leipzig, Germany, [10]Robert DEBRE Univ. Hosp., Paris, France, [11]LHSC, London, ON, Canada, [12]Western Univ., London, ON, Canada, [13]Augusta Unveristy / Univ. of Georgia Athens Med. Partnership, Athens, GA

DNA Methyltransferase 1 Associated Protein 1 (*DMAP1*) encodes a versatile protein involved in different complexes responsible for maintenance of DNA methylation, DNA damage repair, regulation of histone acetylation and catalysis of exchange of histone H2A and H2A.Z. Despite DMAP1's essential roles in multiple transcriptional processes, it has not been implicated in human disease. Through exome sequencing and subsequent reach out to the international matchmaking community, we identified ten individuals from nine families with a syndromic neurodevelopmental disorder carrying homozygous or compound heterozygous variants in *DMAP1*. Among these variants were three splice-altering or frameshift variants and seven missense variants residing in or around the SANT domain, suggesting they may affect interactions with DNA and/or histones. All ten individuals have global developmental delay, intellectual disability, hypotonia, and craniofacial dysmorphisms, although the reported findings varied. Detailed clinical assessment demonstrated that additional features, such as microcephaly, short stature, seizures, and brain malformation, were concomitant in at least three individuals. Utilizing the Gal4-UAS system to perform neural-specific knockdown of the *Drosophila* ortholog *Dmap1*, we observed pupal lethality and structural defects in the mushroom body (MB), highlighting an underappreciated role of Dmap1 in MB development. These phenotypes can be rescued by the wild-type (WT) and the two missense variants of human *DMAP1*, allowed us to conduct the social space assay to further query the high order effect in social behavior. We found that WT DMAP1 restored proper fly distribution, whereas the two missense variants caused clustering and reduced inter fly distance, reflecting impaired social avoidance and conforming the pathogenicity of these missense variants. Additionally, WT DAMP1 but not the two missense variants partially rescued the seizures induced by mechanical stimulation in a sex dependent manner. Transcriptome analyses from *Drosophila* RNAi brains identified dysregulation of hundreds of genes implicated in transcription processing, neuronal function, and brain development. In light of its involvement in DNA methylation, we performed an episignature analysis using blood-derived genomic DNA and identified a specific DNA methylation episignature. Further inversigations on DNA damage repair, histone H2A.Z deposition and histone acetylation are currently underway and will be presented. Taken together, we demonstrate that biallelic variants in *DMAP1* are associated with a novel neurodevelopmental disorder.

Session Title: Mendelian Phenotypes Poster Session II

PB4648 Biallelic loss-of-function variants in *CDK9* cause a CHARGE-like syndrome.

Authors:

D. Marafi[1], R. Alqusaimi[1], A. Othman[2], M. Alshammari[3], A. Saggar[4], W. Eyaid[5], A. Al-Zayed[6], S. Abdulwahhab[7], M. Rocha[8], E. Zonic[8], R. Duan[9], D. Calame[9], S. Omar[10], S. Maddirevula[11], J. Posey[9], A. Bertoli-Avella[12], J. R. Lupski[9], T. Ben Omran[13], F. S. Alkuraya[11]; [1]Kuwait Univ., Jabriya, Kuwait, [2]Adan Hosp., Ministry of Hlth., Hadiya, Kuwait, [3]King Saud Univ. Med. City, King Saud Univ., Riyadh, Saudi Arabia, [4]Intl. Gene Clinic, London, United Kingdom, [5]KING ABDULAZIZ Med. CITY, Riyadh, Saudi Arabia, [6]Coll. of Med., King Saud bin Abdulaziz Univ. for Hlth.Sci., Riyadh, Saudi Arabia, [7]Sidra Hosp., Doha, Qatar, [8]CENTOGENE GmbH, Rostock, Germany, [9]Baylor Coll. of Med., Houston, TX, [10]Kuwait Med. Genetic Ctr., Sabah Area, Kuwait, [11]King Faisal Specialist Hosp. & Res. Ctr., Riyadh, Saudi Arabia, [12]Centogene AG, Rostock, Germany, [13]Sidra Med., Doha, Qatar

Introduction: CHARGE syndrome (Coloboma, Heart anomalies, Atresia of choanae, Retarded growth and mental development, Genital anomalies, and Ear malformations and hearing loss) is a well-established disorder. Diagnostic criteria now mandate the presence of the 3C triad: coloboma, choanal atresia, and abnormal semicircular canals. Only one gene, *CDH7*, has an established association with autosomal dominant CHARGE syndrome. *CDK9* was proposed to cause an autosomal recessive CHARGE-like syndrome after four Arabian families with a homozygous *CDK9* variant [c.673C>T:p. (Arg225Cys)] were identified. One more family has been described since; yet, this disease-gene association remains unestablished. **Methods:** We used GeneMatcher and access to different clinical and research laboratories to identify subjects with biallelic *CDK9* variant alleles and systemically evaluate their phenotype. All subjects had exome sequencing followed by a segregation study. **Results:** We ascertained 13 individuals from 11 families with biallelic *CDK9* deleterious variants including two families with novel variants [c.1118_*18del and c.26A>G; p.(Glu9Gly)] and 9 families (5 new and 4 previously-published) from a large Arabian tribe with the previously described founder variant [c.673C>T:p. (Arg225Cys)]. All subjects had profound global developmental delay/intellectual disability; half had neonatal hypotonia, epilepsy and/or peripheral hypertonia. Dysmorphic features included webbed neck (100%), micrognathia and palatal abnormalities (91%), prominent nasal bridge (81%), curved eyebrows (75%), microcephaly (69%) and midfacial hypoplasia and choanal atresia (55%). Eye abnormalities (83%) included congenital cataract (58%), visual impairment (57%), microphthalmia (56%), and microcornea (33%). Macrotia was observed in 33% and hearing loss in 30%. A third of the subjects had cardiac anomalies and half had genitourinary anomalies including cryptorchidism in all 3 males and absent ovaries with uterine abnormalities in 33% of the female. Abnormal kidneys, hydronephrosis and ectopic ureter were observed in a third each. Neuroimaging findings included vermian hypoplasia (62%), delayed myelination (58%), and dysmorphic large ventricles (55%). None had the 3C triad with coloboma observed in only 17% and external ear abnormalities in 31%. **Conclusion:** Our data upgrade the link between *CDK9* and CHARGE-like syndrome from limited to moderate (GenCC), which allows clinical molecular reporting of this disease. This study underscores the importance of studying diverse populations to aid in disease gene discovery and enhance our understanding of human biology and disease.

Session Title: Mendelian Phenotypes Poster Session III

PB4649 Bi-allelic loss-of-function variants in Replication Factor C 4 (*RFC4*) are associated with a neurological disorder characterized by ataxia, muscular weakness, hearing impairment, and short stature.

Authors:

M. Morimoto[1], E. Ryu[2,3], B. J. Steger[1], N. Hauser[4], F. Kortuem[5], A. Dixit[6], A. van der Ven[5], P. J. Steinbach[7], L. A. Wolfe[1], C. Diamonstein[8], Undiagnosed Diseases Network, A. Y. Huang[9], H. Lee[10,11], M. Nehrebecky[1], C. Wahl[1], C. Lau[1], S. F. Nelson[10,11,12], D. R. Adams[1,13], W. A. Gahl[1,14], S. Kang[2], C. Toro[1], K. Myung[2,3], M. Malicdan[1,14]; [1]NIH Undiagnosed Diseases Program, Common Fund, Office of the Director, NIH, Bethesda, MD, [2]Ctr. for Genomic Integrity, Inst. for Basic Sci., Ulsan, Korea, Republic of, [3]Sch. of Life Sci., Ulsan Natl. Inst. of Sci. and Technology, Ulsan, Korea, Republic of, [4]Div. of Med. Genomics, Dept. of Pediatrics, Inova Hlth.System, Fairfax, VA, [5]Inst. for Human Genetics, UKE Hamburg-Eppendorf, Hamburg, Germany, [6]Nottingham Univ. Hosp., Nottingham, United Kingdom, [7]Bioinformatics and Computational BioSci.s Branch, Natl. Inst. of Allergy and Infectious Diseases, NIH, Bethesda, MD, [8]Simmons Comprehensive Cancer Ctr. Cancer Genetics, Univ. of Texas Southwestern Med. Ctr., Dallas, TX, [9]Inst. for Precision Hlth., David Geffen Sch. of Med., Univ. of California, Los Angeles, Los Angeles, CA, [10]Dept. of Pathology and Lab. Med., David Geffen Sch. of Med., Univ. of California, Los Angeles, Los Angeles, CA, [11]Dept. of Human Genetics, David Geffen Sch. of Med., Univ. of California, Los Angeles, Los Angeles, CA, [12]Dept. of Pediatrics, David Geffen Sch. of Med., Univ. of California, Los Angeles, Los Angeles, CA, [13]Office of the Clinical Director, Natl. Human Genome Res. Inst., NIH, Bethesda, MD, [14]Human Biochemical Genetics Section, Med. Genetics Branch, Natl. Human Genome Res. Inst., NIH, Bethesda, MD

Introduction: DNA replication is vital for cell division and repair, and its precise regulation is critical for maintaining genomic integrity. Errors in DNA replication contribute to the pathogenesis of several Mendelian DNA repair and replication disorders, as well as more common diseases such as cancer. The Replication Factor C (RFC) complex consists of five subunits and loads proliferating cell nuclear antigen (PCNA) onto DNA, facilitating the recruitment of replication and repair proteins and enhancing DNA polymerase processivity. While RFC1 dysfunction has been associated with Cerebellar Ataxia, Neuropathy, and Vestibular Areflexia Syndrome (CANVAS), the role of RFC2-5 in human disease remains unclear. In this study, we identified *RFC4*, which encodes one of the subunits of the RFC complex, as a candidate gene in five individuals with an undiagnosed neurological disorder.

Methods: Each proband underwent extensive clinical evaluation. To identify potential disease-causing genetic loci, we performed a combination of exome/genome sequencing, RNA sequencing, and family-based genomics. Collaborative variant re-analysis was conducted to refine the candidate variants, which were validated by genomic DNA Sanger sequencing. Splicing variants were confirmed by cDNA Sanger sequencing. *RFC4* gene expression and RFC4 protein levels were assessed by quantitative PCR and immunoblot analysis, respectively. Functional studies on protein stability and RFC complex formation were performed using a mutant RFC4 overexpression system.

Results: The affected individuals exhibited ataxia, muscular weakness, hearing impairment, and/or short stature, varying in severity and age at onset. Rare, conserved, likely pathogenic, bi-allelic variants in *RFC4* were identified in all affected individuals and segregated with disease. RNA sequencing and cDNA Sanger sequencing analyses confirmed that the splice site variants lead to exon skipping. All variants were predicted to alter or remove the C-terminal domain, which is indispensable for RFC complex formation. Further, RFC4 protein levels were significantly decreased in cultured patient fibroblasts compared to unaffected controls, and overexpression studies of each RFC4 variant demonstrated decreased protein stability and/or RFC complex formation.

Conclusion: Our findings from in silico analysis and functional assays, together with the overlapping clinical phenotypes observed in CANVAS due to bi-allelic *RFC1* repeat expansions or protein-truncating variants, provide compelling evidence that bi-allelic loss-of-function *RFC4* variants contribute to the pathogenesis of this new neurological disorder.

Session Title: Mendelian Phenotypes Poster Session I

PB4650 † Biallelic loss-of-function variants in *SLC13A1* are associated with skeletal dysplasia phenotypes in humans

Authors:

C. Tise[1], K. Ashton[2], L. De Hayr[3], E. Krzesinski[4], J. Bassetti[5], A. Zankl[6], A. Khanshour[7], J. Rios[7], C. Wise[7], T. Roscioli[2], M. Buckley[2], R. Harvey[3], P. Dawson[8]; [1]Stanford Univ., Palo Alto, CA, [2]NSW Hlth.Pathology, Sydney, Australia, [3]Univ. of the Sunshine Coast, Queensland, Australia, [4]Monash Hlth., Melbourne, Australia, [5]Weill Cornell Med., New York, NY, [6]The Children's Hosp. at Westmead, New South Wales, Australia, [7]Texas Scottish Rite Hosp. for Children, Dallas, TX, [8]Mater Res. Inst., Brisbane, Australia

Sulfate (SO_4^{2-}) is necessary for a variety of physiological processes, including structural and functional maintenance of macromolecules and the formation of sulfur-containing compounds. The key role of sulfate in skeletal growth and development is exemplified by the abnormal bone and cartilage phenotypes of individuals with certain forms of skeletal dysplasias, including diastrophic dysplasia caused by biallelic pathogenic variants in the sulfate transporter gene *SLC26A2*. In this study, we explored a novel association between variants in *SLC13A1* and skeletal phenotypes. SLC13A1 is a sodium-sulfate cotransporter primarily expressed in the kidney where it mediates sulfate reabsorption and maintenance of circulating sulfate levels to meet demands of other tissues. Previous reports described two rare, unlinked nonsense variants in *SLC13A1*, NM_022444.3:c.34C>T[p.R12*] and NM_022444.3:c.144G>A[p.W48*], associated with a ~30% reduction in serum sulfate level in heterozygotes (0.25 mM vs. 0.36 mM in controls) without any obvious physical phenotype. Given the hyposulfatemia and osteochondrodysplasia observed in animals with biallelic loss-of-function (LOF) variants in *Slc13a1*, a skeletal phenotype in humans with biallelic LOF variants has been suspected. Through use of trio exome sequencing and GeneMatcher (https://genematcher.org), a compilation of six children from four families with biallelic variants in *SLC13A1* and a skeletal dysplasia phenotype has emerged. All presented with short stature and were found to have abnormal radiological features, including scoliosis and/or a (spondylo)epimetaphyseal dysplasia. Identified *SLC13A1* variants include the aforementioned nonsense variants and three additional missense variants, NM_022444.3:c.1343G>A[p.G448D], NM_022444.3:c.1547T>C[p.L516P], and NM_022444.4:c.1744T>C[p.Y582H]. Genotypes include homozygous nonsense (n=1), homozygous missense (n=1), and compound heterozygosity for nonsense/missense variants (n=4). The serum sulfate level in the individual with a homozygous missense variant was proportionately reduced compared to individuals with heterozygous *SLC13A1* nonsense variants (0.11 mM vs. 0.25 mM), consistent with these representing LOF alleles of similar effect size. This work defines biallelic LOF variants in *SLC13A1* as a novel cause of skeletal dysplasia phenotypes in humans. Functional studies are underway to characterize the impact of these *SLC13A1* variants on sulfate transport. These findings also highlight the importance of considering sulfate measurements in cases of unsolved short stature with or without obvious skeletal dysplasia features.

Session Title: Mendelian Phenotypes Poster Session II

PB4651 Biallelic *PSTK* variants disrupt selenoprotein synthesis and cause a multisystemic disorder.

Authors:

E. Burke[1], E. F. Macnamara[1], T. S. Chillon[2], J. Bang[3], P. D'souza[1,4], M. J. Elrick[5,6], P. J. Steinbach[7], V. N. Gladyshev[3,8], L. Schomburg[2], W. A. Gahl[1,9], C. J. Tifft[1,4], M. C. V. Malicdan[1,9]; [1]NIH Undiagnosed Diseases Program, Common Fund, Office of the Director, NIH, Bethesda, MD, [2]Inst. for Experimental Endocrinology, Charité-Univ.smedizin Berlin, Corporate Member of Freie Univ. Berlin, Humboldt-Univ. zu Berlin, And Berlin Inst. of Hlth., Berlin, Germany, [3]Div. of Genetics, Dept. of Med., Brigham and Women's Hosp., Harvard Med. Sch., Boston, MA, [4]Office of the Clinical Director, Natl. Human Genome Res. Inst., NIH, Bethesda, MD, [5]Pediatric Neurology and Neuromuscular Med., Johns Hopkins Sch. of Med., Baltimore, MD, [6]Neurology & Dev.al Med., Kennedy Krieger Inst., Baltimore, MA, [7]Bioinformatics and Computational BioSci.s Branch, Natl. Inst. of Allergy and Infectious Diseases, NIH, Bethesda, MD, [8]Broad Inst., Cambridge, MA, [9]Human Biochemical Genetics Section, Med. Genetics Branch, Natl. Human Genome Res. Inst., NIH, Bethesda, MD

Background: Selenoproteins are a distinct class of proteins that contain selenocysteine as a unique amino acid and are involved in a wide range of biological functions, including antioxidant defense, redox regulation, and thyroid hormone metabolism. The highly regulated selenoprotein biosynthesis pathway plays a vital role in the incorporation of the essential micronutrient selenium as selenocysteine (Sec). One critical enzyme in this pathway is phosphoseryl-tRNA$^{[Ser]Sec}$ kinase (PSTK), which phosphorylates seryl-tRNA$^{[Ser]Sec}$. This phosphorylated tRNA facilitates the delivery of selenium to the ribosome for the incorporation of selenocysteine into selenoproteins. Although human diseases have been associated with other key genes in the selenoprotein biosynthesis pathway, including *SECISBP2*, *TRU-TCA-1*, and *SEPSECS*, no human condition has been associated with *PSTK*. Here, we report compound heterozygous variants in *PSTK* in a 2 y/o proband with severe refractory epileptic encephalopathy.
Methods: The proband underwent comprehensive clinical and biochemical evaluation, including research reanalysis of genome sequencing data. Functional validation of the identified variants was performed on patient's biological samples and primary cell lines. Oxidative stress assays were conducted, and selenium and antioxidant therapy were also initiated.
Results: The proband had hypotonia, global developmental delay with regression, sensorineural hearing loss, acquired microcephaly, and seizures, which became refractory and led to encephalopathy and multisystem compromise. Genome sequencing reanalysis identified compound heterozygous variants in *PSTK*. Biochemical studies revealed abnormal thyroid hormone levels, low circulating selenium, and undetectable glutathione peroxidase activity and selenoprotein P levels, consistent with a defect in selenoprotein biosynthesis. The patient's fibroblasts showed poor survival in the absence of high-dose vitamin E supplementation. High levels of reactive oxygen species were detected, suggesting the presence of significant oxidative stress and the potential involvement of ferroptosis.
Conclusion: Collectively, these findings suggest that biallelic *PSTK* variants result in selenoprotein deficiency, contributing to the pleiotropic phenotype observed in this individual. This case highlights the critical role of PSTK in maintaining selenium homeostasis and underscores the importance of selenoproteins in neurological function.

Session Title: Mendelian Phenotypes Poster Session III

PB4652 Biallelic structural variations within *FGF12* detected by long-read sequencing in epilepsy

Authors:

N. Matsumoto, A. Fujita, S. Ohori; Yokohama City Univ. Graduate Sch. of Med., Yokohama, Japan

We discovered biallelic intragenic structural variations (SVs) in *FGF12* by applying long-read whole genome sequencing (LRWGS) to an exome-negative patient with development and epileptic encephalopathy (DEE). We also found another DEE patient carrying a biallelic (homozygous) single nucleotide variant (SNV) in FGF12 that was detected by exome sequencing. *FGF12* heterozygous recurrent missense variants with gain-of-function or heterozygous entire duplication of *FGF12* are known causes of epilepsy, but biallelic SNVs/SVs have never been described. *FGF12* encodes intracellular proteins interacting with the C-terminal domain of the alpha-subunit of voltage-gated sodium channels 1.2, 1.5, and 1.6, promoting excitability by delaying fast inactivation of the channels. To validate the molecular pathomechanisms of these biallelic *FGF12* SVs/SNV, highly sensitive gene expression analyses using lymphoblastoid cells from the patient with biallelic SVs, structural considerations, and Drosophila in vivo functional analysis of the SNV were performed, confirming loss-of-function. Our study highlights the importance of small SVs in Mendelian disorders, which may be overlooked by exome sequencing but can be detected efficiently by LRWGS, providing new insights into the pathomechanisms of human diseases.**Acknowledgments**Akihiko Miyauchi, Hitoshi Osaka, Charles Marques Lourenco, Naohiro Arakaki, Toru Sengoku, Kazuhiro Ogata, Rachel Sayuri Honjo, Chong Ae Kim, Satomi Mitsuhashi, Martin C. Frith, Rie Seyama, Naomi Tsuchida, Yuri Uchiyama, Eriko Koshimizu, Kohei Hamanaka, Kazuharu Misawa, Satoko Miyatake, Takeshi Mizuguchi, and Kuniaki Saito are greatly appreciated for their excellent contribution to this study.

Session Title: Mendelian Phenotypes Poster Session I

PB4653 Biallelic variant in *TGFBR3* is implicated in human craniosynostosis.

Authors:

R. M. Elshafie[1], D. J. Bernard[2], Y-F. Lin[2], I. Pecora[2], E. R. Buddle[2], D. Marafi[1,3], A. Elshafey[1], A. Alshammari[1], H. Alsharhan[1,3], D. Ibrahim[1], L. Bastaki[1]; [1]Kuwait Med. Genetics Ctr., Ministry of Hlth., Sulaibikhat 80901, Kuwait, [2]Dept. of Pharmacology and Therapeutics, McGill Univ., Montreal, QC, Canada, [3]Dept. of Paediatrics, Faculty of Med., Kuwait Univ., Jabriya, P.O. Box 24923, 13110 Safat, Kuwait

Background: Fusion of cranial sutures is a pre-programmed orchestrated process. Disruption of this process in the underlying signaling pathways results in premature fusion and ossification of sutures (craniosynostosis). Knowledge of the molecular mechanisms underlying craniosynostosis remains limited. The transforming growth factor ß (TGF-ß) superfamily is signaling molecules regulating various biological functions in connective and other tissues. Until now, no clear association between the TGF-ß family and a known form of craniosynostosis was established; however, previous studies demonstrated high expression of TGF-ß1, 2, and 3 in cranial sutures and dura matter. The TGF-ß type III receptor (TGFBR3) is a transmembrane proteoglycan that binds TGF-ß isoforms. Half of *Tgfbr3* knockout mouse embryos revealed a generalized reduction in the size and ossification of the craniofacial and appendicular skeletons. No definitive human phenotypes have been associated with *TGFBR3* mutations. Here, we report two Kuwaiti brothers with craniosynostosis and a homozygous variant in *TGFBR3*.Material-Methods: We performed exome sequencing reanalysis on two brothers with craniosynostosis followed by a family segregation study of the identified *TGFBR3* variant. The variant was then introduced into a human *TGFBR3* expression vector using site-directed mutagenesis to study its effect in vitro. Results: We identified a novel homozygous nonsense variant in *TGFBR3* [NM_003243.4: c.2418G>A :p. (Trp806Ter)] in the two brothers. A segregation study for the family confirmed Mendelian autosomal recessive inheritance. Both brothers had posterior sagittal and lambdoid synostosis, undescended testes, hypospadias, and congenital hip dislocation. Mutant *TGFBR3* was expressed at the plasma membrane of heterologous HEK293T cells but was notably smaller due to the loss of its intracellular C-terminus. Mutant TGFBR3 was capable of potentiating responses to TGF-ß1, 2, and 3 in cell-based reporter assays. Collectively, the data show that premature stop codon leads to truncation of TGFBR3 but does not appear to affect its production, subcellular localization, or ability to mediate canonical signaling via the three TGF-ß isoforms. Additional in vitro studies are undergoing to evaluate the potential deleterious effects on non-canonical signaling mediated via the C-terminus of the protein. Conclusion: In conclusion, our study demonstrates the potential association of TGFBR3 signaling dysregulation with syndromic craniosynostosis and is the first report to elucidate such an observation. Identification of additional cases is important to further consolidate our findings.

Session Title: Mendelian Phenotypes Poster Session II

PB4654 † Biallelic variants in *ANKRD6* are a candidate cause for a rare multisystemic congenital anomaly disorder

Authors:

S. Khan[1,2], T. Arbogast[3], L. Ellen[4], H. Cope[4], B. Brita[5], E. Davis[1,6]; [1]Ann & Robert H. Lurie Children's Hosp. of Chicago, Chicago, IL, [2]Human Molecular Genetics Lab., Hlth.Biotechnology Div., Natl. Inst. for Biotechnology and Genetic Engineering (NIBGE), Pakistan., Fasialabad, Pakistan, [3]Ann & Robert H. Lurie children s Hosp. of Chicago, Chicago, IL, [4]Ctr. for Human Disease Modeling, Duke Univ., Durham, NC, [5]Fetal Diagnostic Ctr., Duke Univ. Med. Ctr., Durham, NC, [6]Dept. of Pediatrics and Cell and Dev.al Biology, Fienberg Shchool of Med., Northwestren Univ., Chicago, IL

Wnt signaling regulates crucial aspects of cell fate determination, cell migration, cell polarity, neural patterning, and organogenesis during embryonic development. Diversin, encoded by *ANKRD6* (ankyrin repeat domain-containing 6), acts as a molecular switch between canonical and non-canonical Wnt signaling (planar cell polarity, PCP). As part of our ongoing efforts to identify the genetic basis of congenital anomalies in neonates and children, we performed trio exome sequencing (ES) in healthy parents and a prenatally referred female with multiple congenital anomalies, consistent with VACTERL. Specifically, she has laterality defects, renal agenesis, esophageal atresia, tracheoesophageal fistula, and complex cardiovascular phenotypes, and was enrolled in our research study at 5 days of age. We identified rare compound heterozygous nonsynonymous variants (NM_014942: c.38G>A, p.Arg13His and c.1880G>A, p.Arg627His) in *ANKRD6* that segregate with disease. To establish the physiological relevance of disease phenotypes, we generated CRISPR-Cas9 mutant and transient suppression zebrafish models. The F0 mosaic mutants recapitulate phenotypes observed in the proband, including laterality phenotypes and renal hypoplasia. We validated further the phenotypes observed in our F0 mutants by transient suppression with a splice-blocking morpholino. We evaluated convergent extension and left-right asymmetry defects and observed that morphants displayed shortened body axes, broadening, and thinning of somites, undulating notochords, and a decreased size of anterior structures with concomitant left-right asymmetric defects when compared to controls. Furthermore, *in vivo* complementation assays suggest that p.Arg13His and p.Arg627His variants confer hypomorphic and functional null effects on protein function, respectively. Together, our genetic and functional data implicate *ANKRD6* as an ultra-rare candidate gene for a VACTERL-like syndrome, which highlights a critical role of Wnt/PCP in human development and disease.

Session Title: Mendelian Phenotypes Poster Session III

PB4655 Biallelic variants in ARHGAP19 cause a motor-predominant neuropathy with conduction slowing

Authors:

S. Efthymiou[1], N. Dominik[1], C. Record[1], X. Miao[2], ARHGAP19 Clinical Consortium, D. Osborn[3], J. Jepson[1], N. Lamarche-Vane[2], M. Hanna[1], M. Reilly[1], H. Houlden[1]; [1]UCL Inst. of Neurology, London, United Kingdom, [2]Dept. of Anatomy and Cell Biology, McGill Univ.,, Montréal, QC, Canada, [3]St George Univ. of London, London, United Kingdom

Rho GTPases are members of the large superfamily of small GTPase proteins considered as molecular switches in various cellular events. One of the major regulators of Rho GTPases are Rho GTPase-activating proteins (GAPs). RhoGAPs stimulate intrinsic GTPase activity of Rho GTPases therefore acting as negative regulators of Rho pathway. One of the Rho effectors, the serine/threonine protein kinase ROCK, has important role in actin organisation, cell migration regulation, cell cycle control, and cell adhesion. By using next generation sequencing we identified 17 individuals from 16 unrelated families with biallelic variants in Rho GTPase-activating protein 19 (ARHGAP19) presenting with young age of onset, mild to moderate motor predominant, length-dependent neuropathy with sensory symptoms. Nerve conduction studies reveal chronic denervation and a predominantly axonal neuropathy. We are using various approaches to model these variants; *in-vitro* GAP assays to assess if the GAP activity is affected by expression of proteins carrying ARHGAP19 mutations, complemented by an *in-vivo Drosophila melanogaster* and *Danio rerio* models to test for movement, lifespan and neuromuscular junction integrity; in silico approach to gain an understanding of protein structure changes and its implications. Ongoing studies such as the *in-vitro* GAP assays show that ARHGAP19 has GAP activity towards RhoA but not Rac1 or Cdc42. Three of the mutations found in patients are being tested for their GAP activity and preliminary data suggest a loss of the GAP activity in a frameshift mutations. Visualisation of the endogenous expression pattern of ARHGAP19 ortholog in fly, RhoGAP54D, suggest the protein is expressed in perineural or subperineural glia in the fly brain. Preliminary results indicate that RNAi knockdown of RhoGAP54D in flies reduces both overall movement and startle responses to light-dark transitions. This is a first association of ARHGAP19 with neurological disease and deep phenotyping analysis in conjunction with the in-vivo animal model and the *in-vitro* GAP assay will help highlight he importance of the gene in early human brain development and function.

Session Title: Mendelian Phenotypes Poster Session I

PB4656 Biallelic variants in *BECN1*, an autophagy effector, are associated with a rare neurodevelopmental syndrome

Authors:

F. Ullah[1,2], V. Salpietro[3], M. Coghlan[4], A. M. Sharma[4], A. Mercier[5], A. McRae[5], H. Houlden[6], E. Davis[7]; [1]Lurie Children's Hosp. of Chicago, Chicago, IL, [2]Feinberg Sch. of Med., Northwestern Univ., Chicago, IL, USA., Chicago, IL, [3]UCL, London, United Kingdom, [4]Div. of Neonatology, Dept. of Pediatrics, Ann & Robert H. Lurie Children's Hosp. of Chicago, Northwestern Univ. Feinberg Sch. of Med., Chicago, Illinois., Chicago, IL, [5]Div. of Genetics, Birth Defects and Metabolism, Ann & Robert H. Lurie Children's Hosp. of Chicago, Chicago, IL, USA, Chicago, IL, [6]Inst Neurology, London, United Kingdom, [7]Lurie Children`s Hosp. of Chicago, Dept.s of Pediatrics and Cell and Dev.al Biology, Feinberg Sch. of Med., Northwestern Univ., Chicago, IL, USA., Chicago, IL

Autophagy is a hallmark cellular process involved in the elimination of cytoplasmic materials and degradation of macromolecules via lysosomal machinery. Although impaired autophagy has garnered longstanding attention for involvement in neurodegenerative disorders, pathogenic variants in pathway effectors such as *VPS15* and *ATG7* have more recently emerged as rare contributors to neurodevelopmental disease. As part of our ongoing efforts to identify new causes for rare fetal and pediatric onset central nervous system anomalies, we performed parent-proband exome sequencing and chromosomal microarray in affected individuals from three unrelated families and found rare variants in *BECN1*, encoding Beclin 1. Beclin 1 is a key component of the PI3K-III complex involved in the initiation, recruitment, and activation of autophagosome formation. Ablation of *Becn1* in mice leads to preweaning lethality, but the locus has not been implicated previously in human phenotypes. In family 1, two affected sisters presented with prenatal findings of severe ventriculomegaly, macrocephaly, callosal anomalies, dysplastic cerebellum, and hypoplasia of the vermis, associated postnatally with hypoventilation/apnea ultimately leading to death at <3 months of age. We detected rare, compound heterozygous variants in *BECN1* (p.Cys375Arg and p.Gln326*) co-segregating with the disease. The family 2 proband, harboring a homozygous p.Gly107Ser variant, has macrocephaly, intellectual disability, dysgenesis of the corpus callosum and cortical malformation. The family 3 proband, who has a biallelic p.Gly355Arg change, displays developmental delay and intellectual disability. To assess the functional consequences of *BECN1* biallelic variants, we generated stable mutant and transient suppression zebrafish models. Zebrafish larvae with *becn1* loss of function show a significantly increased head size and a reduced number of commissural axons as proxies for macrocephaly and corpus callosum defects, respectively. We also observe a concomitant increase in cell death in anterior structures in the context of *becn1* loss, and larval lethality in homozygous mutants carrying a frameshifting variant. *In vivo* complementation assays confirm that the missense variants are either loss of function or hypomorphic. Finally, we accompanied our *in vivo* studies with biochemical assays in patient-derived fibroblasts to show that *BECN1* variants induce aberrant AKT-mTOR signaling. Taken together, these data suggest a critical role for Beclin 1 and autophagy-related functions in brain development across different species.

Session Title: Mendelian Phenotypes Poster Session II

PB4657 Biallelic variants in *EPB41L3* underlie a neurodevelopmental disorder with seizures and myelination defects

Authors:

L. Werren[1], P. Majethia[2], L. Burglen[3], A. Afenjar[4], H. Tomoum[5], D. Florence[6], R. Maroofian[7], S. Siddiqui[8], A. Sevilimedu[9], Q. Padiath[10], Y-C. Tsan[1], A. Shukla[11], S. Bielas[1]; [1]Univ. of Michigan, Ann Arbor, MI, [2]Kasturba Med. Coll., Manipal, Udupi, India, [3]APHP-Hosp. Trousseau, 75012 Paris, France, [4]Univ Lausanne, Paris, France, [5]Ain Shams Univ., Cairo, Egypt, [6]Vannes Hosp., Vannes, France, [7]UCL Queen Square Inst. of Neurology, London, United Kingdom, [8]STAR Inst. of NeuroSci.s, Hyderabad, India, [9]Univ. of Hyderabad Campus, Hyderabad, India, [10]Univ. of Pittsburgh, Pittsburgh, PA, [11]Kasturba Med. Coll., Manipal, MAHE, Manipal, Manipal, Udupi, India

In neuronal tissues, the protein encoded by the *EPB41L3* gene, commonly referred to as 4.1B, has been found to play important roles in neuronal development, myelination, and cytoskeletal organization. The close relationship between neurons and oligodendrocytes is essential for neuronal function. Cell-type specificity in expression of 4.1B isoforms is important for both neurons as well as oligodendrocytes during development, which challenges current models of 4.1B function restricted to nodes of Ranvier. Here, we describe six individuals with novel biallelic *EPB41L3* variants (two nonsense and three frameshift) who present with global developmental delay, intellectual disability, epilepsy, hypotonia, and delayed myelination. This represents the first clinical description of an autosomal recessive disorder associated with variants in *EPB41L3*, which we refer to as *EPB41L3*-associated developmental disorder (EADD). The clinical features of EADD do not phenocopy the *Epb41l3* knockout mouse models. To investigate pathogenic mechanisms, we utilize patient-derived fibroblasts and neural progenitor cell lines from human embryonic stem cells harboring patient variants. Our findings indicate greater involvement of oligodendrocyte 4.1B in human EADD neuropathology than has been implicated in *Epb41l3* mouse models. Together, our findings identify *EPB41L3* variants as the basis of a newly described neurodevelopmental disorder.

Session Title: Mendelian Phenotypes Poster Session III

PB4658 Biallelic variants in GET4 cause developmental disabilities with a variable degree of phenotypic spectrum

Authors:

A. Fatima[1], G. Banneau[2], S. Hashami[1], J. Iwaszkiewicz[3], R. Ibrahim[1], M. Jameel[1], A. Khan[4], S. Baig[5], M. Toft[6], Z. Iqbal[7]; [1]Aga Khan Univ., Karachi, Karachi, Pakistan, [2]CHU Toulouse, France, Toulouse, France, [3]SIB Swiss Inst. of Bioinformatics, Lausanne, Switzerland, [4]Natl. Inst. for Biotechnology and Genetic Engineering (NIBGE), Faisalabad, Pakistan, [5]Aga Khan Univ., karachi, Pakistan, [6]Oslo Univ. Hosp, Oslo, Norway, [7]Oslo Univ. Hosp., RH, Oslo, Norway

GET4 is one of the six components of the transmembrane domain recognition complex (TRC) pathway. TRC directs the tail-anchored proteins to the appropriate cellular membranes. A single patient with compound heterozygous variant in the gene GET4 has been reported with global developmental delay, intellectual disability (ID) and seizures. We studied a five-generation Pakistani consanguineous family having four affected members presenting with variable degree of developmental disability including ID, seizures and developmental delay. Through whole exome sequencing, we identified a rare homozygous missense variant c.803G>A; p.R268Q (NM_015949.3) in GET4 gene which segregated with the phenotype. By GeneMatcher, we found another homozygous missense variant c.326C>G; p.A109G in GET4 in a patient of a consanguineous family. The patient presented with global developmental delay, and aspecific dysmorphia. In silico molecular modelling of the identified variants revealed that R268 residue's side chain is forming a hydrogen bond with the Q1035 of BAG6 gene and p.R268Q variant in GET4 possibly influences this interaction. It was predicted that this variant could lead to destabilization of the local protein structure. The p.A109G variant was predicted to decrease the stability of GET4 protein by destabilizing the alpha-helix structure, as Gly is known as a helix breaker. Moreover, we have initiated some functional experiments to study possible defects in TRC pathway and to prove the pathogenicity of these variants.In conclusion, we found two homozygous missense variants, p.R268Q and p.A109G in GET4 causing variable degree of developmental disability by plausibly influencing the stability and the structure of the GET4 protein.

Session Title: Mendelian Phenotypes Poster Session I

PB4659 Biallelic variants in *IFT57* may cause Bardet-Biedl Syndrome.

Authors:

A. Nitoiu[1], Q. Zhang[2], M. Rashid[1], E. Tavares[1], K. Ahmed[1], V. Sheffield[2], A. Vincent[1], E. Heon[1]; [1]The Hosp. for Sick Children (SickKids), Toronto, ON, Canada, [2]Univ. of Iowa, Iowa City, IA

Primary cilia are critical components of most cells, and are involved in many cell signalling pathways that regulate cellular and organ development. When this process goes wrong, patients develop ciliopathies; a genetically and clinically heterogenous group of diseases with a wide range of morbidities. Bardet Biedl syndrome (BBS) is a genetically heterogeneous ciliopathy with retinal degeneration, among other morbidities. The connecting cilium of the photoreceptor is a modified primary cilium sharing proteins and pathways with primary cilia. Approximately 20% of BBS cases remain genetically unsolved. We studied an unsolved case using whole genome sequencing (WGS). Results: A 29-year old male with clinical features suggestive of BBS had negative clinical genetic testing. WGS and comprehensive filtering identified biallelic *IFT57* variants in *trans* (NM_018010.4). The variants included c.675delA; p.(Lys225Asnfs*17), a frameshift deletion in exon 6 of chromosome 3 causing loss of function, and a c.1190T>A:p.(Val397Glu) missense variant in exon 11 of chromosome 3, a variant of unknown significance (VUS). Using patient-derived skin fibroblasts, immunofluorescence revealed a lower cilia count and abnormal cilia appearance. Using IMCD3 *IFT57* knockout cells showed that either a human or mouse Val397Glu variant can rescue ciliogenesis defects, but not in RPE1 knockout *IFT57* cells, suggesting a cell autonomous process. Conclusions: These results suggest that the missense variant Val397Glu acts as a hypomorph and together with the knockout variant likely causes our patient's observed ciliopathy phenotype. This adds complexity to the BBS genetic heterogeneity and demonstrates the value of further exploring VUSs.

Session Title: Mendelian Phenotypes Poster Session II

PB4660 Biallelic variants in SLC4A10 encoding the sodium-dependent chloride-bicarbonate exchanger NCBE lead to neurodevelopmental disorder

Authors:

R. Kaiyrzhanov[1], R. Maroofian[2], M. Zamani[3], L. Liebmann[4], SLC4A10 study group, H. Houlden[5], C. A. Hübner[4]; [1]South Kazakhstan Med. Academy, Shymkent, Kazakhstan, [2]UCL, London, United Kingdom, [3]Shahid Chamran Univ., Faradonbeh, Iran, Islamic Republic of, [4]Inst. of Human Genetics, Jena Univ. Hosp., Friedrich Schiller Univ., Jena, Germany, [5]Inst Neurology, London, United Kingdom

Background: *SLC4A10* encodes a plasma membrane-bound ion transport protein, which mediates Na^+-dependent HCO_3^- import thus mediating acid extrusion. *Slc4a10* knockout (KO) mice show collapsed brain ventricles, an increased seizure threshold, mild behavioral abnormalities, impaired vision, and deafness. Pathogenic variants in *SLC4A10* have hitherto not been associated with monogenic human disease traits. **Methods:** Utilizing proband-only or trio-exome/genome sequencing in families with undiagnosed neurodevelopmental disorders (NDD) and extensive international data sharing, 11 patients from 6 independent families with biallelic predicted-damaging variants in *SLC4A10* were identified. Clinical features, neuroimaging data, and facial images were reviewed by a group of clinical geneticists, neurologists, and a pediatric neuroradiologist. A minigene assay, intracellular pH recordings, and protein modelling were performed to study the possible functional consequences of the variant alleles. **Results**: The worldwide cohort harbours 8 rare/ultra-rare biallelic *SLC4A10* variants (7 missense and 1 splicing), which segregate with the disease within the families. Patients present with a clinical spectrum of global developmental delay/intellectual disability and central hypotonia associated with variable speech delay, microcephaly, cerebellar ataxia, epilepsy, and facial dysmorphism. Neuroimaging features are ranging from some non-specific to distinct neuroradiological findings, including slit ventricles and a peculiar form of bilateral curvilinear nodular heterotopia. All missense variants affect evolutionary highly conserved residues and are predicted to be damaging according to various *in silico* analyses. Intracellular pH recordings supported the disease-causing effect for 4/7 missense variants. **Conclusion**: We provide evidence that pathogenic biallelic *SLC4A10* variants can lead to NDD, which is characterized by variable abnormalities of the central nervous system including altered brain ventricles thus resembling several features observed in KO mice. Our study further highlights the challenges of assessing sequence variants identified in patients and distinguishing pathogenic variants from other potentially functional variants.

Session Title: Mendelian Phenotypes Poster Session III

PB4661 Biochemical Screening of Intellectually Disabled Patients A Stepping Stone to Initiate a Newborn Screening Program in Pakistan

Authors:

M. Wasim[1,2], H. Khan[1,2], H. Ayesha[3], S. Goorden[4], F. Vaz[4], C. Karnebeek[4], F. Arain[1], F. Awan[2]; [1]The Aga Khan Univ., Karachi, Pakistan, [2]Natl. Inst. for Biotechnology and Genetic Engineering, Faisalabad, Pakistan, [3]Faisalabad Med. Univ., Faisalabad, Pakistan, [4]Univ. of Amsterdam, Amsterdam, Netherlands

Inborn errors of metabolism (IEMs) are rare group of genetic disorders comprising of more than 1,000 different types. Around 200 of IEMs are potentially treatable through diet, pharmacological and other therapies, if diagnosed earlier in life. IEMs can be diagnosed early through newborn screening (NBS) programs, which are in place in most of the developed countries. However, establishing a NBS in a developing country is a challenging task due to scarcity of disease related data, large population size, poor economy, and burden of other common disorders. Since, not enough data is available for the prevalence of IEMs in Pakistan; therefore, in this study, we set out to find the prevalence of various treatable IEMs in a cohort of intellectually disabled patients suspected for IEMs, which will help us to initiate a NBS program for the most frequent IEMs in Pakistan. Therefore, a total of 429 intellectually disabled (IQ < 70) patient samples were collected from Pakistan. A subset of 113 patient samples was selected based on the clinical information for the detailed biochemical screening. Advance analytical techniques like, Amino Acid Analyzer, GC-MS, UHPLC-MS, and MS/MS were used to screen for different treatable IEMs like aminoacidopathies, fatty acid b-oxidation disorders and mucopolysaccharidoses (MPS) etc. A total of 14 patients were diagnosed with an IEM i.e., 9 with homocystinuria, 2 with MPS, 2 with Guanidinoacetate methyltransferase (GAMT) deficiency and 1 with sitosterolemia. These IEMs are found frequent in the collected patient samples from Pakistan. Thus, present study can help to take an initiative step to start a NBS program in Pakistan, especially for the homocystinuria having highest incidence among aminoacidopathies in the studied patients, and which is amenable to treatment. This endeavor will pave the way for a healthier life of affected patients and will lessen the burden on their families and society.

Session Title: Mendelian Phenotypes Poster Session I

PB4662 † Breeding scheme affects mouse model for atypical type VI osteogenesis imperfecta with*Ifitm5*/BRIL p.S42L

Authors:

E. Evans[1], M. Jovanovic[1], T. Hefferan[2], C. Locke[3], A. Barnes[1], K. Kozloff[3], J. Marini[1]; [1]Natl. Inst. of Child Hlth.and Human Dev., Bethesda, MD, [2]Mayo Clinic, Rochester, MN, [3]Univ. of Michigan, Dept.s of Orthopedic Surgery and BioMed. Engineering, Ann Arbor, MI

Background: Osteogenesis Imperfecta (OI) or brittle bone disease, is a genetic connective tissue disorder with major findings of bone fractures and deformities, and short stature. OI is a collagen-related disorder, with classical OI types caused by defects in type I collagen structure or quantity, while most rare types are caused by defects in molecules that interact with collagen. Type V, caused by recurrent dominant mutation in *IFITM5*/BRIL, and type VI OI, caused by recessive null mutations in *SERPINF1*/PEDF, have distinct phenotype and biochemistry. Atypical type VI OI (aVI) is caused by an *IFITM5* p.S40L substitution but is similar to type VI in terms of having progressive bone dysplasia, a fish-scale appearance to bone lamellae under polarized light and decreased cellular PEDF secretion. We generated a conditional KI *Ifitm5*/BRIL p.S42L murine model. Female heterozygous (HET) bone has increased BMDD, disordered collagen fibers on SHG, and increased osteocyte lacunar density but decreased osteocyte canalicular networks. However, bone tissue and osteoblasts of mutant mice from HET S42L x HET S42L matings do not replicate aVI OI lamellar organization and PEDF under-secretion.
Methods: Offspring of HET S42L x WT matings were examined as murine models of human aVI OI. Our approach encompasses comparison of male S42L and *Serpinf1* KO mice by RNAseq, X-Ray, and histology, and western blotting analysis of differentiating osteoblasts from newborn pups. Bone mechanics and histomorphometry of *Serpinf1* WT and KO mice were evaluated.
Results: HET S42L mice have decreased static histology parameters Tb. Th, OV/BV, OS/BS, and Ct.W. The dynamic histology parameters analyzed after calcein injections were generally decreased including MAR, BFR/BS, and BFR/BV. Whole-body, spine, and femur BMD were significantly decreased in S42L by DXA. Differentiating calvarial osteoblasts of HET S42L mice from a WT x HET mating exhibited significantly decreased PEDF secretion on days 0 and 10, while BRIL production was significantly increased at day 0. HBM osteoblast collagen secretion was normal except a1(1) secretion was significantly increased while pC/pN a1(1) and proa1(1) were decreased on day 0.
Conclusions: Our findings support an uncharacterized link between PEDF and BRIL revealed by aVI type OI. The HET S42L mice from a HET x WT mating had decreased PEDF secretion, similar to this parameter in aVI OI patients. Collagen produced by cells with BRIL or PEDF defects has normal primary structure and post-translational modification, suggesting these defects cause OI by affecting matrix formation and late osteoblast differentiation; future studies will characterize this mechanism.

Session Title: Mendelian Phenotypes Poster Session II

PB4663 Bringing genetic diagnosis to patients with hereditary sensory neuropathy through a comprehensive analysis of molecular and proteomic data.

Authors:

P. Lassuthova[1], R. Mazanec[2], D. Brožková[1], P. Seeman[3], A. Majcher[4], A. Museer[4], I. Kurth[5], T. Hornemann[4]; [1]Charles Univ., Dept. of Paediatric Neurology, 2nd Faculty of Med., Prague, Czech Republic, [2]Charles Univ., Dept. of Neurology, Faculty of Med., Prague, Czech Republic, [3]Charles Univ., Dept. of Paediatric Neurology, 2nd Faculty of Med., Praha 5, Czech Republic, [4]Univ. Hosp. Zurich, Dept. of Clinical Chemistry, Zurich, Switzerland, [5]Uniklinik RWTH Aachen, Aachen, Germany

Introduction: Hereditary sensory and autonomic neuropathies (HSAN) and congenital insensitivity to pain (CIP) are rare genetic disorders characterised by a lack of pain sensation. **Aim of the study:** To determine the genetic causes of HSAN/CIP. **Patients and methods:** At our centre, we have assembled a cohort of 82 Czech patients with HSAN. Of these, *SPTLC1* variants were found in two families, *SPTLC2* in two families and *RAB7* in three families. Afterwards, most unresolved patients had their exomes sequenced (ES). **Results:** Using ES, additional causal variants were found (always 1 family) in these genes: *NTRK1, SCN11A, SCN10A, MME, DST, and SPTLC2*. A likely pathogenic variant in *SCN11A* in one family was also detected. In addition, a biallelic pathogenic *RCF1* expansion was found in 6 patients. After ES and these analyses, a total of 50 patients remained unresolved. We approached the analysis of ES data based on proteomic data in an attempt to further identify the cause in this group - as the biosynthesis of sphingolipids is crucial in HSAN, genes involved in this pathway were analysed. In one patient we found a variant in the *PHGDH* gene, which is the rate-limiting enzyme in the serine synthesis. The patient is a sporadic case in the family. The first symptoms appeared at the age of 22 after an injury, before which he was an active boxer. Afterwards, he developed complications that led to the amputation of the second digit of the right foot. Currently, at the age of 37, he has undergone amputation of both lower limbs below the knee. A variant in the *PHGDH* gene (NM_006623.4): c.682G>T, p.(Gly228Trp) in the homozygous state was detected in the patient. Both parents are heterozygous carriers. The patient's plasma showed a reduced serine concentration and an increased alanine to serine ratio, which was associated with an increased deoxysphingolipid formation. Biallelic loss-of-function mutations in *PHGDH* have been described to cause Neu-Laxova syndrome and are also associated with macular telangiectasia. The phenotype of our patient is different as he presented with axonal sensory neuropathy. The p.(Gly228Trp) variant is also expected to be loss-of-function, however the residual activity of the enzyme has to be determined. **Conclusions:** Our study presents molecular genetic data from a relatively large cohort of patients with HSAN. The combination of genomic and proteomic data analysis may be helpful in identifying potential targets. All authors declare no conflict of interest. The research follows the criteria of Declaration of Helsinki. Supported by: LX22NPO5107.

Session Title: Mendelian Phenotypes Poster Session III

PB4664 Burden of GM2 gangliosidosis in Kuwait is driven by founder mutations.

Authors:

N. Almoosawy[1,2], F. Mohammed[3], A. Elshafey[3], M. Pavlovic[4], A. Alaqeel[5], I. H. Almashmoom[5], A. Al-Ahmad[6], B. Albash[3], L. Bastaki[3], D. Marafi[7]; [1]Kuwait Univ., Faculty of Med., Kuwait, Hawalli, Jabriya, Kuwait, [2]Dept. of Pediatrics, Mubarak Alkabeer Hosp., Ministry of Hlth., Kuwait, Hawalli, Jabriya, Kuwait, Kuwait, [3]Kuwait Med. Genetics Ctr., Ministry of Hlth., Sulaibikhat, Kuwait, Kuwait, [4]Dept. of Pediatrics, Pediatric Neurology Unit, Alsabah Hosp., Alsabah, Kuwait, Kuwait, [5]Ahmadi Hosp., Kuwait Oil Company, Ahmadi, Kuwait, Kuwait, [6]Kuwait Med. Genetic Ctr., Sulaibikhat, Kuwait, [7]Kuwait Univ., Jabriya, Kuwait

Introduction: GM2 gangliosidoses are a group of rare lysosomal storage disorders resulting from deficiency of the beta-hexosaminidase enzyme including Tay-sachs disease (TD), causing hexosaminidase A (Hex A) deficiency due to biallelic HEXA pathogenic variants; Sandhoff disease (SD), causing hexosaminidase A and B (Hex A and B) deficiency due to biallelic HEXB pathogenic variants; and the AB variant, caused by biallelic GM2A pathogenic variants. In Kuwait, data about GM2 gangliosidoses are lacking. We provide a summary of phenotypic-genotypic features of such disorders in Kuwait showing that their burden is driven by founder mutations. **Methods:** A retrospective review of all cases at Kuwait Medical Genetic Center with a molecular diagnosis of TD, SD or AB variant either by next generation sequencing or targeted testing.**Results:** We identified 12 subjects with genetically-confirmed GM2 gangliosidoses, including seven TD cases, four SD cases, and one AB variant case. All seven subjects with TD were born to five families from the same ancestral background and were homozygous for a single start codon variant c.2T>C;p.(Met1Thr), confirmed pathogenic by enzyme testing, with all having the juvenile form and typical TD features. The four subjects with SD were born to three unrelated families from the same tribe all of which had the same intronic variant c.1169+3_1169+10del, previously reported pathogenic in three SD cases from Saudi Arabia, either in homozygous state, or in compound heterozygous state with another intronic variant c.771+5G>C in trans. The c.1169+3_1169+10del variant was confirmed a tribal founder variant by genealogical analysis and confirmed pathogenic by enzymatic testing. All subjects had the infantile form with typical SD features and abnormal brain MRI imaging. All subjects but one had retinal cherry red spot, macrocephaly and epilepsy. A single subject had absence of cherry red spot at 16 months and microcephaly with no reported seizures. One subject with AB variant had a homozygous 691 kb deletion at 5q33.1 encompassing GM2A and presenting with global developmental delay, immunodeficiency, and failure to thrive. **Conclusion:** Our data expand the phenotypic spectrum of GM2 gangliosidoses and suggest that the burden in Kuwait is driven by tribal and ancestral founder mutations, guiding clinical practice and prevention policies. Clinicians should have high index of suspicion for these disorders in at-risk children, based on ancestral origin, presenting with regression or ataxia. Inclusion of aforementioned founder mutations in the premarital carrier screening may lead to significant reduction in burden of such disorders in Kuwait.

Session Title: Mendelian Phenotypes Poster Session I

PB4665 C2S2: Phenotype-driven clustering for discovery of disease subgroups.

Authors:

D. Danis[1], A. J. M. Dingemans[2], M. Pekar[3,4], P. N. Robinson[1]; [1]The Jackson Lab. for Genomic Med., Farmington, CT, [2]Radboud Univ. Med. Ctr., Nijmegen, Netherlands, [3]Dept. of Physiology, Med. Faculty, Masaryk Univ., Brno, Czech Republic, [4]Vascular and Miniinvasive Surgery Ctr., AGEL Trinec-Podlesi, Trinec, Czech Republic

Deep phenotyping is an increasingly valuable source of information for understanding the genetic and environmental factors that contribute to human health and disease. One key question for precision genetic medicine involves the balance between lumping and splitting. Over 1000 genes are associated with two or more disease entities in the Online Mendelian Inheritance in Man (OMIM) knowledge base. The "splitting" of diseases associated with any given gene into multiple entities has been largely driven by expert opinion. Here, our goal was to use a data-driven approach to assess the optimal number of subgroups (clusters) in cohorts of genetic syndromes. We developed Consensus Clustering with Semantic Similarity (C2S2), a novel algorithm and a software library for unsupervised phenotype-driven clustering of individuals to discover groups with similar phenotypic characteristics and to support expert analysis. Starting from a cohort of individuals encoded as GA4GH Phenopackets and annotated with terms from Human Phenotype Ontology, C2S2 clusters the cohort using a novel semantic distance measure that takes into account both present and excluded phenotypic features. Next, the algorithm determines if the data are best described by one or multiple clusters and in the latter case leverages consensus clustering to determine the best number of clusters. To enhance the interpretability of the results, C2S2 identifies phenotypic features characteristic for individual clusterings and generates figures to visualize the inter-cluster differences. We demonstrate the performance on 10 cohorts of individuals harboring disease-causing variants in genes associated with two or more Mendelian diseases. The interpretable clustering results provided by C2S2 algorithm will aid in the stratification of patients into disease subgroups for tailored clinical management and to deliver on the promise of precision genomic medicine.

Session Title: Mendelian Phenotypes Poster Session II

PB4666 *CACNA1B* gene mutation associated with typical Rett Syndrome patients from India

Authors:

G. Mohan[1], V. Balachandar[2], S. Nachimuthu[3]; [1]Karpagam Academy of Higher Ed., Coimbatore, India, [2]Central Univ. of Punjab, Bathinda, Punjab, India, [3]Mizoram Univ., Aizawl, India

Background/Objectives: *CACNA1B* encodes the alpha-1-subunit of a voltage-dependent N-type calcium channel Cav2.2 expressed in brain. The heterozygous variants of *CACNA1B* have been previously reported in association with ASD and other neurodevelopmental disorders. Here, we report the rare familial case series of three Rett Syndrome siblings with intellectual disability (ID).**Methods:** After signing informed consent from the patients/parents this study was conducted in accordance with the principles of the World Medical Declaration of Helsinki. The phenotypic characteristics were recorded and for the genotypic analysis whole blood of 3 ml was collected and the DNA was extracted and purified using Qiagen kit. Genomic DNA with high integrity, as checked by nanodrop-based spectrophotometric and Qubit based fluorometric method, were used as the input material for exome sequencing library preparation using TruSeq DNA Exome Library Prep Kit (illumina). HaplotypeCaller of Genome Analysis Toolkit (GATK v4.1.4.1) was used for the germline variant calling.**Results:** In three profound patients showed the phenotypic characteristics of microcephaly, short stature, loss of communication, gait abnormalities, apraxia, hypoalgesia, and repetitive movements. We had found the pathogenic heterozygous variant of rs4422842 (*CACNA1B*) associated with Rett Syndrome in three ID patients. There was a correlation of microcephaly and short stature with the *CACNA1B* variants found in 3 siblings.**Conclusion:** The phenotype of rare neurological disorders often overlaps. For proper genetic counselling and treatment, both phenotype and genotype information should be used to make the diagnosis. Whole exome sequencing studies showed that the genetic varients in *CACNA1B*, increase the risk of intellectual disability. Screening of calcium voltage dependent genes are inevitable for the ID patients.

Session Title: Mendelian Phenotypes Poster Session I

PB4668 *CAMK2*, from mice to men and back: insights in the CAMK2 expertise center where clinicians and scientists work side-by-side.

Authors:

G. van Woerden, D. C. M. Veenma; Erasmus Med. Ctr., Rotterdam, Netherlands

With the advancing technological developments in the field of DNA diagnostics for rare congenital neurocognitive developmental disorders (NDD), the number of newly identified disorders has greatly increased in the past 10 years. For a large portion of these rare NDDs, very little is known about the underlying mechanism, which severely hampers the development of a possible treatment, or optimal care. The ENCORE Expertise Center in Rotterdam, The Netherlands, is an international expertise center for various rare NDDs, where a multidiciplinary team of clinicians and fundamental scientists, closely collaborate to gain more insights into these NDDs and strive to achieve optimal care and identify optimal treatment.
One of the relatively novel rare NDDs for which ENCORE is the only expertise center worldwide, is the CAMK2 syndrome. CAMK2 is a highly abundant protein kinase in the brain, known to play a critical role in synaptic plasticity, and learning and memory. Despite extensive fundamental research on CAMK2 from ourself and others, not much is known about disease etiologie, prognosis or potential treatments. For this reason we have initiated a natural history study in the clinic, and functional genomic studies in the lab to understand the pathogenicity of CAMK2 variants identified in patients. Additionally, in the lab we focus on potential therapeutic options, which we then can bring to the clinic to initiate Phase I clinical trials. Based on this joint effort of the multidisciplinary team of the expertise center, we will be able to identify genotype - phenotype correlations and aim to have in place optimal care and potential treatment options in the near future. During the presentation, I will use the CAMK2 expertise center as an example, sharing preliminary data, for what we believe is the best way forward from diagnosis of rare NDDs towards treatment options for these patients.

Session Title: Mendelian Phenotypes Poster Session II

PB4669 Candidate genes identified in a large UK family associated with Familial Thoracic Aortic Aneurysm and Dissection (FTAAD) from a Whole Exome Sequencing (WES) study.

Authors:

J. Aragon-Martin[1,2,3], V. Khurana[1], S. La Penna[1], S. Babu[1], D. Morris-rosendahl[4], C. Nienaber[4], N. Naqvi[4], F. Dijk[4], N. Ghali[4], A. Reed[4], J. A. R. Gaer[4], J. Pepper[4], U. Rosendahl[4], J. Sneddon[5], E. Fisher[6], J. Dean[7], A. Saggar[8], M. Jahangiri[8], E. Behr[9], M. Simpson[10], A. Smith[10], M. Y. B. Salmasi[4,2], A. Child[11,12], M. Moosajee[3,13], T. Athanasiou[2,14], P. Deloukas[15]; [1]Queen Mary Univ. of London, London, United Kingdom, [2]Imperial Coll., London, United Kingdom, [3]Univ. Coll. London, London, United Kingdom, [4]Royal Brompton Hosp., Guy's and St. Thomas' NHS Fndn. Trust, London, UK, London, United Kingdom, [5]East Surrey Hosp., Redhill, United Kingdom, [6]The Park Surgery, Horsham, United Kingdom, [7]NHS Grampian, Aberdeen, United Kingdom, [8]St. George's Univ. Hosp. NHS Fndn. Trust, London, United Kingdom, [9]St. George's, Univ. Hosp., NHS Fndn. Trust, London, United Kingdom, [10]King's Coll. London, London, United Kingdom, [11]Imperial Coll., London, London, United Kingdom, [12]Marfan Trust, London, United Kingdom, [13]Moorfiels Eye Hosp. NHS Fndn. Trust, London, United Kingdom, [14]Imperial Coll. Hlth.care NHS Trust, London, United Kingdom, [15]Queen Mary Univ. London, London, United Kingdom

BACKGROUND

Thoracic Aortic Aneurysm & Dissection (TAAD) can arise at any time without previous warning and with fatal consequences. About 20% of non-syndromic TAAD is inherited with variable penetrance and severity. It has been crucial to make collaborations with TAAD networks around the world to facilitate novel gene discovery like LMOD1 gene. Our database has collected (over >30 years) 91 probands (53 with a family history and 38 sporadic cases). The proband of one of these TAAD families had an emergency aortic root replacement for Type A dissection (age 52); 2/3 of his sons (at age 25 [4cm] and 28 [4.3cm], both of them on losartan medication), and his cousin (at age 37 [4.2cm]) have an aortic aneurysm. Full features suggesting Marfan, Loeys-Dietz and Ehlers-Danlos were not met upon examination and were excluded by subsequent testing of *ACTA2*, *FBN1*, *PRKG1*, *SMAD3*, and *TGFBR1/2* genes. Negative results from an "aortopathy gene panel" were recorded in 2014. Proband has a deletion picked up by microarray-CGH: 12p12.23 (DNA size deletion [455 kb], which did not segregate with disease in the family. There is no family history of joint dislocations, slipped disc, knee problems or ocular disease.

METHODS

From the family pedigree, the 2 most distantly related affected individuals were processed for WES. Sequence alignment to the GRCh38, variant calling and annotation were performed. The complete variant list was filtered using our in-house pipeline that included the removal of all intronic and intergenic polymorphisms (except splice-site disrupting variants), heterozygous variants with MAF of >0.001 (gnomAD & ExAC), CAD score <20, those that did not segregate with the disease, and all homozygous variants. Then, the list was further filtered based on the frequency of the most common causative variant in *ACTA2* (R258C, MAF=0.00032).

RESULTS

The final list contains 5 candidate genes (3 non-synonymous [*GRIFIN*, *RMND1*, *SYNE1*], 1 frameshift insertion [*LAPTM4B*], and 1 stop-gain [*TSPAN10*] variants). Two promising candidate genes (*SYNE1* and *LAPTM4B*) were found to have a high expression in the ascending aorta tissue (Bgee), and in-silico analysis classified them as being disease-causing variants. *SYNE1* stands out in Phenolyzer with a 0.017 score, while *LAPTM4B* & *TSPAN10* have a 0 score. Exomiser prioritised *SYNE1*, *MANF* (filtered out after exclusion of 5'UTR variants), and *MUC12*, *MUC16* & *MAN2C1* (filtered out after CADD exclusion step).

CONCLUSIONS

This study is still in progress, but the variant in candidate gene *SYNE1* has maximum priority. Family segregation analyses are being performed for these 5 candidate genes to confirm their association with the TAAD phenotype in this family.

Session Title: Mendelian Phenotypes Poster Session III

PB4670 Candidate genes in a hEDS large UK Caucasian family from whole genome sequencing (WGS) data.

Authors:

S. La Penna[1], J. A. Aragon-Martin[1,2,3], V. Khurana[1], S. Babu[1], I. Nasrullah[4], D. Morris-rosendahl[5], C. Nienaber[5], N. Naqvi[5], F. Dijk[5], N. Ghali[5], A. Reed[5], J. A. Gaer[5], J. Pepper[5], U. Rosendahl[5], J. Sneddon[6], E. Fisher[7], J. Dean[8], A. Saggar[9], M. Jahangiri[10], E. Behr[10], M. Y. B. Salmasi[2], N. Vadgama[11], C. Brookes[12], J. Nasir[12], A. Child[2,13], M. Moosajee[3,14], T. Athanasiou[2,15], P. Deloukas[16]; [1]Queen Mary Univ. of London, London, United Kingdom, [2]Imperial Coll., London, United Kingdom, [3]Univ. Coll. London, London, United Kingdom, [4]Imperial Coll. of London, London, United Kingdom, [5]Royal Brompton Hosp., Guy's and St. Thomas' NHS Fndn. Trust, London, United Kingdom, [6]East Surrey Hosp., Surrey, United Kingdom, [7]The Park Surgery, Surrey, United Kingdom, [8]NHS Grampian, Aberdeen, United Kingdom, [9]St. George's Univ. Hosp. NHS Fndn. Trust, London, United Kingdom, [10]St. George's Univ., London, United Kingdom, [11]Stanford Sch. of Med., Palo Alto, CA, [12]Univ. of Northampton, Northampton, United Kingdom, [13]Marfan Trust, London, United Kingdom, [14]Moorfiels Eye Hosp. NHS Fndn. Trust, London, United Kingdom, [15]Imperial Coll. Hlth.care NHS Trust, London, United Kingdom, [16]Queen Mary Univ. London, London, United Kingdom

BACKGROUND

Hypermobile Ehlers-Danlos syndrome (hEDS) seems to be an autosomal dominant type of connective tissue disorder within the family of EDS. The molecular basis of this disease is still unmapped. Because of this the diagnosis of the disease is based solely on family history discussion and its clinical manifestations. These features include joint hypermobility, proneness to muscle weakness and general musculoskeletal involvement, abnormal skin hyperextensibility and tissue fragility. Some EDS phenotypic features have been linked with altered collagen producing genes like the *COL5A1* gene; however, hEDS is most likely not correlated to collagen and appears to be non-fatal. From a large family with hEDS (recruited - Rheumatology Clinic), a female member, age 54, was presented to the Chest Pain Clinic with chest pain diagnosed as costochondritis, which resolved spontaneously. On examination she is mildly myopic, but with no lens or retinal problems, with normal palate and teeth. She has no scoliosis, pectus deformity or arachnodactyly. Her Beighton Score is 4/9 and she has 90˚little fingers and thumb touching forearm bilaterally. Her echocardiogram was normal, with no aortic root aneurysm. Her sister was known to experience hypermobile joint pain, as is the proband's maternal aunt; a second maternal aunt, her daughter and her granddaughters were also affected. A specific marker of this family is a hypermobile 90˚interphalangeal thumb joint as also demonstrated in 1 male and 2 female family members in the last generation of the pedigree.

METHODOLOGY

From a cohort of hEDS probands, 1 family was entered in this study and 2 related family members were screened by whole genome sequencing; known EDS genes were ruled out. The variants were filtered according to an in-house protocol, which includes removal of all intronic and intergenic polymorphism (except splice site related), heterozygous with MAF>0.001 (gnomAD, ExAC), CADD<20, non-segregating with the disease, and all homozygous variants. Candidate genes were reviewed using databases like OMIM, GeneCards, GTex and Bgee.

RESULTS

Out of 15 variants, 9 are non-synonymous, 2 stop-gain, 2 frameshift insertion, 1 non-frameshift insertion and 1 splicing. Candidate genes: high expression scores in cartilage tissue (Bgee) (*PABPC3*, *SRA1*, *IL32*, *ITGAE*). Phenolyzer top genes: *DGKA* and *HLA-DRB1*. Genomiser prioritized *HLA-DRB1*, *GAB2* (filtered out after MAF) and *FOXD4L4* (filtered out after CADD) as top genes.

CONCLUSION

Priority lies on 6/14 genes based on expression levels in relevant tissues (Bgee). Family segregation studies are being performed to confirm the association of the candidate genes with hEDS phenotype.

Session Title: Mendelian Phenotypes Poster Session I

PB4671 *CAPRIN1*-linked neurodevelopmental disorder: understanding the role of CAPRIN1 loss on neuronal differentiation, neurogenesis, and proliferation.

Authors:

L. Pavinato[1,2,3], A. Delle Vedove[4,5], E. Campus[6], E. R. Vecchi[6], A. Verrillo[2,3], A. Angius[7], V. Rallo[7], E. Ferrero[1], G. Ferrero[8], L. Conti[6], B. Wirth[4], A. Baggiolini[2,3], A. Brusco[1,9]; [1]Dept. of Med. Sci., Univ. of Turin, Turin, Italy, [2]Inst. of Oncology Res. (IOR), BIOS+, Bellinzona, Switzerland, [3]Università della Svizzera Italiana, Lugano, Switzerland, [4]Inst. of Human Genetics, Ctr. for Molecular Med. Cologne, Ctr. for Rare Diseases Cologne, Univ. Hosp. Cologne, Univ. of Cologne, Köln, Germany, [5]Inst. for Genetics, Univ. of Cologne, Köln, Germany, [6]Lab. Of Stem Cell Biology, Dept. Cellular, Computational and Integrative Biology – CIBIO, Univ. of Trento, Trento, Italy, [7]Inst. of Genetic and BioMed. Res., Natl. Res. Council, Monserrato, Cagliari, Italy, [8]Dept. of Clinical and Biological Sci., Univ. of Turin, Orbassano, Turin, Italy, [9]Med. Genetics Unit, Città della Salute e della Scienza Univ. Hosp., Turin, Italy

Cell cycle-associated protein 1 (*CAPRIN1*; OMIM* 601178) is a ubiquitous protein involved in cell proliferation and migration. In neurons, it likely regulates the transport and translation of mRNAs involved in synaptic plasticity. We have recently associated *CAPRIN1* loss-of-function variants with an autosomal dominant neurodevelopmental disorder characterized by language impairment/speech delay (n:12; 100%), intellectual disability (83%), ADHD (82%), and ASD (67%). In this study, we have generated CRISPR/Cas9-engineered *CAPRIN1*$^{+/-}$ hiPSCs-lines and differentiated them into neuronal cells, observing several defects including decreased processes length, overall disruption of the neuronal structure, enhanced neuronal death, impaired calcium signalling, increase protein translation and oxidative stress (PMID:35979925).To further define the mechanisms behind those neuronal defects, we performed transcriptomics analysis in early and late differentiated hiPSCs-derived neurons. The obtained data explained the physiological defects we previously observed, highlighting impaired oxidative phosphorylation, neuronal outgrowth, and translation. We also found a significant increase in genes involved in cell cycle control, DNA replication, proliferation, and dopaminergic neurogenesis and a decrease in GABA pathway-related genes. Preliminary data on *CAPRIN1*$^{+/-}$ neuronal progenitor cells, showed an increased proliferative rate and expression of PAX6, NESTIN and SOX2, further suggesting a possible loss of balance in neuronal proliferation/differentiation. To explore these features, we are generating forebrain cortical organoids (fBOs) from *CAPRIN1* patient-derived hiPSCs lines, to analyse their cellular architecture, neuronal activity and the cellular composition combining immunofluorescence analysis and a multicolor immunophenotyping panel (NeurOMIP assay, PMID:33852833). First results show an important impact of *CAPRIN1* loss on the development of neuronal rosettes and fBOs structure. In addition, we previously reported defective translation in *CAPRIN1*$^{+/-}$ neurons; these defects are detectable also in fBOs models at day 30-50 of differentiation, and preliminary data show an effect on some specific neuronal populations, including MAP2B and GFAP positive cells. Further experiments are ongoing to assess the impact of *CAPRIN1* loss on certain neuronal populations, and to establish the specific role of the defective protein translation. In conclusion, our study will unravel the impact of *CAPRIN1* loss in neurodevelopment, shedding light on its role in neuronal differentiation, neurogenesis and proliferation.

Session Title: Mendelian Phenotypes Poster Session III

PB4673 Cardiomyopathy, an uncommon phenotype of congenital disorders of glycosylation: Characterization of the spectrum and recommendations for follow-up.

Authors:

R. Zemet Lazar[1], A. Edmondson[2], R. Shah[3], M. Patino[2], K. D. Hope[1], A. M. Yesso[1], K. Sarafoglou[4], A. Larson[5], C. Lam[6], E. Morava[3], F. Scaglia[7]; [1]Baylor Coll. of Med., Houston, TX, [2]Children's Hosp. of Philadelphia, Philadelphia, PA, [3]Mayo Clinic, Rochester, MN, [4]Univ. of Minnesota Med. Sch., Minneapolis, MN, [5]Univ. of Colorado, Denver, CO, [6]Univ. of Washington, Seattle, WA, [7]Baylor Col Med, Houston, TX

Background Congenital disorders of glycosylation (CDGs) are a continuously expanding group of monogenic disorders of glycoprotein and glycolipid biosynthesis that cause multi-system diseases. Cardiac involvement in CDGs can vary significantly, not only based on the specific disorder, but also between patients with the same disorder. In this study, we aimed to identify common cardiomyopathy patterns among patients with a broad spectrum of CDGs, and to propose follow up recommendations for implementation in clinical practice. **Methods** Patients were enrolled in the Frontiers in Congenital Disorders of Glycosylation Consortium (FCDGC) natural history study. We analyzed clinical and molecular history, imaging characteristics of cardiac involvement, type of cardiomyopathy, and clinical outcomes of individuals with CDG also found to have cardiomyopathy. **Results** Out of 282 patients with CDG enrolled in the FCDGC study, we identified 16 individuals, 9 females and 7 males, diagnosed with CDG confirmed by molecular testing, who also exhibit cardiomyopathy. The majority of these patients are diagnosed with PMM2-CDG (n=10), however, cardiomyopathy was seen in other diagnoses including PGM1-CDG (n=3), ALG3-CDG (n=1), DPM1-CDG (n=1), and DPAGT1-CDG (n=1). Most patients (ten with PMM2-CDG) presented with hypertrophic cardiomyopathy, three with dilated cardiomyopathy (two with PGM1-CDG and one with ALG3-CDG), two with left ventricular non-compaction cardiomyopathy (one with PGM1-CDG and one with DPAGT1-CDG), and one patient with DPM1-CDG presented with an unspecified cardiomyopathy type. The estimated median age of cardiomyopathy onset was 5 months (range: prenatal-23 years). Seven patients were deceased, including five with PMM2-CDG, one with DPAGT1-CDG, and one with ALG3-CDG, though cardiomyopathy was not necessarily the primary cause of death. Cardiac improvement was observed in three patients with PMM2-CDG and one patient with DPM1-CDG, four patients demonstrated a progressive course of cardiomyopathy, while the condition remained static in eight individuals. One patient with PGM1-CDG underwent cardiac transplantation. **Conclusion** Cardiomyopathy can manifest in patients with CDG. Interestingly, some of these cases do not appear to exhibit progression. We advocate that clinicians caring for patients with CDGs should be aware of the risk of cardiomyopathy, especially in PMM2 and PGM1-CDG, and obtain baseline and periodic cardiac evaluation, especially if clinical concerns arise. Furthermore, patients diagnosed with cardiomyopathy should be under periodic cardiological supervision for optimal management and monitoring of their condition

Session Title: Mendelian Phenotypes Poster Session I

PB4674 Carriers of pathogenic variants in causal cardiomyopathy genes show incomplete disease penetrance in UK Biobank

Authors:

S. Klasfeld, H. Kim, R. Moccia, J. Berghout; Pfizer, Cambridge, MA

Identifiable rare, pathogenic mutations explain a significant proportion of intrinsic cardiomyopathy risk. More than twenty genes have been curated by the ClinGen cardiomyopathy working group as having a strong or definitive causal relationship to cardiomyopathies. However, individual carriers of the same or similar rare variants can exhibit variable penetrance and expressivity of disease. We designed a computational filtering strategy to classify UK biobank (UKBB) observed variants as known or predicted pathogenic by using a combination of frequency, literature/ClinVar assertions, and functional consequence prediction (VEP/LOFTEE, Mutpred2, etc). We applied this pipeline to UKBB (data release 450K participants) which yielded genetically defined case (n=10,896 carriers of risk variants for dilated cardiomyopathy (DCM), 6,248 hypertrophic cardiomyopathy (HCM), 3,245 arrhythmogenic right ventricular cardiomyopathy (ARVC)) and control cohorts. Using electronic health records data to identify cardiomyopathy-relevant phenotypes, we determined that the overall rate of phenotype positivity in carriers of DCM risk variants was 22.02%, HCM risk variants was 21.91% and ARVC risk variants carriers was 21.79%. In ongoing work, we will use this genotype-first strategy together with UKBB genetic and clinical information to explain a proportion of gene and gene-set penetrance variability.

Session Title: Mendelian Phenotypes Poster Session II

PB4675 Case report of a novel *FGFR2* gene mutation resulting in mild clinical manifestations of Crouzon syndrome.

Authors:

N. Cruz-Perez[1], E. Albino[2], N. Arciniegas[3], A. Cornier[4], S. Carlo[1,3,5]; [1]Ponce Hlth.Sci. Univ., Ponce, PR, [2]Univ. of Puerto Rico - Sch. of Hlth.Professions, San Juan, PR, [3]Mayaguez Med. Ctr., Mayaguez, PR, [4]San Jorge Children s & Women s Hosp., San Juan, PR, [5]Ponce Res. Inst., Ponce, PR

Craniosynostosis refers to the early closure of cranial sutures during infancy and can be classified as syndromic or non-syndromic based on its clinical presentation. Syndromic craniosynostosis consists of autosomal dominant genetic syndromes and has been linked to specific mutations in the fibroblast growth factor receptor (*FGFR*) genes. Crouzon syndrome (CS) is one of the most common forms of syndromic craniosynostosis, presenting with a flattened forehead, proptosis, hypertelorism, a beaked nose, and midface hypoplasia. It is known that CS is caused by mutations in the *FGFR2* gene, and many variants have been found to be characteristic of this syndrome. Here, we present a 5-year-old Puerto Rican male with craniosynostosis, bilateral exophthalmos, inattention, and episodes of non-responsiveness who is suspected of having attention deficit/hyperactivity disorder. A Blueprint Genetics Whole Exome Plus test revealed a heterozygous missense variant in the *FGFR2* gene (Leu357Ser). This mutation is predicted to be likely pathogenic and possibly caused by a de novo variant. Moreover, head computer tomography revealed the closure of the sagittal, coronal, and lambdoid sutures, which is congruent with craniosynostosis. This case report describes a patient with a novel *FGFR2* mutation resulting in mild phenotypical characteristics quite similar to those of CS, the mildness of which prevents an accurate diagnosis. Genetic testing is vital to correctly diagnose syndromic craniosynostosis, particularly when its variable expressivity results in minimal clinical manifestations.

Session Title: Mendelian Phenotypes Poster Session III

PB4676 Case Report: CAMK2B variant with tetralogy of Fallot, developmental delay, and growth retardation.

Authors:

Y. Kuroda, Y. Saito, Y. Enomoto, T. Naruto, K. Kurosawa; Kanagawa Children's Med. Ctr., Yokohama, Japan

CAMK2B encodes the beta-subunit of calcium/calmodulin-dependent protein kinase II (CAMK2). CAMK2B is associated with intellectual disability and developmental delay, and more than 19 patients have been reported. Here, we described a novel de novo CAMK2B variant in a patient with tetralogy of Fallot (TOF), developmental delay, and growth retardation. The patient was a 2-year-old female and the second child of healthy nonconsanguineous parents. She was delivered at 36 weeks 6 days gestational age by caesarean section due to non-reassuring fetal status, with birth weight of 1,680 g (-2.0 SD), birth length of 43.5 cm (-1.5 SD), and occipital-frontal head circumference (OFC) of 29.4 cm (-1.7 SD). She was diagnosed with Tetralogy of Fallot (TOF) by her fetal echocardiogram. Blalock-Taussig shunt was performed at 9 months of age. Head MRI revealed no other anomaly than slightly white matter volume loss and hypoplastic corpus callosum at 13 months of age. She was referred to our department at the age of 18 months for assessment of developmental delay, microcephaly, and growth retardation. At referral, her height was 68.8 cm (-3.9 SD), weight was 5.7 kg (-4.2 SD), and OFC was 40 cm (-4.3 SD). She controlled her head position at four months, rolled at six months, sat at 13 months, crawled at 18 months, and walked with support at 21 months. She began speaking words at 2 years old. Her karyotype and chromosomal microarray were normal. Her dysmorphic features included puffy eyelids, puffy cheeks, and cup ears.In the patient's exome sequencing, the mean depth of coverage was 103.13 per base, and 96.3 % of the coding region was covered by at least 20 reads. A de novo missense CAMK2B variant (NM_172079.2:c.895A>G (p.Lys299Glu) NC_000007.14:g.44241708T>C (hg38)) was identified by proband exome sequencing. The de novo variant was confirmed by Sanger sequencing. According to ACMG 2015 variant interpretation guidelines, this variant was classified as "Likely pathogenic (PS2+PM2_supporting+PP2+PP3+PP4)". No other candidate variants have been detected.CAMK2A and 2B-related neurodevelopmental disorders (NDDs) are characterized by various degrees of developmental delay and intellectual disability, autistic behavior, epilepsy, and dysmorphic features. More than 36 patients have been reported as CAMK2A and 2B-related NDDs. Congenital heart defect was an atypical phenotype in these NDDs, and only one patient had ventricular septal defect with CAMK2A variant. TOF has never described in previously reported patients. As a de novo CAMK2B (p.Leu443Val) variant have been found in a cohort of TOF, CAMK2B may be associated with TOF.

Session Title: Mendelian Phenotypes Poster Session I

PB4677 Cell type specific network analysis in Diversity Outbred mice identifies genes potentially responsible for human bone mineral density GWAS associations.

Authors:

L. Dillard, G. Calabrese, L. Mesner, B. Al-Barghouthi, A. Abood, C. Farber; Univ. of Virginia, Charlottesville, VA

BACKGROUND: Genome-wide association studies (GWAS) have identified over 1,100 statistically independent associations for bone mineral density (BMD), the most significant clinical predictor of osteoporotic fracture. We previously performed network analyses based on bulk transcriptomic data from bone tissues and heterogenous primary bone cells to predict candidate causal genes for BMD GWAS associations. Here, we applied the same network approaches to transcriptomic profiles of specific osteogenic cell types with the goal of identifying additional candidate causal GWAS genes. This approach also provides information regarding the cellular context for putative causal genes which is essential for future functional studies.
METHODS: Using the Diversity Outbred (DO) mouse population (N=80, 32 females, 48 males), we performed scRNA-seq on bone marrow-derived stromal cells cultured under osteogenic differentiation. After stringent processing and quality control, the dataset consisted of 139,392 total cells from 15 cell clusters ranging from cell types of the osteogenic lineage to various immune cells (e.g., macrophages and lymphocytes). To prioritize cell types, we used CELLECT with the largest BMD GWAS performed to date to identify cell types enriched for BMD heritability. For each prioritized cell type, we generated co-expression networks. Next, Bayesian networks were generated for each co-expression network whose module eigengene was correlated with at least one of 55 bone strength-related traits collected in the same DO mice. Key driver analysis was performed to identify critical network genes. Mouse key drivers with a corresponding human homolog were prioritized based on proximity to BMD GWAS associations, colocalizing expression quantitative trait loci (eQTL) or splicing QTL (sQTL) from the Genotype-Tissue Expression (GTEx) Project, and evidence of impacting BMD based on mouse knockout data from the International Mouse Knockout Consortium (IMPC).
RESULTS: We identified 148 cell type-specific modules that were correlated with at least one bone strength trait in the DO mice. Key driver analysis identified 834 network “key driver” genes on average per cell type. Of the key driver genes across all cell types, 394 had a colocalizing eQTL and/or sQTL. Of these, 61 demonstrated a statistically significant effect on BMD in IMPC knockout mice.
SUMMARY: Using our network-based approach and prioritization scheme leveraging multiple data modalities, we discovered many genes, including novel targets such as, *Prpf40a*, *Tmem97*, *Atf7*, *Creld2*, *Zzef1*, and *Eif5b*, that are putatively implicated in variation in BMD, bone strength, and risk of osteoporosis.

Session Title: Mendelian Phenotypes Poster Session II

PB4678 Cerebral vasculopathy in Rubinstein-Taybi syndrome

Authors:

K. Kurosawa, Y. Saito, Y. Kuroda; Kanagawa Children's Med. Ctr., Yokohama, Japan

Rubinstein-Taybi syndrome (RSTS) is a rare malformation syndrome characterized by broad thumbs and toes, intellectual disability, developmental delay, and distinctive facial appearance. The causative genes comprise CREBBP and EP300 through loss of function for the disorder. Mice with homozygous mutation of CREBBP show defects of vasculo-angiogenesis and hematopoiesis, leading to severe hemorrhage in telencephalon and mesencephalon, resulting in embryonic lethality. However, the defects of vasculo-angiogenesis, shown by the previous studies in mice, have been reported in a few cases. We report Moyamoya disease in the patient with RSTS associated with t(2;16), reported by Imaizumi and Kuroki [1991]. At age 34, brain magnetic resonance angiography (MRA) revealed moyamoya vessels in the right hemisphere. The neurological investigation was within normal ranges, including deep tendon reflex, diadochokinetic signs, and muscle tonus. Sequence analysis of RNF213 identified the most common susceptibility variant p.(R4810K) in East Asian. After that, we continued conservative follow-up for her. At age 37, she underwent neurosurgery for subarachnoid hemorrhage with no apparent origin of bleeding. Prior information about moyamoya disease enabled prompt neurosurgery, and she recovered without paralysis. At 48, she still has cerebrovascular abnormalities as moyamoya disease but has not had any stroke or other ischemic symptoms. The previous reports of a cerebrovascular anomaly in RSTS are so limited. The cerebrovascular anomaly or strokes were not noted in the study of adult individuals with RSTS. Park et al. [2021] suggested that having the pathogenic variants both in the EP300 gene and the PLAU gene, which is involved in angiogenesis, predisposed to abnormal angiogenesis in their patient. Combined with studies of CREBBP/EP300-deficient mice and limited case reports, RSTS individuals may develop cerebral vascular hemorrhage and abnormal angiogenesis when another genetic background related to the vasculopathy is added. Therefore, cerebrovascular disease should also be included in the medical management of RSTS in adulthood.

Session Title: Mendelian Phenotypes Poster Session III

PB4679 Characteristics of *JAG1* and *NOTCH2* variants in patients with Alagille syndrome in Japan.

Authors:

T. Tanimura[1], T. Togawa[1], S. Ito[1], K. Ohashi[1], K. Imagawa[2], H. Hayashi[3], K. Ito[1], T. Endo[1], T. Sugiura[1], S. Saitoh[1]; [1]Nagoya City Univ., Nagoya, Japan, [2]Univ. of Tsukuba, Tsukuba, Japan, [3]The Univ. of Tokyo, Tokyo, Japan

Alagille syndrome (ALGS; OMIM 118450) is an autosomal dominant multisystem disorder that affects the liver, heart, face, eyes, skeleton, and other organs. Its prevalence is estimated to be 1 in 30,000 live births. ALGS is caused by pathogenic variants in one of two genes in the Notch signaling pathway, *JAG1* or, rarely, *NOTCH2*. Pathogenic variants in *JAG1* are most commonly protein-truncating, such as frameshift, nonsense, exon level deletions, and splice site. *NOTCH2* variants are predominantly missense. Pathogenic variants in *JAG1* have been identified in approximately 90% of clinically diagnosed ALGS patients. To date, *JAG1* and *NOTCH2* variants, over 800 and 100, respectively, have been reported in Western countries and East Asia, while *JAG1* and *NOTCH2* variants in patients with ALGS have not been reported in a large series in Japan, except in our previous report (n = 28, Ohashi et al., Acta Paediatr. 2017). We aimed to elucidate the diversity of *JAG1* and *NOTCH2* variants in Japanese patients with ALGS. We carried out molecular genetic analyses with targeted next-generation sequencing of *JAG1* and *NOTCH2* and/or multiplex ligation probe amplification covering the entire exons of *JAG1*. We detected 81 *JAG1* pathogenic variants in unrelated patients who were diagnosed or highly suspected as ALGS, and 73 of those variants were unique. Regarding the *JAG1* variant types, 22 frameshift variants (32%), 17 nonsense variants (28%), 11 splice site variants (14%), 8 gross deletions (10%), and 10 missense variants (16%) were detected. We observed fewer frameshift variants and more nonsense variants than those in the largest previous study (Gilbert MA et al., Hum Mutat. 2019). Forty-nine (60%) of all the variants were presented in 8 exons of *JAG1*; exon 2, 4, 6, 10, 12, 17, 23, and 24. In contrast, there was no variant in 8 exons; exon 1, 7, 8, 9, 14, 19, 20, and 25. The distribution of variants in *JAG1* was more similar to that reported in Western countries (Warthen DM et al., Hum Mutat, 2006) than that from China (Li L et al., PLoS One. 2015). Three pathogenic variants in *NOTCH2* were detected: one frameshift, one nonsense, and one missense variant. These findings could provide the characteristics of *JAG1* and *NOTCH2* variants in the patients with ALGS in Japan.

Session Title: Mendelian Phenotypes Poster Session I

PB4680 † Characterization of a chromosome 5q35 Alzheimer disease risk locus in African Ancestry families using short and long read whole genome sequencing

Authors:

D. Van Booven[1], S. Fazal[1], A. Griswold[1], K. Nuytemans[1], F. Rajabli[1], M. Jean-Francois[1], L. Adams[1], T. D. Starks[2], P. Whitehead[1], B. Kunkle[1], M. Cuccaro[1], J. Vance[1], J. Haines[3], C. Reitz[4], G. S. Byrd[2], G. Beecham[1], M. Pericak-Vance[1]; [1]Univ. of Miami, Miami, FL, [2]Wake Forest Univ., Winston-Salem, NC, [3]Case Western Reserve Univ., Cleveland, OH, [4]Columbia Univ., New York, NY

Background: Genetic risk for Alzheimer's disease (AD) can vary between genetic ancestral populations, with the best example being the significantly lower risk conferred by *APOEe4* in African relative to European or East Asian ancestries. Thus, identification of other ancestry specific risk or protective factors is critical to understand the etiology of AD in diverse populations.
Method: Multipoint linkage was performed across 51 African American (AA) multiplex AD families, consisting of 160 AD and 318 cognitively unimpaired (CU) individuals, as part of the Research in African American Alzheimer Disease Initiative (REAAADI) and Late-Onset AD Family Study (LOAD) initiatives. Fine mapping of linked loci was performed using short-read whole genome sequencing (WGS) followed by variant annotation for putative function including protein changes for coding variants and evidence for regulatory activity (ENCODE, RoadMap Epigenome) and chromatin interactions (publicly available HiC and promoter capture C) for noncoding variants followed by co-segregation analysis. To identify potential structural variants (SV) underlying signals we also performed long-read WGS (LRS) using the Oxford Nanopore PromethION followed by alignment with minimap2 and variant calling with Sniffles2.
Result: Linkage analyses identified a peak LOD score (HLOD=3.20) on chr5q35. Co-segregation analysis of the short-read WGS identified a 3'UTR and synonymous variant in *INSYN2B* as well as a promoter variant in *WWC1* (all MAF<0.01) segregating with disease in the AD individuals of the family with the highest LOD score contributing to the linkage peak. Interestingly, four other AA families contributing to the chr5 linkage signal harbor different within-family shared variants located in *INSYN2B*'s promoter or in enhancer regions with evidence for interaction with *INSYN2B*'s promoter. Co-segregation analysis of the LRS SVs in the highest LOD score family revealed further potential candidate genes with all AD individuals sharing eight SVs (six deletions, 2 insertions). These include a deletion of an exon in *KCNMB1*. Combining ongoing functional validation in this region will help us better understand the contribution of these genes to AD development.
Discussion: Our analyses provide evidence for candidate genes *INSYN2B*, encoding an inhibitory synaptic factor active in oligodendrocytes and neurons, and *KCNMB1,* encoding a potassium calcium-activated channel, to play a role in AD genetics in the AA population. This finding highlights the importance of diversifying population-level genetic data to better understand the genetic contributions to AD on a global scale.

Session Title: Mendelian Phenotypes Poster Session II

PB4681 Characterization of lower limb differences in individuals with Cornelia de Lange syndrome.

Authors:

K. Arnold[1], S. Raible[2], L. Conway[1], I. Krantz[2,3]; [1]Master of Sci. in Genetic Counseling Program, Perelman Sch. of Med., Univ. of Pennsylvania, Philadelphia, PA, [2]Roberts Individualized Med. Genetics Ctr., Div. of Human Genetics, Children's Hosp. of Philadelphia, Philadelphia, PA, [3]Dept. of Pediatrics, Perelman Sch. of Med., Univ. of Pennsylvania, Philadelphia, PA

Cornelia de Lange syndrome (CdLS) is a genetic, multisystem diagnosis that has been associated with morphologic differences of the upper limbs such as small hands, proximally placed thumbs, fifth finger clinodactyly, radioulnar synostosis and variable limb reductions. Lower limb involvement commonly includes small feet and 2-3 toe syndactyly with structural/reduction difference being much less frequently reported. Recent observations suggest that lower limb involvement in individuals with CdLS can be more severe, and is not fully represented in the literature. This study aims to: 1) review the existing literature on lower limb involvement in CdLS, 2) characterize lower limb anomalies in CdLS individuals seen at the Children's Hospital of Philadelphia (CHOP), and 3) identify additional CdLS individuals with lower extremity structural differences via an IRB-exempt, RedCap administered survey sent to the National CdLS Foundation. A PubMed search to identify previously reported individuals with a diagnosis of CdLS and lower limb differences yielded 82 articles. Of these, 10 articles contained sufficient information on 13 cases identified to have CdLS and unique lower limb differences. Anomalies were often associated with the fibulae and tibulae, variants in the *NIPBL* gene (when tested for), and significant upper limb involvement. For the CHOP cohort, information for 2,006 probands was obtained through an existing cohort of individuals previously enrolled under an IRB approved protocol of informed consent. Of the 2,006 affected individuals, 137 (6.8%) had photos of the lower extremities available for review and at least one lower limb difference. Minor lower limb differences such as small feet (76%), 2-3 toe syndactyly (40%), brachydactyly (38%), and lateral deviation of the first toe (18%) were commonly seen in CdLS while 6 (0.3%) individuals had more involved lower limb differences such as oligodactyly with clubfoot or 1-2 syndactyly. For the CdLS survey, 32 survey responses were received; however, only 7 individuals were identified as having less common lower limb features including limb length discrepancies (6/7) often occurring below the knee and one individual with bilateral in-toeing, bent tibiae and fibulae, and a left clubfoot. A majority (6/7) of individuals had an intervention for their lower limb involvement including occupational (3/7) or physical therapy (4/7), surgery (2/7) or shoe modification (3/7). These results provide a more complete phenotype of CdLS limb anomalies, guide management options, and improve clinical diagnostics and targeted genetic testing for patients and families prenatally and postnatally.

Session Title: Mendelian Phenotypes Poster Session III

PB4682 Characterization of the impact of NSD2 loss on the transcriptional and epigenetic landscape in the brain

Authors:

S. Kinoshita[1,2], K. Kojima[1], E. Ohnishi[1], Y. Takayama[1], H. Kikuchi[1], S. Takada[1], K. Nakabayashi[1], T. Kawai[1], K. Hata[1,3]; [1]Natl. Res. Inst. for Child Hlth.and Dev., Tokyo, Japan, [2]Tokyo Med. and Dental Univ. Graduate Sch. of Med. and Dental Sci., Tokyo, Japan, [3]Gunma Univ. Graduate Sch. of Med., Gunma, Japan

Epigenetic disruptions have been implicated in neurodevelopmental disorders. Recent studies have shown that *NSD2* is associated with developmental delay/intellectual disability; however, its role in brain development and function remains unclear. To better understand the role of NSD2 in the brain, we performed transcriptomic and epigenetic analyses using *Nsd2* knockout mice. We found that the loss of NSD2 caused dysregulation of genes related to synaptic transmission and formation in the brain. Genome-wide analysis showed that NSD2-mediated H3 lysine 36 dimethylation (H3K36me2) marked quiescent regions in the embryonic brain. Finally, changes in H3K36me2 were partially associated with changes in gene expression in *Nsd2* knockout brains. These results suggest that NSD2 is involved in the regulation of genes that are important for neural function via H3K36me2. Our findings provide insights into the role of NSD2 in the brain and provide a better understanding of the pathogenesis of NSD2-associated neurodevelopmental disorders.

Session Title: Mendelian Phenotypes Poster Session I

PB4683 Characterizing the phenotypic abnormalities of a mouse model of Snyder Robinson-syndrome: a valuable tool for therapeutic development

Authors:

O. Akinyele[1], A. Munir[2], M. Johnson[1], M. Perez[2], Y. Gao[1], J. Foley[3], Y. Wu[1], T. Murray-Stewart[3], B. Casero[3], H. Bayir[1], D. Kemaladewi[4]; [1]Univ. of Pittsburgh, Pittsburgh, PA, [2]Univ. of Pittsburgh, PITTSBURGH, PA, [3]Johns Hopkins Sch. of Med., Baltimore, MD, [4]Univ. of Pittsburgh Sch. of Med., Pittsburgh, PA

The polyamines (putrescine, spermidine, and spermine) are essential for normal cellular functions and are subject to strict metabolic regulation. Mutations in the gene encoding spermine synthase (SMS) lead to the accumulation of spermidine in an X-linked recessive disorder known as Snyder-Robinson syndrome (SRS), which manifests with a spectrum of symptoms including intellectual disability, developmental delay, thin habitus, and low muscle tone with no available treatment. Development of therapeutic interventions for SRS would require a suitable disease-specific animal model that recapitulates many of the abnormalities observed in affected individuals. Here, we study the molecular, behavioral, and neuroanatomical features of a mouse model with a missense mutation in the *Sms* gene that results in a glycine-to-serine substitution at position 56 (G56S) of the SMS protein. The G56S mice exhibit a complete loss of SMS protein and elevated spermidine/spermine ratio in the skeletal muscles and the brain, as well as increased anxiety, impaired learning, and decreased explorative behavior. Furthermore, these mice fail to gain weight over time and exhibit abnormalities in brain structure and bone density. Transcriptomic analysis of the cerebral cortex revealed the downregulation of genes associated with ribosomal protein synthesis and mitochondrial oxidative phosphorylation, which were functionally recapitulated in fibroblasts. Collectively, our findings establish the first in-depth characterization of an SRS preclinical mouse model that identifies cellular processes that could be targeted for future therapeutic development. Current experiments focus on developing an effective gene therapy approach for SRS via viral-mediated expression of WT *Sms* transgene. We anticipate that the successful completion of this study will open the door to a gene therapy approach for SRS and other polyamine-related genetic diseases

Session Title: Mendelian Phenotypes Poster Session II

PB4684 Chimeric *TNXB/TNXA* gene mutations in hypermobile Ehlers Danlos Syndrome.

Authors:

M. Galindo[1], R. E. Handsaker[2,3], J. G. Andrews[1], C. M. Laukaitis[4]; [1]Univ. of Arizona, Tucson, AZ, [2]Broad Inst., Cambridge, MA, [3]Harvard Med. Sch., Dept. of Genetics, Boston, MA, [4]Carle Illinois Coll. of Med., Urbana, IL

Introduction: Biallelic pathogenic mutations in *TNXB* are known to cause classic-like Ehlers Danlos Syndrome (EDS), but less described heterozygous mutations in *TNXB* can cause hypermobile EDS (hEDS). *TNXA* is a pseudogene of *TNXB* that is non-functional due to two mutations: a 120 base pair deletion that removes a splice junction for exon 35 and an amino acid substitution in exon 40. The paralogy of *TNXB* and *TNXA* precludes use of next generation sequencing to investigate the region, as short-reads do not unambiguously map to either the gene or pseudogene. The similarities also predispose to misalignment of DNA, which can produce a chimeric dysfunctional *TNXB/A* chimera that deletes material from *TNXB* (copy number loss) and incorporates *TNXA* mutations. There are two *TNXB/A* chimeric configurations that have been described in individuals with hEDS, but these mutations have not been explored in a cohort of individuals with hEDS to determine frequency in a population.
Methods: The hEDS Gene Study enrolled and obtained biosamples from 100 people diagnosed with hEDS using the 2017 diagnostic criteria. A validated TaqMan assay was used to identify wild type *TNXB* copies in genomic DNA by targeting two sites that distinguish it from *TNXA*: the deletion (exon 35) and single nucleotide variant (exon 40; C4060W). This assay was run in triplicate on DNA from study participants using *HBB* as an endogenous control. The cycle threshold mean was calculated per individual to determine *TNXB* copy number.
Results: Of the 100 participants, 91.0% are female and the average age is 44.5 years (s.d. 14.9). 71/100 individuals had the expected two copies of *TNXB* exon 40 and 35, representing the wild type allele. More than two *TNXB* exon 40 alleles were seen in conjunction with only two *TNXB* exon 35 sequences in 23/100 individuals. Finally, 6/100 had three or more alleles reflecting wild type *TNXB*. There were no people with copy number losses causing chimeric *TNXB/A*.
Conclusion: Chimeric *TNXB/A* genes with a loss of expressed *TNXB* are seen in individuals with hEDS, but were not observed in our study, suggesting this may not be a frequent cause of hEDS. Observed copy number discrepancies between *TNXB* exons 35 and 40 suggest there may be chimeric genes creating a copy number gain, but this assay is not able to determine if the chimeric gene is expressed or could contribute to the hEDS phenotype. Since copy number variation in this region is observed in the general population, further analysis is needed to determine whether the variation seen in this population with hEDS differs from the general population.

Session Title: Mendelian Phenotypes Poster Session III

PB4685 Cleidocranial Dysplasia Family Case Report and identification of *RUNX2* and *SULF1* variants

Authors:

A. Rivera-Cameras[1], A. V. Ruiz-Ramírez[2], M. A. Aceves-Aceves[3], M. C. Morán-Moguel[4], M. Salazar-Paramo[5], I. P. Dávalos-Rodríguez[1]; [1]División de Genética, Centro de Investigación Biomédica de Occidente, IMSS, Guadalajara, México, Doctorado en Genética Humana, CUCS, Univ. de Guadalajara, México., Guadalajara, México, Mexico, [2]Univ. Autónoma de Guadalajara, Departamento de Neurociencias, Programa Internacional de la Facultad de Med., Guadalajara, Mexico, Mexico, [3]Unidad Medica Alta Especialidad, HGO, CMNO, IMSS, Guadalajara, México, Mexico, [4]Departamento de Biología Molecular y Genómica, CUCS, Univ. de Guadalajara, México., Guadalajara, México, Mexico, [5]Departamento de Fisiología, CUCS, Univ. de Guadalajara, Guadalajara, México, Mexico

INTRODUCTION: Cleidocranial dysplasia (CCD), OMIM #119600, is a rare autosomal dominant disorder characterized principally by delayed fusion of cranial sutures, hypoplastic or aplastic clavicles, and dental abnormalities. CCD is caused by pathogenic variants in the RUNX2 (Runt-related transcription factor 2) gene (6p21.1) involved in the differentiation of osteoblasts and bone formation. OBJECTIVE: To describe a family with clinical and radiological data of CDD and molecular diagnosis. RESULTS: Heterozygous pathogenic variant was identified in the RUNX2 gene c.852del (p.Ile285Leufs*23) in three family members, and additionally, in one of them, a heterozygous uncertain variant was identified in the SULF1 gene c.643G>C (p.Val215Leu). We performed a bioinformatic analysis and predicted that the SULF1 variant is in an important functional domain, and it could probably affect the activity of the protein. Since SULF1 is implicated in the Wnt and BMP signaling pathways as RUNX2, it will ultimately increase the affectation in developing tissues such as chondrogenesis and skeletal bone formation. CONCLUSIONS: We propose that the SULF1 gene could act as a modulator gene of the CDD phenotype by intervening in the same signaling pathways, which would affect the development of tissues such as chondrogenesis and skeletal bone formation. We suggest a more detailed analysis of variants in potential modifiers genes that could contribute to clinical variable expression in CDD patients.

Session Title: Mendelian Phenotypes Poster Session I

PB4686 Clinical and genetic characteristics of 78 Chinese subjects with *POU3F4*-related hearing loss.

Authors:

H. Yuan; Inst. of Rare Diseases, West China Hosp. of Sichuan Univ., Chengdu, China

This study aims to enhance our understanding of X-linked hearing loss (DFNX2) associated with the *POU3F4* gene by investigating its clinical and genetic spectrum, as well as the underlying pathophysiological mechanisms. While previous reports have described DFNX2 in numerous male patients, most of the identified *POU3F4* variants have been documented in case reports with limited sample sizes. We identified 75 pathogenic variants, including single nucleotide variants, insertions/deletions, and copy number variants, within the *POU3F4* gene in a cohort of 78 male patients. The clinical manifestations and severity of the hearing loss displayed significant variability among male patients and female carriers. Missense variants were found to cluster in the POU-specific and POU homeodomain regions, demonstrating an association with highly and moderately conserved residues among POU family proteins (paralog-conservation). Stop-loss variants primarily disrupt the normal function of *POU3F4* by impairing its DNA-binding ability and exhibiting additional effects such as mislocalization. This study expands the clinical and genetic spectrum of the *POU3F4* gene.

Session Title: Mendelian Phenotypes Poster Session II

PB4687 Clinical and genetic spectrum in a cohort of Lithuanian patients with limb girdle muscular dystrophy

Authors:

B. Burnyte[1], G. Asadauskaite[2], A. Morkuniene[1], A. Matuleviciene[1], L. Cimbalistiene[1], A. Utkus[1]; [1]Inst. Of BioMed. Sci., Faculty Of Med., Vilnius Univ., Vilnius, Lithuania, [2]Faculty Of Med., Vilnius Univ., Vilnius, Lithuania

Introduction Limb-girdle muscular dystrophies (LGMDs) encompass a group of genetic disorders characterized by predominant proximal muscle weakness and wasting. LGMD type R1, resulting from pathogenic variants in the *CAPN3* gene, is the most common subtype worldwide. This study aims to investigate the clinical and genetic spectrum of LGMD in a cohort of Lithuanian patients. Methods In this retrospective cohort study, we collected data from 45 patients with LGMD who were genetically diagnosed between 2012 and 2021. We analyzed clinical, demographic and molecular genetic data. Results The study cohort included 19 (42%) males and 26 (58%) females with a mean age of 31.27 years. Proximal muscle weakness of the upper and lower limbs was the most common presenting symptom. The most frequently identified causative genes were *CAPN3* (n=32), *DYSF* (n = 4), *FKRP* (n = 4), *ANO5* (n = 2), and *SGCB* (n = 1). Additionally, we identified a c.598_612del variant in the *CAPN3* associated with dominant inheritance in one family. The *CAPN3* c.550del, p.(Thr184fs) variant was the most commonly observed, and two new *CAPN3* variants (c.1342C>T and c.1746-3C>A) were identified. Conclusion Our study suggests that a significant proportion of LGMD cases in the Lithuanian cohort are associated with *CAPN3* variants.

Session Title: Mendelian Phenotypes Poster Session III

PB4688 Frequency of genetic disorders associated with pseudoanodontia, revisiting classification, and suggestive management.

Authors:

M. Mostafa[1], M. Mehrez[2], I. Sayed[3], N. Hassib[4]; [1]Natl. Res. Ctr., Dokki, Egypt, [2]Natl. Res. Ctr., Cairo, Egypt, [3]Natl. Res. Ctr., Dokki, Cairo, Egypt, [4]Natl. research centre, Cairo, Egypt

Absence of teeth in the oral cavity is one of the most common dental abnormalities. Anodontia is a term used to describe congenitally missing teeth.This condition can be either true, due to tooth agenesis or false (pseudo) due to impaction, or failure of eruption. In this single-center seven-year retrospective study, we report and describe the frequency of genetic disorders associated with pseudoanodontia in patients who presented to the outpatient clinic of Oro-dental Genetics Department at the National Research Center, Cairo, Egypt.The recorded patients were classified into; primary pseudoanodontia (failure of teeth eruption) and secondary pseudoanodontia (early loss of teeth). Cleidocranial dysplasia was the most frequent disorder in the primary pseudoanodontia followed by GAPO syndrome, while Papillon-Lefèvre syndrome was the most frequent disorder in the secondary pseudoanodontia followed by congenital insensitivity to pain with anhydrosis and hypophosphatasia. Dental mangement of primary pseudoanodontia is challenging, in this study we reported The dental management of 19-year old male suffered from Cleidocranial dysplasia depended on preservation of the unerupted teeth by orthodontic traction after their surgical exposure and 11-year old male suffered from GAPO syndrome by conventional complete denture.

Session Title: Mendelian Phenotypes Poster Session I

PB4689 Clinical exome sequencing efficacy and phenotypic expansions involving anomalous pulmonary venous return.

Authors:

E. Huth[1], X. Zhao[1], P. N. Luna[1], I. Vogel[2], I. L. H. Dorf[3], S. Joss[4], J. Clayton-Smith[5], M. J. Parker[6], J. J. Louw[7], M. Gewillig[8], J. Breckpot[9], A. Kraus[10], E. Sasaki[11], U. Kini[12], T. Burgess[13], T. Y. Tan[13], R. Armstrong[14], K. Neas[15], G. B. Ferrero[16], A. Brusco[16], W. S. Kerstjens-Frederikse[17], J. Rankin[18], L. R. Helvaty[19], B. J. Landis[19], G. C. Geddes[19], K. L. McBride[20], S. M. Ware[21], C. A. Shaw[1], S. R. Lalani[1], J. A. Rosenfeld[1], D. A. Scott[1]; [1]Baylor Coll. of Med., Houston, TX, [2]Aarhus Univ., Aarhus, Denmark, [3]Aarhus Univ. Hosp., Aarhus, Denmark, [4]Queen Elizabeth Univ. Hosp., Glasgow, United Kingdom, [5]Manchester Univ. Hosp., Manchester, United Kingdom, [6]Sheffield Children's Hosp., Sheffield, United Kingdom, [7]Pediatric Cardiology Maastricht Univ. Med. Ctr., Maastricht, Netherlands, [8]Dept. of Cardiovascular Sci. KU Leuven, Leuven, Belgium, [9]Catholic Univ. Leuven, Leuven, Belgium, [10]Chapel Allerton Hosp., Leeds, United Kingdom, [11]Oxford Ctr. for Genomic Med., Oxford, United Kingdom, [12]Oxford Univ. Hosp. NHS FT, Oxford, United Kingdom, [13]Murdoch Children's Res. Inst., Melbourne, Australia, [14]Addenbrooke's Hosp., Cambridge, United Kingdom, [15]Genetic Hlth.Service NZ, Wellington, New Zealand, [16]Univ. of Torino, Torino, Italy, [17]Univ. Med. Ctr. Groningen, Groningen, Netherlands, [18]Peninsula Clinical Genetics Service, Exeter, United Kingdom, [19]Indiana Univ. Sch. of Med., Indianapolis, IN, [20]Univ. of Calgary, Calgary, AB, Canada, [21]Indiana Univ., Indianapolis, IN

Anomalous pulmonary venous return (APVR) frequently occurs with other congenital heart defects (CHDs) or extra-cardiac anomalies. While some genetic causes have been identified, the optimal approach to genetic testing in individuals with APVR remains uncertain, and the etiology of most cases of APVR is unclear. Here, we analyzed molecular data from 49 individuals who received clinical exome sequencing from Baylor Genetics to determine the diagnostic yield of clinical exome sequencing (ES) for non-isolated APVR. A definitive or probable diagnosis was made for 8 of those individuals yielding a diagnostic efficacy rate of 16.3%. We then analyzed molecular data from 62 individuals with APVR accrued from three databases (Baylor Genetics clinical exome, Cytogenetics of Cardiovascular Malformations Consortium Registry, and DECIPHER) to identify novel APVR genes. Based on data from this analysis, published case reports, mouse models, and/or similarity to known APVR genes as revealed by a machine learning algorithm, we identified 3 genes—*EFTUD2*, *NAA15*, and *NKX2-1*—for which there is sufficient evidence to support phenotypic expansion to include APVR. We also provide evidence that 3 recurrent copy number variants contribute to the development of APVR: proximal 1q21.1 microdeletions involving *RBM8A* and *PDZK1*, recurrent BP1-BP2 15q11.2 deletions, and central 22q11.2 deletions involving *CRKL*. Our results suggest that ES and chromosomal microarray analysis (or genome sequencing) should be considered for individuals with non-isolated APVR for whom a genetic etiology has not been identified, and that genetic testing to identify an independent genetic etiology of APVR is not warranted in individuals with *EFTUD2*-, *NAA15*-, and *NKX2-1*-related disorders.

Session Title: Mendelian Phenotypes Poster Session II

PB4690 Clinical features of 110 patients with Angelman syndrome.

Authors:

M. Fujimoto[1], Y. Nakamura[1], T. Iwaki[1], E. Sato[2], D. Ieda[1], Y. Negishi[1], A. Hattori[1], S. Saitoh[1]; [1]Nagoya City Univ., Nagoya, Japan, [2]Nagoya city Univ., Nagoya, Japan

Background: Angelman syndrome (AS) is a severe neurodevelopmental disorder caused by loss of function of maternal *UBE3A*. The clinical features of AS include severe neurodevelopmental delay, easily provoked laughter, and seizures. Williams (Am J Med Genet 2006) reported that all AS patients exhibited developmental delay, movement or balance disorder, behavioral uniqueness and speech impairment. Additionally, over 80% of AS patients displayed seizures, abnormal electroencephalogram (EEG) and microcephaly. Sleep disorders and hypopigmentation were observed in 20-80% of AS. However, occasionally we experienced AS patients who did not match these features and might miss the opportunities to examine appropriate genetic tests. Here, we investigated clinical features of substantially large number of AS patients. **Methods**: We investigated the clinical features and genetic causes of AS patients through surveys from physicians treating AS patients. **Results**: We investigated 110 AS patients, of whom 28 were caused by a maternal deletion in 15q11-q13, 35 by uniparental disomy (UPD) or imprinting disorder (ID) and 47 by a *UBE3A* mutation. Developmental delay was present in all AS patients. But movement or balance disorders were observed in 91% of patients. Behavioral uniqueness was observed in 97% of patients, and language impairment was present in 80% of patients. Notably, some AS patients caused by UPD or ID tended to exhibit loss of these features. All AS patients with a deletion showed seizures and abnormal EEG. However, approximately 30% of AS patients with UPD/ID or a *UBE3A* mutation did not experience seizures. Microcephaly was observed in less than 50% of AS patients. Sleep disorders were present in over 80% of all AS patients, and more than 50% did not exhibit hypopigmentation. **Conclusion**: Our study demonstrated clinical features of 110 AS patients. Although developmental delay was present in 100%, movement or balance disorder and behavioral uniqueness were not present in all patients. Patients with UPD/ID or a *UBE3A* mutation tended to show less specific phenotypes as reported in previous literature. Therefore, genetic test should not be limited to patients with typical clinical features of AS.

Session Title: Mendelian Phenotypes Poster Session III

PB4691 Clinical features of eight cases of neuronal intranuclear inclusion disease (NIID) in Kochi prefecture, Japan

Authors:

H. Furuya[1,2], D. Yoshimoto[1], S. Ohturu[1], T. Shogase[1], Y. Nishikawa[1], Y. Hashimoto[1], Y. Morita[1], Y. Osaki[1], H. Nakazawa[2], T. Matsushita[1]; [1]Kochi Med. Sch., Kochi Univ., Kochi, Japan, [2]Dept. of Neurology, Nankoku Hosp., Kochi, Japan

Background : Neuronal intranuclear inclusion disease (NIID) (MIM 603472), caused by an expansion of GGC repeats in the 5'-untranslated region of *NOTCH2NLC*, is a neurodegenerative disorder characterized by eosinophilic intranuclear inclusions in neuronal cells and other tissues. Such inclusions are also found in non-neuronal cells. The clinical features and pathological findings in patients with NIID are highly varied, except for progressive dementia.**Objective**: In this study, we present eight cases of NIID in Kochi Prefecture, Japan. Three of the cases deceased after an average clinical course of 11.6 years.**Results**: All cases underwent skin biopsy, MRI imaging, and genetic analysis. Eosinophilic inclusion bodies and increased signals in the corticomedullary junction curvilinear lesions on diffusion-weighted imaging (DWI) were observed in all cases. Expanded GGC repeats were found in all patients. The median age at symptom onset was 61.8 years (range, 57-72 years), and the initial clinical symptoms included dementia (50%), syncope or loss of consciousness (25%), dizziness (25%), parkinsonism (gait disturbance, resting tremor, rigidity) (50%), depression (25%), and character change (25%). All patients exhibited dementia, and six patients (75%) showed parkinsonism during the clinical course. Five patients underwent MIBG cardiac scintigraphy and DAT scans, which showed no abnormalities. Levodopa treatment was administered to all patients with parkinsonism, and some patients showed positive drug effects. In neurological examinations, six NIID patients (75%) showed laterality or severely increased Pulmomental Reflex.**Conclusion:** NIID is considered a clinically heterogeneous disease entity with various clinical features. Dementia is a common symptom, but parkinsonism also commonly occurs during the course of the disease.**Achknowledgement:** We also thank to Prof. Sone J (Department of Neurology, Aichi Medical University, Japan) and Prof. Toda T (Department of Neurology, Tokyo University, Japan) for immune staining of biopsy specimen and genetic analysis.

Session Title: Mendelian Phenotypes Poster Session I

PB4692 Clinical Features of Individuals with Rauch-Steindl Syndrome due to *NSD2* pathogenic variant.

Authors:

E. Nishi[1], K. Yanagi[2], N. Okamoto[1], T. Kaname[3]; [1]Osaka Women s and Children s Hosp., Osaka, Japan, [2]Natl. Ctr. for Child Hlth.and Dev. (NCCHD), Tokyo, Japan, [3]Natl Ctr Child Hlth Dev (NCCHD), Tokyo, Japan

Rauch-Steindl syndrome (RAUST) (MIM#619695) is a rare genetic syndrome characterized by prenatal and postnatal growth retardation, distinctive facial features, and psychomotor delay caused by heterozygous pathogenic variants in *NSD2*. RAUST is similar to Wolff-Hirschhorn syndrome (WHS), caused by a deletion of the chromosome 4p16.3 region where *NSD2* is located. *NSD2*, also known as *WHSC1*, has been recognized as important for the WHS phenotype. While both RAUST and WHS share growth failure and psychomotor delay, there are differences between the two syndromes. RAUST presents with a milder phenotype compared to WHS, and characteristic symptoms such as seizures, cleft palate, and heart disease commonly observed in WHS are less frequent in RAUST. Moreover, facial features of individuals with RAUST have been reported to have sufficient specificity to distinguish it as a distinct syndrome (Zanoni et al., 2021). Here, we present the clinical details of two individuals with *de novo* heterozygous pathogenic variants in *NSD2*. Individual 1: A 7-year-old boy, born at 41 weeks and 3 days via normal delivery. He did not experience respiratory or feeding difficulties in the neonatal period but had feeding problems as an infant. He exhibited moderate psychomotor delay (began walking independently at 28 months, first speech at 4 years; DQ 35). Short stature and delayed bone age were also present. We identified a novel nonsense variant in *NSD2*. Individual 2: A 5-year-old boy, delivered by cesarean section at 36 weeks and 1 day due to umbilical cord torsion. He required ventilator support for 1 day due to respiratory failure. He also showed moderate psychomotor developmental delay (began walking independently at 24 months, first speech at 2 years; DQ 37). Short stature and delayed bone age were evident. We identified a reccurent frameshift variant in *NSD2*. Both individuals exhibited craniofacial features consistent with previously reported RAUST cases, including relative macrocephaly, triangular face, broad arches, and sparse horizontal eyebrows. Neither individual had a history of epilepsy, although individual 1 had abnormal EEG findings. Chiari malformation type I was detected in individual 2 through brain MRI. Cleft lip and palate or congenital heart disease were not present in either individuals. The ages of gait and speech acquisition also suggest that the psychomotor delay seen in RAUST is milder than that in WHS.These findings support previous reports. The accumulation of more RAUST individuals is expected to contribute to the establishment of medical management for RAUST.

Session Title: Mendelian Phenotypes Poster Session II

PB4693 Clinical findings of the developmental disorder by *KDM5C* variants

Authors:

H. Futagawa[1], K. Fukuda[1], H. Yamanaka[1], M. Kuroda[1], S. Ito[1], M. Honda[1], M. Yamada[2], H. Suzuki[2], T. Takenouchi[3], K. Kosaki[4]; [1]Tokyo Metropolitan Children's Med. Ctr., Tokyo, Japan, [2]Ctr. for Med. Genetics, Keio Univ., Shinjuku-ku, Tokyo, Japan, [3]Keio Univ, Tokyo, Japan, [4]Keio Univ. Sch. of Med., Tokyo, Japan

<Introduction> Loss-of-function variants in *KDM5C* (Xp11.22) is known to cause Intellectual developmental disorder, X-linked syndromic, Claes-Jensen type (MIM#300534: MRXSCJ) in which one of the most common causative gene of X-linked intellectual disability. Here, we report additionally two Japanese patients of MRXSCJ.<Patient 1> He was a 5-year-old male born to non-consanguineous healthy Japanese parents. Other than a dilated renal pelvis noted in the second trimester, the perinatal history was normal. The age of two, delayed speech emerged. At five years of age, he had moderate developmental delay. His anthropometry values were within normal range. Dysmorphic features included, short and down slanted palpebral fissure, long philtrum and micrognathia. He suffered recurrent epilepsy and had overfriendly. Trio exome analysis revealed de novo hemizygous missense *KDM5C* variant which was categorized as “Likely Pathogenic” according to ACMG/AMP guideline: NM_004187.5:c.1354G>A p.(Gly452Ser).<Patient 2> He was a 5-years-old male born to non-consanguineous healthy Japanese parents. He was born after an uneventful pregnancy. The age of two, delayed developmental milestone was noted. At five years of age, he had moderate developmental delay. His height was 92.2cm (-3.5SD) which indicated short stature. Dysmorphic features included down slanted palpebral fissure, prominent nasal tip, micrognathia, short neck. Trio exome analysis revealed de novo hemizygous nonsense *KDM5C* variant which was categorized as “Pathogenic” according to ACMG/AMP guideline: NM_004187.5:c.1153C>T p.(Arg385Ter).<Discussion>The patients reported here had facial gestalt similar to Rubinstein-Taybi syndrome (RTS), especially Patient 1 had a happy demeanor compatible to RTS. However, common features suggestive of RTS, a downward nasal septum and broad thumb, were not present. In male patients with atypical RTS, careful physical evaluation may lead to a clinical diagnosis of MRXSCJ. MRXSCJ may require genetic counseling regarding X-linked inheritance and is important as one of the differential diagnosis for RTS, which is generally autosomal dominant inheritance.*KDM5C* is the epigenetic regulator which removes methylation of H3K4me2/3 through the JmjC domain. The pathogenic variant of the Patient 1 was located at JmjC domain and presumed to affect the KDM5C demethylase activity. This variant was novel variant and would expand genotypic knowledge of the *KDM5C*.<Conclusion>In male patients with intellectual disability, distinctive facial dysmorphism and happy demeanor, the X-linked disorder MRXSCJ would be considered.

Session Title: Mendelian Phenotypes Poster Session III

PB4694 Closing the gap: overcoming the challenges of *RPGR* ORF15 sequencing, a target for gene therapy of X-linked retinitis pigmentosa.

Authors:

S. Van De Sompele[1], M. De Bruyne[1], M. Van Heetvelde[1], B. P. Leroy[2], J. De Zaeytijd[3], B. Guillemyn[1], K. Claes[1], T. Rosseel[1], K. De Leeneer[1], M. Bauwens[1], E. De Baere[1]; [1]Dept. of Biomolecular Med. & Ctr. for Med. Genetics Ghent, Ghent Univ. and Ghent Univ. Hosp., Ghent, Belgium, [2]Dept. of Ophthalmology & Ctr. for Med. Genetics Ghent, Ghent Univ. and Ghent Univ. Hosp., Ghent, Belgium, [3]Dept. of Ophthalmology, Ghent Univ. and Ghent Univ. Hosp., Ghent, Belgium

Retinitis pigmentosa (RP) is a group of inherited retinal diseases (IRD) leading to vision impairment in 2 million people worldwide. Representing approximately 10% of all cases, X-linked RP (XLRP) is the most severe form, being generally associated with an earlier disease onset and more rapid progression up to legal blindness by the third-fourth decade. The major gene responsible for XLRP is *RPGR*, encoding a protein involved in the regulation of ciliary transport, harboring up to 80% of pathogenic variants underlying XLRP. Interestingly, the majority of pathogenic *RPGR* variants are located in ORF15, an exon unique to a retina-specific *RPGR* isoform that is expressed in the cilium and basal bodies of the photoreceptor cells.

During recent years, several gene therapy clinical trials have been conducted, aimed at halting the degenerative processes in *RPGR*-IRD. To be eligible for such innovative therapeutic interventions, an accurate and fast molecular diagnosis is required. Due to its highly repetitive and purine-rich nature, ORF15 does however pose substantial technical challenges for genetic testing, including PCR amplification artifacts, insufficient coverage, sequence misalignments, and inaccurate variant calling. Here, we evaluated the performance of different sequencing technologies to overcome these challenges.

We observed that whole exome (WES) and short-read whole genome sequencing (SR-WGS) (Illumina) provide a minimum coverage of <20X for the purine-rich ORF15 subregion (amino acid 700-1030), resulting in inaccurate capture of genetic information within this region. Variants in ORF15 generally need to be confirmed, which requires unique reagents, multiple pairs of overlapping primers, and specific PCR conditions, thus hampering its application in routine diagnostics. Long-read WGS (LR-WGS) using Oxford Nanopore Technologies (ONT) on the other hand demonstrated drastic improvements in the coverage uniformity of the entire ORF15 region, albeit at a ten times higher cost. To overcome these limitations, we designed a KAPA HyperCap probe set (Roche) for the targeted enrichment of ~400 kb clinically relevant regions, including *RPGR*. Sequencing on a NovaSeq 6000 platform (Illumina) resulted in an overall higher coverage, with ORF15 demonstrating a minimum coverage of >120X. Moreover, this strategy allowed us to accurately detect variants in both the purine-rich ORF15 subregion and a broad selection of other IRD genes, for almost half the cost of WES.

Overall, these findings accelerate the molecular diagnosis of *RPGR*-IRD in a robust and cost-effective way, improving patients' eligibility for therapeutic interventions.

Session Title: Mendelian Phenotypes Poster Session I

PB4695 Clusters of rare skeletal dysplasias in Brazil

Authors:

D. Cavalcanti[1], L. Bonadia[1], C. A. Moreno[1], I. Monlléo[2], M. Santos[2], S. Oliveira[3], K. Nunes[4], F. Gameleira[5]; [1]Univ. of Campinas (UNICAMP), Campinas, SP, Brazil, [2]Federal Univ. of Alagoas, Maceió, AL, Brazil, [3]Univ. of Brasilia, Brasília, DF, Brazil, [4]Univ. of São Paulo, São Paulo, SP, Brazil, [5]Prefeitura de Ouro Preto, Ouro Preto, MG, Brazil

In the last 13 years, the Skeletal Dysplasia Group at UNICAMP found evidence of clusters for at least four different osteochondrodysplasias (OCD). Three proven clusters are summarized here. The first was recently found in a small community with less than 300 inhabitants in the South of Minas Gerais, a State of the Brazilian Southeast. In an inbred family in this village, three affected individuals with typical Omodysplasia were found, all of them carriers of a homozygous variant in *GPC6* (c.878-2A>G). The other two clusters were found in Northeastern Brazil, both involving increased bone density. Pycnodysostosis was the 1st one reported by our group (DOI 10.1186/s40001-016-0228-7*)*, and the 2nd cluster is here reported and involves a lethal phenotype - Blomstrand Dysplasia (BD). This OCD is associated with biallelic mutations in the *PTHR1* gene, and fewer than 20 patients have been reported. All available members of three nuclear families were studied using Sanger sequencing, microsatellite analysis, and a high-density SNP array (~600K). The kinship analyses were performed by the method of the moment SNPRelate. The mutation age was inferred using the Gamma method. Finally, genetic ancestry was carried out through the genomic ancestry average (Admixture) and ancestry observed in the genomic mutation region (RFMIX). The same variant (c.639-2 A>G) in *PTHR1* identified in HMZ in two fetuses with BD from two families was found in HTZ in the parents and one unaffected sib of a fetus with BD previously reported (Pediatr Radiol 1999 29:842-845). Microsatellite analysis confirmed the founder effect. Analysis of the SNP array confirmed a second cousin's kinship between the three genealogies and a minimum haplotype shared of ~13 Kb between carriers' individuals. The time since the most recent common ancestor between them was inferred in 5.9 generations (CI=2.3-15.4), meaning the mutation has been segregated in this family for approximately 150 years (CI = 50-375 years). The average genomic ancestry was 72% European, 19% African, and 9% Native American. All carriers have the mutation in a segment of European ancestry. In conclusion, these results confirm our 3rd cluster of OCD due to a founder effect of the c.639-2A>G variant in *PTHR1* that probably emerged in Brazil over 150 years ago or was brought by a European colonizer. Grants : CNPq #590148/2011-7/ Faepex #136/17/ FAPESP # 98/16006-6; 2015/22145-6

Session Title: Mendelian Phenotypes Poster Session II

PB4696 *CNP* variants cause non-syndromic high myopia in humans

Authors:

T. Qi[1], K. Rizwan[1], T. Runyi[1], Z. Liyu[1], M. Sigurbjörn[2], R. Arne[2], F. Chaolan[1], T. Ping[1], K. Petri[2], X. Kun[1], H. Zhengmao[1]; [1]Central South Univ., Changsha, China, [2]Univ. of Bergen, Bergen, Norway

High myopia (HM) is a common form of refractive error with high heritability and one of the leading causes of blindness worldwide. Only a few genes have been implicated in HM, while the genetic basis for most cases remains elusive. This study identified a heterozygous missense variant (c.59A>C, p.K20T) on the *CNP* gene (which encodes the 2',3'-cyclic nucleotide 3'-phosphodiesterase, CNPase) co-segregated with HM in a three-generation family from China. Subsequent screening of *CNP* in the HM cohort led to the discovery of three additional variants; two were recurrent with low frequency (c.59A>G, p.K20R; c.1015G>A, p.V339I), and the remaining one was novel (c.1034G>A, p.G345D). Single-cell RNA sequencing and qPCR demonstrated that *Cnp* was abundantly expressed in the progenitor cells of developing mouse retina, indicating the importance of CNPase in retinal development. Under transmission electron microscopy, the mitochondria size was increased, and the cristae number per mitochondria was decreased in the *CNP* knockout Hela cell lines, implicating the swelling of the mitochondria and disturbed oxidative phosphorylation. After overexpression of CNP wild type and mutants in the knockout cell lines, we detected an increased mitochondrial permeability transition pore opening using the mPTP assay in the mutant groups, indicating that the mitochondria's swelling may attribute to the mPTP opening in knockout cell lines. Then we analyzed the interactomes of CNP wild type and K20T mutant using label-free mass spectrometry. We found that interactive PMPCB (a catalytic subunit of the essential mitochondrial processing protease) was increased, and NFDUFB4 (a subunit of the complex I in the mitochondrial electron transport chain) was decreased in the K20T mutant group. Moreover, in the overexpression experiment, the mitochondrial targeting sequence (MTS) cleavage of K20T was reduced, indicating that K20T's MTS cannot be cleaved normally and play its role subsequently, such as regulating oxidative phosphorylation. We hypothesize that CNP mutations may disturb its pre-peptide processing or its function in the oxidative phosphorylation regulating, thus affecting retina development through impaired mitochondria function and causing high myopia ultimately. To conclude, we found that *CNP* is a novel causal gene in the monogenic form of HM. Its defects may cause HM through impaired mitochondrial function during eye development.

Session Title: Mendelian Phenotypes Poster Session III

PB4697 Comparative RNAseq analysis for the study of motoneuron diseases in multi-omics approaches

Authors:

S. Chen[1], U. Ulmer[2], M. Synofzik[3], R. Schule[4], S. Zuchner[1], M. Danzi[1]; [1]Univ. of Miami, Miami, FL, [2]Univ. of Tübingen, Tübingen, Germany, [3]Hertie Inst. for Clinical Brain Res., Tuebingen, Baden Wã¼rttemberg, Germany, [4]Ctr. for Neurology and Hertie Inst. for Clinical Brain Res., Tübingen, Germany

Nearly half of patients with suspected monogenic Mendelian diseases still remain undiagnosed. Deep intronic non-coding variation contributes to mis-splicing, thereby constituting pathogenic changes that are difficult to discern when analyzing whole genome data alone. The integration of genome sequencing and RNA sequencing can reveal the functional significance of rare changes. However, this approach is challenged when studying neurological and motoneuron diseases. Obtaining affected tissue samples from living patients is typically not feasible. Therefore, it is important to compromise on the availability of type of tissue and its reflection of the target tissue expression patterns. Here we explore the utility of typically available sources of material for RNAseq studies that can empower genome analysis. We found that fibroblasts detectably express 76.8%,81.2%, and 73.6% of genes known to cause the monogenic diseases CMT, HSP, and ataxia, respectively. This outperformed other peripheral tissues such as whole blood and lymphocytes, thereby making fibroblasts a valuable tissue for studying motoneuron diseases. making them a valuable tool for studying HSP. Furthermore, we analyzed RNA-seq data from fibroblasts of two HSP patients carrying a POLR3A splicing variant to evaluate the sensitivity and specificity of several alternative splicing detection tools for diagnostic purposes. Our results highlight the potential of fibroblast RNA-seq data for diagnosing and studying HSP and other motoneuron diseases using peripheral tissue.

Session Title: Mendelian Phenotypes Poster Session I

PB4698 Complex genetic architecture underlies craniofacial microsomia.

Authors:

N. Gogate[1], A. Jolly[1], J. Punetha[1], H. Streff[2], A. Ayaz[3], E. Macnamara[4], M. Racobaldo[5], H. Du[6], H. Du[1], S. Jhangiani[7], D. Muzny[7], R. Gibbs[7], D. Pehlivan[1], S. Antonarakis[8], R. Lewis[9], J. Lupski[1], J. Posey[1]; [1]Baylor Coll. of Med., Houston, TX, [2]Texas Children s Hosp., Houston, TX, [3]Dept. of Med. Genetics, Adana City Training and Res. Hosp., Dept.s of Med. Genetics, Sch. of Med., Istanbul Medipol Univ.,, Adana and Istanbul, Turkey, [4]NIH, Bethesda, MD, [5]Univ. of South Florida, Tampa, FL, [6]BCM, Houston, TX, [7]Baylor Coll. Med., Houston, TX, [8]Univ. of Geneva, Geneva, Switzerland, [9]Baylor Col Med, T-803 Taub Building, Houston, TX

Craniofacial microsomia (CFM, also known as Goldenhar syndrome), is the second most frequent congenital malformation of craniofacial development with an incidence of 1 in 3500 to 5600 births. CFM has variable expressivity with a recognizable clinical set of abnormalities, such as asymmetric development of the ear (with preauricular tags), nose, soft palate, lip, and mandible. CFM has been suggested to be a genetically heterogeneous condition from studies of candidate genes and variant alleles, but only a handful of these alterations seem to be recurrent in unrelated patients with CFM. Most CFM cases are sporadic and largely the molecular basis remains elusive. Through the Baylor Hopkins Center for Mendelian Genomics (BHCMG) and Baylor College of Medicine Genomic Research to Elucidate the Genetics of Rare (BCM-GREGoR) databases, we investigated a cohort of 18 unrelated families with 20 affected individuals diagnosed with CFM using exome sequencing and rare variant, family-based, genomic analyses. Previously, we had identified *FOXI3* as a candidate gene associated with CFM, and affected individuals presented with types II and III microtia or with ear deformity and preauricular tags. Additionally, we identified a homozygous *MACF1* variant (c.4084G>A, (p.E1362K)) in a case with hemifacial microsomia along with cortical dysplasia and focal epileptic findings. Three more individuals with biallelic variants (c.5569C>T, p.(R1857W); c.485C>T, p.(T162I), c.5557T>C, p.(S1853P)) were ascertained through GeneMatcher and provide supplementary evidence for phenotypic expansion of CFM with neurological features. These biallelic *MACF1* variant alleles occur closer to the N-terminus of the encoded protein and present a partially overlapping but unique disease trait from that caused by monoallelic variants within the GAR domain of *MACF1*. Furthermore, through retrospective cohort-wide analysis, we identified putatively damaging variants in candidate genes *SF3B2* and *SIX5*. *SF3B2* haploinsufficiency has been ascertained recently as one cause of CFM, while *SIX5* has been associated with craniofacial microsomia phenotypes in Branchio-Oto-Renal Syndrome 2. These results provide insight into the complex genetic architectures underlying CFM and highlight an evolving picture of genetic heterogeneity contributing to the CFM trait. Furthermore, comprehensive genotype-phenotype studies may better delineate the phenotypic spectrum of CFM or help to identify syndromic presentations driven by unique molecular etiologies.

Session Title: Mendelian Phenotypes Poster Session II

PB4699 Compound Heterozygous Variants in *PRDM5* Causing Brittle Cornea Syndrome Type 2 - A Case Report.

Authors:

C. Lyon[1], J. Black[1], R. Lebel[2]; [1]SUNY Upstate Med. Univ., Syracuse, NY, [2]SUNY Upstate Med Univ, Fayetteville, NY

Background: Brittle cornea syndrome type 2 (OMIM #614170) is an autosomal recessive condition caused by variants in the *PRDM5* gene (OMIM #614161). Typical manifestations are blue sclera, corneal rupture after minor trauma, keratoconus, hyper-elasticity of the skin, and hypermobility of the joints. The *PRDM5* gene plays a role in regulating the development and thickness of the cornea.
Case presentation: A 26-year-old female, born in a non-consanguineous union, had typical neurodevelopment and was in good general health when she presented with progressive sensorineural hearing loss, keratoconus, joint laxity, patellofemoral instability, dystrophic scarring, easy bruising, *pes planus*, blue sclerae, and heterochromia. Whole exome sequencing revealed heterozygous pathogenic variant in *PRDM5* (c. 1513dup) and heterozygous likely pathogenic variant in (c.301-5T>G) in trans. Ophthalmological exam revealed thinning of the central cornea, superficial punctate keratitis that is mild but more in the left than in the right eye, chorioretinal scar right eye, keratoconus, left eye myopia, right eye hypermetropia, bilateral astigmatism, and normal vitreo-retinal imaging, scanning laser ophthalmoscopy, and field studies.
Conclusion: Brittle cornea syndrome type 2 is usually associated with consanguinity, which are not present in this patient. In addition, this patient does not present with a history of corneal perforation, typically a hallmark of this disease. This atypical phenotypic presentation for brittle cornea syndrome type 2 is likely due to the compound heterozygous nature (rather than homozygosity by common descent) of its inheritance. Due to the novel genetic mutations and phenotypic characteristics, this case offers an expansion of the genotype-phenotype correlations in this rare condition.

Session Title: Mendelian Phenotypes Poster Session III

PB4700 Comprehensive Analysis of Low-Frequency Genetic Variants through Exome Sequencing in Familial Multiple Sclerosis

Authors:

E. Turanli[1,2], A. Bulbul[3], E. Everest[4], U. Voyvoda[5], U. Sezerman[3], K. Bilguvar[6,7], C. Boz[8], C. F. Demir[9], F. Yetkin[10], H. Efendi[11], S. D. Bunul[11], T. Duman[12], M. Tecellioglu[13], M. Terzi[14], S. Sen[14], R. Karabudak[15], M. A. Tuncer[16], C. Uzunkopru[17], O. F. Turan[18], S. Demir[19], U. Aluclu[20], H. Gumus[21], M. Tutuncu[22], U. Uygunoglu[22], S. Saip[22], A. Siva[22]; [1]Acibadem Univ., Dept. of Molecular Biology and Genetics, Istanbul, Turkey, [2]Acibadem Univ., Graduate Sch. of Natural and Applied Sci., Dept. of Molecular Biology and Genetics, Istanbul, Turkey, [3]Acibadem Univ., Graduate Sch. of Hlth.Sci., Dept. of Bioistatistics and Bioinformatics, Istanbul, Turkey, [4]Istanbul Technical Univ., Faculty of Sci. and Letters, Dept. of Molecular Biology and Genetics, Istanbul, Turkey, [5]Acibadem Univ., Graduate Sch. of Natural and Applied Sci., Dept. of Molecular and Translational Biomedicine, Istanbul, Turkey, [6]Yale Sch. of Med., Dept.s of Neurosurgery and Genetics, New Haven, CT, [7]Acibadem Univ., Sch. of Med., Dept. of Med. Genetics, Istanbul, Turkey, [8]Karadeniz Technical Univ., Dept. of Neurology, Trabzon, Turkey, [9]Firat Univ., Faculty of Med., Dept. of Neurology, Elazig, Turkey, [10]Erciyes Univ., Faculty of Med., Dept. of Neurology, Kayseri, Turkey, [11]Kocaeli Univ., Faculty of Med., Dept. of Neurology, Kocaeli, Turkey, [12]Mustafa Kemal Univ., Faculty of Med., Dept. of Neurology, Antakya, Turkey, [13]Inonu Univ., Faculty of Med., Dept. of Neurology, Malatya, Turkey, [14]Ondokuz Mayis Univ., Faculty of Med., Dept. of Neurology, Samsun, Turkey, [15]Yeditepe Univ., Faculty of Med., Dept. of Neurology, Istanbul, Turkey, [16]Hacettepe Univ., Faculty of Med., Dept. of Neurology, Ankara, Turkey, [17]Izmir Katip Celebi Univ., Faculty of Med., Dept. of Neurology, Izmir, Turkey, [18]Uludag Univ., Faculty of Med., Dept. of Neurology, Bursa, Turkey, [19]Univ. of Hlth.Sci., Dept. of Neurology, Istanbul, Turkey, [20]Dicle Univ., Dept. of Neurology, Diyarbakir, Turkey, [21]Selcuk Univ., Faculty of Med., Dept. of Neurology, Konya, Turkey, [22]Istanbul Univ.-Cerrahpasa, Cerrahpasa Med. Faculty, Dept. of Neurology, Istanbul, Turkey

Multiple sclerosis (MS) is a chronic, immune-mediated neurological disease with a complex inheritance pattern. In this study, we conducted exome sequencing to explore genetic variants in 45 multiply affected MS families (101 patients and 53 unaffected family members). In 33 families (73 patients and 53 unaffected family members), the pVAAST burden test and Exomiser gene prioritization were performed using 206 MS-related Human Phenotype Ontology terms, and variants with a minor allele frequency (MAF) < 0.01 were selected for both dominant and recessive segregation patterns. This resulted in 527 genes, which then underwent functional enrichment analysis through g:Profiler in multiple databases. These genes were associated with several immunological and cell signaling adhesion pathways, among which the Laminin Subunit Alpha 5 (*LAMA5*) gene, involved in blood-brain barrier (BBB) processes such as cell adhesion, differentiation, migration, and signaling, was highlighted. Variations in *LAMA5* were found as well as in other laminin genes essential for BBB function, such as *LAMA1, LAMC5, LAMA2, LAMB1,* and *LAMA3*. This co-occurrence pattern was identified in patients exhibiting protein-changing variants with a MAF < 0.05 in the gnomAD exome database. Among 40 variants identified in the *LAMA5* gene in the patients (all with a MAF < 0.05 in gnomAD and causing amino acid changes), rs79319629 (p.Lys812Thr; c.2435A>C; ENST00000252999.7; CADD score: 17.18) was found in nine patients across families. This variant was detected in an affected mother-son duo, a nephew and aunt, a sibling pair, and three patients in different families. Other significant rare variants with a global MAF of 0.003 (*LILRA1*) and 0.0006 (*LILRB1*) were identified. *LILRA1* and *LILRB1* are associated with immunoregulatory interactions, including response pathways to natalizumab, which is used in MS treatment. The *LILRA1* variant rs75416770 (p.His164Tyr; c.490C>T; ENST00000251372.8) was detected in one family exhibiting an autosomal dominant inheritance. Selected variants are currently being studied to assess their functionality in co-culture models using genome-edited BBB cells together with peripheral blood mononuclear cells. Acknowledgement: We thank Clifton L. Dalgard and Uniformed Services University, Laboratory Core of the Collaborative Health Initiative Research Program, and Yale Center for Genome Analysis, for sequencing services.

Session Title: Mendelian Phenotypes Poster Session I

PB4701 Comprehensive compendium of Joubert syndrome gene variants and associated phenotypes highlights genetic and phenotypic overlap in ciliopathies.

Authors:

B. Henning[1], H. Loucks[1], A. K. Christman[1], P. Panchal[1], J. Du[1], V. Bowman[1], J. Dempsey[2], D. Doherty[2]; [1]Univ. of Washington, Seattle, WA, [2]Univ of Washington, Seattle, WA

Joubert syndrome (JS) is a largely recessive neurodevelopmental condition that overlaps clinically and genetically with Meckel syndrome (MKS), Nephronophthisis (NPH) and other ciliopathies. >40 genes have been associated with JS, while ~20 have been associated with MKS, and ~40 with NPH. Understanding the genotype-phenotype associations within ciliopathies is essential for interpreting genetic testing. Here, we compile a comprehensive dataset of variants in JS-genes linked to individual phenotypes extracted from the literature. We aim to identify gene-diagnostic category associations and variant-phenotype associations; and create a comprehensive database of JS-genes variants. We searched PubMed for papers reporting known and candidate JS-associated gene variants in patients with ciliopathy or neurodevelopmental phenotypes between 05/2004-07/2021. We extracted variant information, and used the Ensembl Variant Effect Predictor to determine pathogenicity based on predicted functional effect, Genome Aggregation Database (gnomAD) allele frequencies, and Combined Annotation Dependent Depletion (CADD) scores. We also extracted individual phenotypic features, which we mapped to Human Phenotype Ontology (HPO) terms using an in-house algorithm. We computed the percentage of individuals with variants in each phenotypic category and functional category. Data curation is partially done. So far, we have compiled 902 unique variants in 1437 individuals assigned to 12 diagnostic groups: JS (38%), retinal dystrophy (21%), NPH (18%), MKS (12%), and oral-facial-digital syndrome (7%). Other diagnostic groups each represented ~1%. Of the 43 JS-associated genes, 15 were only associated with JS, while 19 were associated with JS and MKS, and 7 with JS-NPH. *CEP290* variants were associated with the largest number of conditions (NPH, SLS, JS, MKS, RD, OCRS, NDC-NOS). About 60% of the variants had allele frequencies <= 0.001 in gnomAD. The top 25% most common variants had allele frequencies between 0.01-0.04. ~98% of the variants had CADD scores >15, 94% >20, 80% >25, 49% >30. 67% of individuals with JS had frameshift variants only, 17% truncating variants only, 4% missense variants only; 1% CNV variants only. In summary, we assembled a comprehensive and highly curated database of JS-gene variants and the associated phenotypes, which will aid in determining the pathogenicity of future variants. Going forward, we will analyze genotype-phenotype correlations, identify all variants that might be amenable to antisense oligonucleotide treatment, and contribute the curated variants to publicly accessible databases.

Session Title: Mendelian Phenotypes Poster Session II

PB4702 Comprehensive Genetic Evaluation of Witteveen-Kolk Syndrome in a 6-Year-Old Puerto Rican Girl: Identification of a Novel *SIN3A* Variant

Authors:

C. Rodriguez[1], E. Albino[2], F. Velez-Bartolomei[1], S. Carlo[3], A. Cornier[4]; [1]San Jorge Children & Women's Hosp., San Juan, PR, [2]Univ. of Puerto Rico - MSC, San Juan, PR, [3]Ponce Hlth.Sci. Univ., Cabo Rojo, PR, [4]San Jorge Children s & Women s Hosp., San Juan, PR

Witteveen-Kolk syndrome is a rare neurodevelopmental disorder characterized by developmental delay and intellectual disability. With less than 50 reported cases worldwide, the syndrome occurs in approximately 1 in 10,000 individuals. We present the case of a 6-year-old Puerto Rican girl referred for genetic evaluation due to speech and developmental delay since the age of four, accompanied by mild hypotonia. A comprehensive genetic assessment was performed on the patient, starting with a metabolic newborn screening panel that revealed a heterozygous pathogenic mutation in the *MMACHC* gene (c.481C>T; p.Arg161), associated with autosomal recessive methylmalonic aciduria with homocystinuria due to cobalamin deficiency. Further evaluation using a neurodevelopmental disorders gene panel identified a novel variant in the *SIN3A gene* (c.269C>T; p.Ala90Val) located on chromosome 15. This variant, replacing alanine with valine, has not been reported in the literature for individuals affected with *SIN3A*-related conditions but is present in population databases (gnomAD 0.0009%). The patient exhibits clinical features overlapping with Witteveen-Kolk syndrome, including characteristic facial features such as a broad and tall forehead, thin upper lids, and a pointed chin. Despite the variant's classification as uncertain significance, the patient's clinical manifestations, including facial features, hypotonia, developmental delay, and intellectual disabilities, strongly suggest pathogenicity. Further neurological and hearing evaluations are recommended to assess potential seizures and hearing impairment commonly associated with this syndrome. The identification of this variant expands our understanding of the syndrome's genetic landscape and emphasizes the need for further clinical assessments to manage potential complications. Enhancing our knowledge of this rare disorder will facilitate improved diagnosis, prognosis, and potential therapeutic interventions for affected individuals within the Puerto Rican population.

Session Title: Mendelian Phenotypes Poster Session III

PB4703 Comprehensive genomic analysis of a trios family study with an index child diagnosed with autism spectrum disorders

Authors:

D. Manjegowda[1], U. Radhakrishna[2], S. Chettiar[3]; [1]Dayananda Sagar Univ., Bangalore, India, [2]Green Cross Pathology & Molecular Lab., Ahmedabad, India, [3]Genexplore Diagnostics and Res. Ctr., Ahmedabad, India

Autism Spectrum Disorder (ASD) is a complex neurodevelopmental disorder characterized by impaired social interaction, communication difficulties, and restricted repetitive behaviors. ASD is a complex and heterogeneous disorder, and it is likely that multiple genes, as well as environmental factors, contribute to its development, however, the underlying genomic landscape and the specific genetic variants associated with the disorder remain largely unknown. We performed a comprehensive genomic analysis to identify potential genetic factors contributing to the disorder. A total of 27 Indian trios between the ages of 8 and 12 with clinically diagnosed ASD index children were recruited and their blood genomic DNAs were prepared. Microarray analysis and Next-Generation Sequencing (NGS) were used to analyze the samples. The data was integrated using programmatic series of analyses (bioinformatics, statistical, computational biology, and pathway analyses), to explore the relation of genomic variants to ASD to lead to identification pathophysiology of disease and new therapeutic targets for preventing ASD. Microarray analysis revealed copy number variations (CNVs) less than 1MB duplications among chromosomal critical region 6p21.32, 7p21.2, 8p23.3, 14q12.14, Xp22.11. Whole-exome sequencing (WES) revealed single nucleotide variants (SNVs) in *CDK5RAP2, CC2D1A, AP4M1, SYNGAP1, STXBP1, ACSL4, ASTN1, MBD5, CDK5RAP2, CNTNAP2, SCN8A, SATB1, CREBBP, SHANK3, SHANK2, GNG13 and PPP3CB.* While most of these variants are novel and may contribute to the development of ASD, a few were previously reported in ways related to synaptic function, neuronal connectivity, and neurodevelopmental processes. Analyses of pathways and functional enrichment revealed dysregulation of glutamatergic synaptic transmission. Glutamate is the most abundant excitatory neurotransmitter in the brain and plays a crucial role in synaptic transmission and abnormalities in glutamatergic neurotransmission may contribute to the pathophysiology of ASD.Through the integration of clinical data with comprehensive genomic analysis in trios, genetic variations associated with ASD can be more precisely identified. As a result of these findings, genetic counseling may be able to develop targeted therapeutic interventions and personalized treatment strategies for individuals with ASD.

Session Title: Mendelian Phenotypes Poster Session II

PB4705 Consistent long-term clinical benefit with govorestat treatment: results of the ACTION-Galactosemia Kids trial

Authors:

R. Perfetti, E. Bailey, S. Wang, S. Shendelman; Applied Therapeutics Inc., New York, NY

Classic Galactosemia (CG) is an autosomal recessive disease caused by deficiency in the GALT enzyme, which is required for galactose metabolism. At abnormally high levels, galactose becomes an aberrant substrate for the enzyme aldose reductase, resulting in conversion to a toxic metabolite, galactitol. Newborn screening and implementation of a galactose-restricted diet has reduced newborn fatalities. However, despite early dietary intervention, children continue to develop progressive long-term complications. Govorestat (AT-007) is an oral, CNS penetrant aldose reductase inhibitor, which prevents conversion of galactose to the toxic metabolite galactitol.The ACTION-Galactosemia Kids Ph 3 study was designed to evaluate the impact of govorestat vs. placebo on clinical outcomes in 47 children age 2-17 with CG.The primary endpoint was a composite sum of change across four endpoints: OWLS-2 Oral Expression and Listening Comprehension, BASC-3 Behavior Symptoms Index and Activities of Daily Living. Cognition was included as a pre-specified sensitivity analysis to the primary endpoint. Additional secondary endpoints included adaptive skills and tremor. Data was assessed every 6 months by a firewalled independent DMC.Treatment with govorestat provided sustained clinical benefit on activities of daily living, behavioral symptoms, cognition, adaptive behavior and tremor over 18 months of treatment. While statistical significance ($p<0.05$) was not met on the primary endpoint, systematic improvement over time was demonstrated for the overall primary endpoint ($p=0.1030$) and for the sensitivity analyses including cognition ($p=0.0698$). Speech components were not impacted, which is suspected to be due to lack of progression in the placebo group and concomitant speech therapy received by almost all children in the trial. A post-hoc analysis of the primary endpoint excluding speech components demonstrated a highly statistically significant benefit of govorestat vs. placebo ($p=0.0205$). Govorestat provided significant benefit on tremor ($p=0.0428$), and adaptive skills (BASC-3 Adaptive Skills Index) ($p=0.0265$). Improvement in galactitol level was sustained throughout the trial with no impact on Gal-1p or galactose. Govorestat was generally safe and well tolerated, with no treatment-related SAEs.

Session Title: Mendelian Phenotypes Poster Session III

PB4706 Contribution of rare genetic variants to aortic valve stenosis in Quebec French-Canadians.

Authors:

E. MacKenzie[1], U. Houessou[2], K. Desbiens[3], L. Dufresne[3], K. Trajanoska[1], V. Mooser[4], S. Thériault[2], G. Thanassoulis[3], J. Engert[1]; [1]McGill, Montréal, QC, Canada, [2]Université Laval, Québec, QC, Canada, [3]McGill Univ. Hlth.Ctr. and Res. Inst., Montreal, QC, Canada, [4]McGill Genome Ctr., Westmount, QC, Canada

Background

Aortic valve stenosis (AS) is a severe disease affecting 2 to 5% of the North American population over age 65. Typically viewed as a polygenic condition, several common variants have been associated with AS through GWAS, yet monogenic contributors to AS have not been studied in the general population. Here, we aimed to identify rare genetic variations with a strong impact on AS in the Quebec French-Canadian population, which is genetically unique due to a limited number of original French colonisers, creating a founder effect.

Methods

Whole exome sequencing data from 310 AS patients of French-Canadian ancestry were analyzed. Rare variants (MAF < 0.001) were scored based on their predicted pathogenicity, minor allele frequency, and involvement in other diseases with similar phenotypes using Exomiser. Two independent datasets were then queried for the presence of high-scoring variants: UK Biobank (415,346 participants including 4,762 cases, imputed array and exome sequencing data) and the Quebec-based CARTaGENE cohort (11,587 participants including 263 cases, imputed array data). The association with AS of variants available in these datasets was assessed using logistic regression.

Results

A total of 722 variants found in the AS patients passed our score thresholds. Of these, 369 and 541 were present in CARTaGENE and UK Biobank respectively. 13 and 29 variants in the respective datasets were also nominally associated with AS ($p < 0.05$). One missense variant (rs368711105) in the elastin gene (*ELN*) was associated with AS in both CARTaGENE (p=0.023, OR=11.95) and UK Biobank (p=0.014, OR=10.80), and two additional genes each contained at least one associating variant in each cohort (myosin heavy chain (*MYH7*) rs149193520: p=0.0065, OR=5.23 in UK Biobank, rs200939753: p=0.0064, OR=50.53 in CARTaGENE; titin (*TTN*) rs190041566: p=0.016, OR=15.46 in UK Biobank, rs200843338: p=0.022, OR=33.65 in CARTaGENE). Of note, the *ELN* variant was observed at a higher frequency in CARTaGENE than in any other cohort listed in the dbSNP database.

Conclusion

These results support a potential role for elastin, myosin heavy chain and titin in the development of AS, as well as point to possible monogenic contributors to AS in a subset of patients. Variants in *ELN* have previously been associated with supravalvular aortic stenosis and play a role in valve pathogenesis in mice; variants in *MYH7* and *TTN* are involved in dilated and hypertrophic cardiomyopathy. Future work will attempt to identify the mechanism underlying these novel genetic associations, and seek to identify additional rare variants associated with AS, both Quebec-specific and those shared with other populations.

Session Title: Mendelian Phenotypes Poster Session I

PB4707 Copy number and non-coding variants leading to *NMNAT1*-related retinopathy

Authors:

E. Place; Mass Eye & Ear, Boston, MA

Purpose: Mutations in *NMNAT1* are associated with autosomal recessive retinal dystrophy. The most recognized retinal phenotype is Leber congenital amaurosis (LCA), which is characterized by poor visual acuity in the first year of life with characteristic colobomatous macular atrophy. Variants in this gene have also been associated with milder and later onset pan-retinal disease as well as cone and cone-rod dystrophy. Over 90 variants have been reported to be disease causing with the vast majority single nucleotide variants (SNV) and small insertions/deletions. For inherited retinal disorders, it has been shown that structural variation including copy number as well as non-coding variants contribute to disease causality. The purpose of this study was to assess the role and contribution of copy number variants (CNV) and non-coding variants in *NMNAT1*-related retinopathy.
Methods: Over 3000 subjects with retinal dystrophy were screened for *NMNAT1* variants through Next-generation sequencing (NGS) targeted gene sequencing, whole exome sequencing, or whole genome sequencing. NGS data was used to detect CNVs with gCNV. Likely causal CNVs were validated by quantitative polymerase chain reaction (qPCR). Likely causal non-coding variants were assessed with in silico predictions as well as in vitro studies. Individuals were eligible for inclusion if they had either one CNV or non-coding variant, or, if they had a less typical phenotype.
Results: 7 individuals with biallelic *NMNAT1* variants were identified. Three unrelated subjects had at least one structural variant. Two individuals both had a tandem duplication spanning exon 2 to 3'UTR , and a single exon deletion was identified in the third. Two of these individuals had clinical diagnoses of LCA while the third had a diagnosis of macular dystrophy. Non-coding likely causal variants were identified in 2 subjects: one with a clinical diagnosis of LCA and the other with macular dystrophy. Additionally, analysis identified two unrelated subjects with cone-rod dystrophy and compound heterozygous causal missense coding variants.
Conclusions: Inherited retinal diseases (IRDs) are phenotypically heterogenous and it is well established that a single IRD gene can be associated with a spectrum of retinal phenotypes. The phenotypic and molecular findings from this study adds to the growing evidence supporting the phenotype spectrum associated with *NMNAT1* and highlights the contribution copy number and non-coding variants in *NMNAT1*-related retinopathy.

Session Title: Mendelian Phenotypes Poster Session II

PB4708 CRISPR/Cas9 gene-edited zebrafish model for Noonan syndrome.

Authors:

P-X. Lin, C-Y. Li, Y-C. Wang, P-C. Chen; Inst. of Clinical Med., Coll. of Med., Natl. Cheng Kung Univ., Tainan, Taiwan

Noonan syndrome (NS) is an autosomal dominant genetic disease characterized by short stature, facial dysmorphia, congenital heart disease, and developmental delay. In our previous study of identifying disease-causing genes for Noonan syndrome, we identified variants in *RASIP1*, which encodes RAS interacting protein 1, in familial and sporadic NS cases. RASIP1 plays an important role in maintaining endothelial cell functions that is essential throughout vasculogenesis. Patients with *RASIP1* variants developed atypical cardiovascular phenotypes of NS, such as aortic coarctation and aortic aneurysm. To prove the causality of RASIP1, a zebrafish model was generated through CRISPR/Cas9. Early morphogenesis at different embryonic developmental stages and craniofacial features in the mutant zebrafish were monitored. Additionally, double transgenic zebrafish *Tg(fli1:EGFP; gata1a:dsRed)*, which expresses the fluorescent protein in endothelial and blood cells, was used to investigate vascular development. Our results suggested that the deficiency of *rasip1* in the zebrafish leads to developmental delay, abnormal craniofacial features, and impaired vasculogenesis that recapitulate NS patients' phenotypes.

Session Title: Mendelian Phenotypes Poster Session III

PB4709 CRISPR-Cas9 knock-out model of NGLY1 Deficiency in *Danio rerio* (zebrafish) harbors mild phenotypic changes detectable with automated phenotyping

Authors:

M. Mooney, B. Figueroa, A. Bartkowiak; Univ. of North Florida, Jacksonville, FL

Background: Loss-of-Function variants in NGLY1 result in NGLY1 Deficiency, a rare recessive disorder with variable phenotypes. Human symptoms can include developmental delay, visual impairment, and deficits in neurological function. Animal models of NGLY1 Deficiency from invertebrates to mice are often lethal or infertile, though different genetic backgrounds can ameliorate these phenotypes. In contrast, zebrafish models of NGLY1 Deficiency are viable and fecund and display milder phenotypes. **Methods:** We created *ngly1-/-* knock-out zebrafish using CRISPR-Cas9 targeted to exon 2 to generate a premature stop codon just outside the conserved PUB domain. These fish have been maintained past generation F5 and produce stable homozygotes from Het x Het crosses in the expected 1:2:1 genotypic ratio. Using these crosses and automated microscopy produced large image sets from single clutches with randomized genotypes. Phenotypes assessed included body length, eye size, and brain metrics (cortical area, ventricular area, optic tectal area, neuronal cell density), using measures made both manually and via automated processing with ImageJ macros. **Results:** Automated microscopy and image processing revealed significant phenotypes with small effect between wild-type and *ngly1-/-* genotypes from the same clutch. Body length displayed dose-dependent reduction in *ngly1-/-* fish: the Kruskal-Wallis H test/post-Hoc Dunn's test, α=0.05, indicated the mean rank between *ngly1+/+* vs *ngly1-/-* is significantly different (p=0.013). There is also significant reduction in eye area, with mean ranks 43.13 for *ngly1+/+*, 26.93 for *ngly1+/-*, 28.85 for *ngly1-/-* (p=.008). The eye radius measure returned significant reduction in *ngly1-/-* fish without a detectible phenotype in heterozygotes (*ngly1+/+* vs *ngly1-/-*, p=0.019; *ngly1+/-* vs *ngly1-/-*, p=0.002). Likewise, *ngly1-/-* fish also had reduced cortical area (p=0.002), but no difference in neuronal density, or in ventricular and optic tectal area. **Conclusion:** This *ngly1-/-* zebrafish model reveals new, mild morphological phenotypes detectable through automated microscopy. The reduced body length of *ngly1-/-* fish suggests developmental delay, as in humans. Differences in the alignment of heterozygous fish with either wild-type or *ngly1-/-* eye phenotypes and the small effect size suggest correlation with developmental delay. Likewise, cortical area reduction in *ngly1-/-* fish may correlate with delay, however neuronal density appears unaffected at 3dpf. Comparative studies with other models of NGLY1 Deficiency that express extreme phenotypes might suggest key compensatory mechanisms at work in *ngly1-/-* zebrafish.

Session Title: Mendelian Phenotypes Poster Session I

PB4710 *CTSA* variants c.430del (p.Leu144Trpfs*49) and c.587G>C (p.Ser196Thr) as a cause of late infantile Galactosialidosis in a Hispanic American patient: a case report

Authors:

C. Hollas[1], Z. Yu[1], J. Gamblin[1], S. Palle[1], Y. Enchautegui-Colon[2]; [1]OUHSC, Oklahoma City, OK, [2]Oklahoma Hlth.Sci. Ctr., Oklahoma City, OK

We present a 3-year-old girl who was referred to genetics evaluation due to elevated liver enzymes, hepatomegaly, and abnormal liver biopsy with foamy hepatocytes and Kupffer cells concerning for glycogen storage disease (GSD). At birth, the patient had a 2-week NICU stay due to direct hyperbilirubinemia, elevated TSH, withdrawal from intrauterine exposure to illicit drugs, and feeding problems. The patient's bilirubin normalized. Then at 19 months of age, her liver enzymes were found to be elevated. A cholestasis molecular panel was negative. Workup for infectious and autoimmune causes of hepatobiliary disease was negative. Physical examination was remarkable for coarse facial features, thick eyebrows, flat nasal bridge, hepatomegaly, and extensive Mongolian spots, raising high concern for a lysosomal storage disorder (LSD). A skeletal survey revealed "oar shape ribs" and vertebral body beaking. She was identified to have macular cherry spots by ophthalmology evaluation. She also has a history of global developmental delays. Molecular testing for GSDs and LSDs was performed. Two variants in the *CTSA* gene were identified: c.430del (p.Leu144Trpfs*49) classified as pathogenic, and c.587G>C (p.Ser196Thr) classified as a variant of uncertain significance (VUS). Given the high suspicion for Galactosialidosis, urine oligosaccharides were obtained and were abnormal. Neu5Ac1Hex3HexNAc2 was significantly elevated and Hex3HexNAc2 was mildly elevated. No variants were identified in the *NEU1* gene. The patient is adopted and biological parents are not available. The clinical features, presence of these two variants in the *CTSA* gene, the liver biopsy with foamy hepatocytes and Kupffer cells, and abnormal urine oligosaccharides support the diagnosis of Galactosialidosis in this patient. Galactosialidosis is an extremely rare disorder. This patient is unique because her combination of variants has not been reported before and the majority of patients have the juvenile/adult form. Additionally, the pathogenic variant in this patient is novel. The VUS found in this patient has been reported in trans with a pathogenic variant in at least one individual affected with Galactosialidosis. This, along with our patient's clinical findings, supports the pathogenicity of the *CTSA* c.587G>C (p.Ser196Thr) variant. Given that this disorder has been found to have genetic heterogeneity and a spectrum of severity, knowledge of the variants in this patient will help to further classify the spectrum and natural history of disease for future patients with this disorder or similar molecular findings.

Session Title: Mendelian Phenotypes Poster Session II

PB4711 Data sharing in the GREGoR Consortium to support rare genetic disease research.

Authors:

B. Heavner[1], S. I. Berger[2], J. X. Chong[1], S. N. Jhangiani[3], A. O'Donnell-Luria[4], C. Austin-Tse[5], S. Baxter[4], M. Conomos[1], Z. Coban-Akdemir[6,7], E. C. Délot[8], P. C. Goddard[9], S. M. Gogarten[1], A. Jolly[7], M. Mahmoud[7], S. Marwaha[10], D. E. Miller[1], J. E. Posey[7], A. Sanchis-Juan[11], A. M. Stilp[1], R. A. Ungar[12], Genomics Research to Elucidate the Genetics of Rare Diseases (GREGoR) Consortium; [1]Univ. of Washington, Seattle, WA, [2]Children's Natl. Hosp., Washington, DC, [3]Baylor Coll. Med., Houston, TX, [4]Broad Inst. of MIT and Harvard, Cambridge, MA, [5]Broad Inst. of MIT and Harvard Cambridge, Cambridge, MA, [6]Univ. of Texas Sch. of Publ. Hlth., Houston, TX, [7]Baylor Coll. of Med., Houston, TX, [8]George Washington Univ., Washington, DC, [9]Stanford Univ., Menlo Park, CA, [10]Stanford Univ., Palo Alto, CA, [11]Ctr. for Genomic Med., Massachusetts Gen. Hosp., Boston, MA, [12]Stanford, Stanford, CA

The members of the Genomics Research to Elucidate the Genetics of Rare Diseases (GREGoR) Consortium are working collaboratively to advance rare genetic disease research. To succeed, this effort requires rapid Consortium-wide sharing of data generated by participating Research Centers and controlled-access release of a coherent Consortium data set to the broader scientific community. The GREGoR Consortium has developed a flexible, expandable, and publicly available data model that serves as a robust and scalable metadata framework, which can be extended to support data from a wide range of experiment types.

The GREGoR Consortium leverages NHGRI's AnVIL platform for a secure and scalable cloud environment for Consortium data sharing and analysis. Since each participating Research Center uses the Consortium metadata standards defined by the Consortium Data Model for data they upload to AnVIL, the Consortium is able to quickly share and update data on a regular quarterly cycle. In this cycle, each Research Center uploads data to an AnVIL workspace, where it can be validated for conformance to the data model using a workflow-based validation process. Once the data is validated and uploaded into AnVIL workspaces, the Data Coordinating Center can merge the data submitted by the Research Centers into a single, shared workspace for Consortium-wide sharing, collaborative analysis, and regular, periodic release for controlled access by the broader scientific community.

Through this process, the GREGoR Consortium has produced a comprehensive and growing data set that is available to the broader scientific community. The Consortium Data Set currently includes short-read and long-read genome sequencing data, RNA-seq data, family structure, phenotype information, and genetic findings from our collaborating Research Centers. We anticipate that the data set will expand to include additional data types produced by participating Research Centers. The GREGoR Consortium Data Set is broadly consented and allows researchers worldwide to access a wealth of information, accelerating their own investigations into rare diseases or other areas of research. Researchers may obtain access to the GREGoR Consortium Data Set by submitting a Data Access Request via dbGaP.

Session Title: Mendelian Phenotypes Poster Session III

PB4712 De novo and inherited variants in *DDX39B* cause a novel syndrome characterized by neurodevelopmental delay, congenital hypotonia, and short stature.

Authors:

M. Chui[1], K. T. A. Booth[2], S. V. Jangam[3], K. Treat[2], K. White[2], C. K. Christensen[4], T. Lynnes[2], S. Yamamoto[3], O. Kanca[3], M. H. Y. Tsang[1], S. A. Lynch[5], S. V. Mullegama[6], Y. Yuen[7], D. Iancu[8], S. K. Joss[5], J. Baptista[9], C. C. Y. Mak[1], A. K. Y. Kwong[1], H. J. Bellen[3], M. F. Wangler[3], BCM Center for Precision Medicine Models, E. Conboy[2], B. H. Y. Chung[1], F. Vetrini[2]; [1]The Univ. of Hong Kong, Hong Kong, China, [2]Indiana Univ. Sch. of Med., Indianapolis, IN, [3]Baylor Coll. of Med., Houston, TX, [4]Riley Dev.al Pediatrics, Indianapolis, IN, [5]Queen Elizabeth Univ. Hosp., Glasgow, United Kingdom, [6]GeneDx, Gaithersburg, MD, [7]Hong Kong Children's Hosp., Hong Kong, China, [8]Univ. Coll. of London, London, United Kingdom, [9]Univ. of Plymouth, Plymouth, United Kingdom

DDX39B is a member of the DEAD-box family of ATP-dependent RNA helicases. DEAD-box proteins are ubiquitously expressed from yeast to humans and perform essential functions associated with mRNA metabolism. *DDX39B* is also a crucial component of the TRanscription-EXport (TREX) super protein complex, which recent studies have highlighted the important role of its subunits in neurodevelopmental disorders. Here, we describe six individuals from five families, four with novel de novo missense variants in *DDX39B*, and one carrying an inherited splicing variant, all presenting with mild to severe global developmental delay, congenital hypotonia, epilepsy, short stature, skeletal abnormalities and variable dysmorphic features. 3D molecular modeling predicts these variants would alter protein structure. *DDX39B* is a conserved gene and Drosophila melanogaster (fruit flies) studies were conducted. We generated a new Hel25E Kozak-GAL4 allele which disrupts the fly gene and allows expression of transgenes. We also generated transgenic *DDX39B*-reference and variant flies. Human transgenic *DDX39B* was unable to rescue *Hel25E* loss in the fly. However, human reference *DDX39B* when overexpressed ubiquitously leads to lethality but the variants found in the patients do not recapitulate the lethality suggesting that the mutants are loss of function alleles. Blood transcriptomics of the patients revealed an excess of aberrant splicing events, indicating a disrupted mRNA processing as anticipated from the role of *DDX39B* in mRNA metabolism. Our human genetic data, coupled with *in silico* and *in vivo* data supports that *DDX39B* is a novel candidate gene in a potential group of disorders named TREX-complex related neurodevelopmental syndrome.

Session Title: Mendelian Phenotypes Poster Session I

PB4713 *De novo* dominant negative variants of *IKZF2* cause ICHAD syndrome, a new disorder characterized by Immunodysregulation, Craniofacial Anomalies, Hearing Impairment, Athelia, and Developmental Delay.

Authors:

A. Mohajeri[1], M. Vaseghi-Shanjani[1], J. A. Rosenfeld[2], G. X. Yang[1], H. Lu[1], M. Sharma[1], S. Lin[1], A. Salman[1], M. Waqas[1], M. Sababi Azamian[2], K. C. Worley[2], K. L. Del Bel[1], F. K. Kozak[1], R. Rahmanian[1], M. Biggs, Catherine[1], K. J. Hildebrand[1], S. R. Lalani[2], S. K. Nicholas[3], D. A. Scott[2], S. Mostafavi[1], C. van Karnebeek[1], E. Henkelman[1], J. Halparin[1], C. L. Yang[1], L. Armstrong[1], Undiagnosed Diseases Network, Care4Rare Canada Consortium, S. Turvey[1], A. Lehman[1]; [1]The Univ. of British Columbia, Vancouver, BC, Canada, [2]Baylor Coll. of Med., Houston, TX, [3]Texas Children's Hosp., Houston, TX

Background: *IKZF2* encodes Helios, a transcription factor and chromatin remodeler that plays roles in the development and regulation of immune function. Germline and somatic variants in this gene have been implicated in an inborn disorder of immunity and in hematopoietic malignancies, respectively. The present study reports two *de novo* dominant negative variants of *IKZF2* causing a novel syndrome in two probands with T cell lymphopenia, recurrent respiratory infections, craniofacial differences, cleft palate, sensorineural hearing impairment, athelia, and developmental delay. **Methods:** Proband 1 was enrolled in the C4R-SOLVE project, and Proband 2 was identified via GeneMatcher. Genome sequencing and exome sequencing was conducted for Proband 1 and 2, respectively. To investigate the functional impact of *IKZF2* variants, RT-PCR, Sanger sequencing, Western blot analysis and Luciferase assays were undertaken. **Results:** Our analysis revealed a *de novo* tandem duplication and a *de novo* missense variant in exon 5 of *IKZF2*, which encodes the critical Zinc fingers 2 and 3 that are responsible for the DNA binding activity of Helios. Tandem duplications of exon 5 have been previously reported in cases of T-cell lymphoma, and a similar missense variant affecting a homologous residue in family member Aiolos has been associated with a disorder of immunity. RT-PCR and Sanger confirmed an in-frame duplication of exon 5. Western blot analysis demonstrated expression of both altered proteins. Luciferase assay verified that co-expression of the altered proteins with the wild-type protein significantly reduced the suppressor ability of the wild-type on *IL2* promoter, indicating a dominant negative effect of these variants. **Conclusion:** Previous studies have reported germline variants in *IKZF2* and other Ikaros family members resulting in immunodeficiency. The present study is the first to report two *de novo* dominant negative *IKZF2* variants that cause ICHAD syndrome, a multiorgan disorder Characterized by Immunodysregulation, Craniofacial Anomalies, Hearing Impairment, Athelia, and Developmental Delay.

Session Title: Mendelian Phenotypes Poster Session II

PB4714 De novo variant in TLK2: clinical evaluation & genotype-phenotype of a neurodevelopmental disorder

Authors:

T. Carrion[1,2]; [1]Hosp. universitario son espases, palma, Spain, [2]Centro medico Teknon, Barcelona, Spain

We report the case of a 22- year -old woman, followed for genetics, psychiatry and neurologist due to intellectual disability, dysmorphic symptoms and attention deficit hyperactivity disorder. She was born of non-consanguineous parents. No history of intellectual disability or neurological diseases. Two healthy male brothers aged 29 and 26. She was delivery at 38 weeks without complications. Weight 2.400 gr, T: 44 cm and PC 31 cm. Delay in acquiring developmental milestones. Examination at 22-yo showed peculiar fascies, microcephaly, broad forehead, hypertelorism. Palpebral fissures short, oblique upwards. Bilateral epicanthal folds, ears of low implantation. Large nose with bulbous tip, small mouth, with thin lips, macroglossia, broad lower jaw, short and flat philtrum, prognathism and high palate. Throughout years, extensive investigations were performed, metabolic studies and skeletal series were normal. CT, MRI and MR angio 2002 it has a ventricular size in high limits. 2nd MRI and MR angio 2005 moderate dilation of the ventricular system, in the same, which may be significant for some subcortical atrophy. Karyotype 46 XX - X-fragile: negative. Most common microdeletion syndromes (MLPA) : normal.Molecular study to detect microdeletion in region 22q11.2 : normal.Smith-Magenis SD and congenital central hypoventilation syndrome: negative FISH: no deletion of subtelomeric regions. Array-CGH 60 Kb : Normal High resolution karyotype 46XX Microarray CGH 1000K arr 8p23.3 1,504-084-1,509,519 x1 female karyotype with microdeletion in 8p23.3 de novo. Study of family segregation: negative. Whole exome sequencing detected LP heterozygous variant p**.** Arg546Gly in the TLK2 gene. The variant found p.Arg546Gly is not found in the databases of general population consulted. It has not been described in the scientific literature or classified in databases of genetic variants of clinical relevance. Predictors indicate a deleterious effect on the protein encoded by this gene. This variant is located in a functionally relevant domain of the protein involving a hot spot. TLK2 is a protein-coding gene. It encodes a nuclear serine/threonine kinase. The encoded protein functions in regulating chromatin assembly in the S phase of the cell cycle by regulating the levels of a histone H3-H4 chaperone. Associated with repairing double-stranded breaks of DNA damage caused by radiation. Associated diseases include autosomal dominant intellectual developmental disorder 57. Among its related pathways are the regulation of miRNA, the response to DNA damage and the regulation of Chks at checkpoints.

Session Title: Mendelian Phenotypes Poster Session III

PB4715 *De novo* variants in *FRYL* are associated with developmental delay, intellectual disability, and dysmorphic features.

Authors:

X. Pan[1,2], A. M. Tao[3], S. Lu[1,2], M. Ma[1,2], R. Slaugh[4], S. Drewes Williams[5], L. O'Grady[6,7], O. Kanca[1,2], M. T. Carter[8], K. Platzer[9], A. Revah-Politi[10], J. Granadillo De Luque[4], S. P. A. Stegmann[11], A. Accogli[12,13], V. Salpietro Damiano[14], L. Ghaloul-Gonzalez[5,15,16], M. E. H. Simon[17], D. A. Sweetser[6,18], K. E. Glinton[1,19], S. E. Kirk[20,21], Baylor College of Medicine Center for Precision Medicine Models, M. F. Wangler[1,2], S. Yamamoto[1,2,22], W. K. Chung[23], H. J. Bellen[1,2,22]; [1]Dept. of Molecular and Human Genetics, Baylor Coll. of Med., Houston, TX, [2]Jan & Dan Duncan Neurological Res. Inst., Texas Children's Hosp., Houston, TX, [3]Vagelos Sch. of Physicians and Surgeons, Columbia Univ., New York, NY, [4]Div. of Med. Genetics and Genomics, Dept. of Pediatrics, Washington Univ. in St. Louis Sch. of Med., St. Louis, MO, [5]Div. of Genetic and Genomic Med., UPMC Children's Hosp. of Pittsburgh, Pittsburgh, PA, [6]Div. of Med. Genetics & Metabolism, Massachusetts Gen. Hosp. for Children, Boston, MA, [7]MGH Inst. of Hlth.Professions, Charlestown, MA, [8]Dept. of Genetics, Children's Hosp. of Eastern Ontario, Ottawa, ON, Canada, [9]Inst. of Human Genetics, Univ. of Leipzig, Leipzig, Germany, [10]Inst. for Genomic Med. and Precision Genomics Lab., Columbia Univ. Irving Med. Ctr., New York, NY, [11]Dept. of Clinical Genetics, Maastricht Univ. Med. Ctr., Maastricht, Netherlands, [12]Div. of Med. Genetics, Dept. of Med., McGill Univ. Hlth.Ctr., Montreal, QC, Canada, [13]Dept. of Human Genetics, McGill Univ., Montreal, QC, Canada, [14]Dept. of Neuromuscular Disorders, Univ. Coll. London Inst. of Neurology, Queen Square, United Kingdom, [15]Div. of Genetic and Genomic Med., Dept. of Pediatrics, Univ. of Pittsburgh, Pittsburgh, PA, [16]Dept. of Human Genetics, Graduate Sch. of Publ. Hlth., Univ. of Pittsburgh, Pittsgurgh, PA, [17]Dept. of Clinical Genetics, Erasmus Med. Ctr. Rotterdam, Rotterdam, Netherlands, [18]Ctr. for Genomic Med., Massachusetts Gen. Hosp., Boston, MA, [19]Dept. of Genetics, Texas Children's Hosp., Houston, TX, [20]Section of Hematology/Oncology, Dept. of Pediatrics, Baylor Coll. of Med., Houston, TX, [21]Texas Children's Cancer and Hematology Ctr., Houston, TX, [22]Dept. of NeuroSci., Baylor Coll. of Med., Houston, TX, [23]Dept.s of Pediatrics and Med., Columbia Univ., New York, NY

FRYL belongs to a Furry protein family which is evolutionarily conserved from yeast to human. The functions of FRYL in mammals are largely unknown, and variants in *FRYL* have not previously been associated with a Mendelian disease. Here, we report fourteen individuals with heterozygous variants in *FRYL* who present with developmental delay, intellectual disability, dysmorphic features, and other congenital anomalies. The variants are confirmed *de novo* in all individuals except one whose mother has not been tested. Among the variants, nine are likely gene disrupting variants (stop gain, frameshift, or splicing variants) and five are missense variants. Human genetic data suggest that *FRYL* is intolerant to loss of function (LoF). We found that the fly *FRYL* ortholog, *furry* (*fry*), is expressed in multiple tissues including the central nervous system where it is present in neurons but not in glia. We demonstrated that *fry* LoF mutant flies are lethal at early developmental stages and loss of *fry* in mutant clones causes defects in the wings and compound eyes. We next modeled four out of the five missense variants using gene knock-in alleles. The knock-in alleles were generated using prime editing and a novel Recombinase-Mediated Cassette Exchange (RMCE)-based strategy, which introduce amino acid changes analogous to the human variants into the endogenous *fry* locus in flies. One variant allele (*fry* p.Phe2746Ser, analogous to *FRYL* p.Phe2295Ser) fails to complement *fry* severe LoF alleles and phenocopies a severe LoF allele in homozygous clones, suggesting that it causes a severe LoF. Two variants (*fry* p.Phe2024Leu and p.Tyr3410Cys, analogous to *FRYL* p.Phe1628Leu and p.Tyr2951Cys, respectively) complement severe LoF alleles. However, both cause developmental delay in compound heterozygous animals when crossed to a severe LoF allele, whereas one allele (*fry* p.Phe2024Leu) causes defects in the response to light in compound eyes. These data suggest that the two variants behave as partial LoF variants. One variant (*fry* p.Ser2910Ile) does not cause any observable defect in flies and the corresponding human variant (*FRYL* p.Ser2397Ile) is not confirmed *de novo*, suggesting the variant is benign and may not be the cause of the clinical features of the affected individual. In summary, our findings support that *fry* is required for proper fly development, and that the LoF variants in *FRYL* may cause a disorder with developmental and neurological symptoms.

Session Title: Mendelian Phenotypes Poster Session I

PB4716 Deciphering the genetic architecture of autism spectrum disorder using chromosomal microarray and whole exome sequencing in patient-parent trios in India

Authors:

J. Shah[1], D. Jain[2], S. Shah[3], H. Patel[4], K. Patel[5], D. Solanki[6], A. Iyer[7], B. Menghani[8], P. Mhatre[9], S. Mehta[3], S. Bajaj[10], V. Patel[11], M. Pandya[12], D. Dhami[13], D. Patel[14], J. Sheth[1], F. Sheth[1], H. Sheth[1]; [1]FRIGE s Inst. of Human Genetics, Ahmedabad, India, [2]Shishu Child Dev. And Early Intervention Ctr., Ahmedabad, India, [3]Royal Inst. of Child NeuroSci.s, Ahmedabad, India, [4]Zydus Hosp., Ahmedabad, India, [5]Speciality Homeopathy Clinic, Ahmedabad, India, [6]Mantra Child Neurology and Epilepsy Hosp., Bhavnagar, India, [7]NeuroKids, Ahmedabad, India, [8]Children's Inst. for Dev. and Advancement Ctr., Vadodra, India, [9]Tender Kids Ctr. for Child Dev., Navi Mumbai, India, [10]The Purple Gene Clinic, Mumbai, India, [11]Little Brain Pediatric Neurocare Ctr., Vadodra, India, [12]Kadam Maternity Home, Ahmedabad, India, [13]Axon Child Neurology and Epilepsy Ctr., Rajkot, India, [14]Charotar Inst. of ParaMed. Sci., Charotar Univ. of Sci. and Technology, Changa, India

Autism spectrum disorder (ASD) is a heterogeneous group of neurodevelopmental disorders (NDD) and is characterized by impaired social communication along with repetitive behavior or restricted interests which can persist throughout lifetime. Guidelines put forth a decade ago by the American College of Medical Genetics suggests using chromosomal microarray (CMA) as a first line test for genetic diagnosis of ASD. However, latest results from whole exome sequencing (WES) studies suggests higher diagnostic yield compared to CMA due to detection of *de novo* SNVs; a significant cause for NDD. No study has been performed to delineate the genetic architecture of ASD in the Indian population which can help in selection of first-tier test. We present results of the first systematic study to assess the genetic architecture and diagnostic yields of karyotype, Fragile-X testing, CMA and WES in a cohort of 101 patient-parent trios with ASD from India. No positive genetic diagnosis was obtained from karyotyping and Fragile-X testing. Of 101 trios, 3 (2.9%) and 30 (29.7%) trios received a confirmed genetic diagnosis from CMA and WES, respectively. Amongst diagnosis from WES, SNVs were detected in 27 cases (90%) and CNVs in 3 cases (10%), including the 3 CNVs detected from CMA. Segregation analysis showed 66.6% (n=3 for CNVs and n=17 for SNVs) and 16.6% (n=5) cases had *de novo* and recessive variants, respectively, which is in concordance with the distribution of variant types and mode of inheritance observed in ASD patients of non-Hispanic white/ European ethnicity. *De novo* SNVs were found primarily in previously known ASD genes- *MECP2, SCN2A, KCNQ2, TBL1XR1, CNTNAP2, TCF4, CAMK2A, NF1, AUTS2, FOXP2* and *NLGN3.* Of 17 *de novo* SNVs, 6 were predicted to be loss of function (35.2%) whereas the remaining were missense variants. *MECP2* gene was the most recurrently mutated gene (n=6; 20%). Additionally, WES identified 22 variants of uncertain significance in 21 patients (n=21/101; 20.8%). These were identified in genes that have previously been associated with or implicated in ASD etiology. Of these, majority of the probands were detected with heterozygous variants (66.6%) which were inherited from either of the unaffected parents with equal distribution. Lastly, 4 genes (*NEUROG1*, *LRFN1*, *UNC13A* and *UNC79*) were identified as potential novel candidates for ASD. Majority of the detected genes are involved in synaptic formation, transcription and its regulation, ubiquitination and chromatin remodeling. Ours is the first study to suggest *de novo* variants as a major cause of ASD and provide evidence for implementation of WES as a first-tier test for genetic diagnosis of ASD in the Indian population.

Session Title: Mendelian Phenotypes Poster Session II

PB4717 Defining *BSN*: a novel and ultrarare synapse disorder

Authors:

S. Guzman[1], S. Ruggiero[2], S. Jain[3], C. Jensen[4], M. McDonald[4], K. Butler[5], M. Lyons[6], Z. Stark[7], N. Brown[8], A. Ngoh[9], M. Yeganeh[10], B. Keren[11], J. Tenorio[12], P. Lapunzina[13], J. Nevado[14], S. Heide[11], K. R. Sullivan[2], M. P. Fitzgerald[2], S. Ganesan[2], C. Ellis[2], S. Parthasarathy[2]; [1]Univ. of Pennsylvania, Philadelphia, PA, [2]Children's Hosp. of Philadelphia, Philadelphia, PA, [3]Alberta Hlth.Services, Edmonton, AB, Canada, [4]Duke Univ., Chapel Hill, NC, [5]Greenwood Genetic Ctr., Greenwood, SC, [6]Greenwood Genetic Ctr., Torrance, CA, [7]Murdoch Children's Res. Inst., Melbourne, Australia, [8]Murdoch Children's Res. Inst., Melbourne, Australia, [9]UCL Great Ormond Street Hosp. Inst. of Child Hlth., London, United Kingdom, [10] Le Ctr. Hosp.ier de l'Université Laval, Quebec, QC, Canada, [11]Pitié-Salpêtrière Univ. Hosp., Paris, France, [12]1-CIBERER, Madrid, Spain, [13]INGEMM-IdIPAZ, Madrid, Spain, [14]INGEMM-HULP, Madrid, Spain

Disease-causing variant in genes involved in synaptic function are common causes of severe neurodevelopmental disorders and epilepsy. However, in contrast to other genetic causes of neurodevelopmental disorders, milder presentations have largely not been observed so far. Here, we report on 9 individuals identified through Genematcher with *de novo* disruptive variants in *BSN*, encoding the presynaptic protein bassoon. All individuals presented with mild developmental differences and variable types of epilepsy. In order to aggregate available information on the trajectory of *BSN*-related disorders, we screened clinical and genetic information from the Penn Medicine Biobank (PMBB), a large biorepository with over 60,000 individuals and retrieved clinical information on 8 individuals that carry protein-truncating *BSN* variants. Only 1 individual had a neurological diagnosis code, suggesting a generally mild presentation in adulthood. Reconstruction of phenotypic presentations were completed using the Human Phenotype Ontology (HPO), a standardized phenotypic dictionary, which allows us to map heterogeneous clinical data to a common framework. We identified seizure disorders present in 7 individuals, with the majority having febrile seizures (n=5) followed by bilateral tonic-clonic seizures (n=3). Most cases presented with behavioral abnormalities, such as autistic (n=4) and hyperactive (n=4) behavior. We applied a previously reported phenotypic similarity algorithm to derive a similarity score (sim score) based on pairwise phenotypic similarities between individuals through various combinations of the most specific terms shard by both individuals. When comparing sim scores of different genetic conditions, *BSN* altogether is more significant than expected by chance through phenotypic terms alone. In summary, we identify *de novo BSN* variants as the cause for a new class of synapse disorders with an unusually mild clinical presentation. In addition, we demonstrate how existing data repositories can be leveraged to maximize available information on rare neurodevelopmental conditions that would otherwise remain unexplored.

Session Title: Mendelian Phenotypes Poster Session I

PB4719 Deleterious ZNRF3 germline variants as a novel cause of neurodevelopmental disorders with mirror brain phenotypes due to distinct domain-specific effects on Wnt/β-catenin signaling

Authors:

P. Boonsawat[1], R. Asadolahi[1,2], D. Niedrist[1], K. Steindl[1], A. Begemann[1], M. Zweier[1], P. Joset[3], M. Papik[1], E. Bhoj[4], D. Li[5], E. Zackai[6], A. Vetro[7], R. Guerrini[7], S. Whalen[8], B. Keren[9], A. Khan[10], D. Jing[11], M. Bralo[12], Q. Hao[13], G. Siegel[1], A. Horn[1,14], H. Sticht[14], A. Rauch[1,15]; [1]Inst. of Med. Genetics, Univ. of Zurich, Schlieren-Zurich, Switzerland, [2]Faculty of Engineering and Sci., Univ. of Greenwich London, Medway Campus, Chatham Maritime, ME4 4TB, London, United Kingdom, [3]Med. Genetics, Univ. Hosp. Basel, Basel, Switzerland, [4]Children`s Hosp. of Philadelphia, Philadelphia, PA, [5]Children s Hosp. of Philadelphia, Philadelphia, PA, [6]Children S Hosp. of Philadelphia, Philadelphia, PA, [7]Pediatric Neurology, Neurogenetics and Neurobiology Unit and Lab., Meyer Children's Hosp., Univ. of Florence, Florence, Italy, [8]Groupe Hosp.ier Pitié Salpêtrière, Paris, France, [9]Département de génétique, Hôpital de la Pitié-Salpêtrière, Paris, France, [10]Faculty of Sci., Dept. of Biological Sci. (Zoology), Univ. of Lakki Marwat, Khyber Pakhtunkhwa, Pakistan, [11]Shenzhen Children's Hosp., Shenzhen, Guangdong, China, [12]Inst. de Genética Médica y Molecular (INGEMM), Unidad de trastornos del neurodesarrollo, Hosp. Univ.rio La Paz, Madrid, Spain, [13]Odense Univ. Hosp., Odense, Syddanmark, Denmark, [14]Inst. of Biochemistry, Friedrich-Alexander-Univ. Erlangen-Nürnberg (FAU), Erlangen, Germany, [15]Univ. Hosp. Zurich, Zurich, Switzerland

Zinc RING finger 3 (ZNRF3) is a negative-feedback regulator of the Wnt/β-catenin signaling, which plays an important role in the development of the human brain. Although somatic mutations are frequently observed in cancer, germline variants in ZNRF3 for neurodevelopmental disorders have not been reported so far. Here, we describe mirror brain phenotypes in a total of seven patients harboring de novo deleterious missense ZNRF3 variants and demonstrate that these variants led to distinct domain-specific opposing effects on the Wnt/β-catenin signaling. Using in vitro transcriptional reporter assays and structural modelling we found that one missense variant located in the protease-associated (PA)-domain in a patient with microcephaly attenuated the Wnt/β-catenin signaling through loss of the binding affinity to R-spondin, a positive regulator of the Wnt/β-catenin signaling. Two of six other missense variants located in the RING domain found in patients with macrocephaly were studied by the same assays. They showed enhanced Wnt/β-catenin signaling in a dominant negative manner by disrupting the ubiquitin ligase function and presumably compromising the Wnt receptor turnover. Taken together, we provide evidence for mirror brain size phenotypes caused by distinct mechanism of action on the Wnt/β-catenin signaling through protein domain-specific deleterious germline gene variants in ZNRF3.

Session Title: Mendelian Phenotypes Poster Session II

PB4720 Delineation of*AMMECR1*as a candidate gene for idiopathic short stature

Authors:

C. Thiel, A. Schulz, A. Fink, C. Kraus, S. Uebe, E. Arif, M. Krumbiegel, A. Reis; Inst. of Human Genetics, FAU Erlangen-Nürnberg, Erlangen, Germany

Short stature is a common medical concern in childhood and affects 3 % of the general population. It is thought to be controlled by a multitude of genetic variants with a broad range of effect sizes. We recently demonstrated that rare variants play significant role in the etiology of idiopathic short stature by confirming variants in known short stature associated genes in more than 16 % of these patients. In addition, we identified 63 candidate genes for short stature. In a study group of 605 patients with idiopathic short stature we identified an affected male with an *de novo* InDel-Variant in the X-linked *AMMECR1* gene, leading to a frameshift p.(Arg152Aspfs*11). The variant was detected in the patient's transcriptome by reverse transcriptase RT-PCR. Using quantitative real-time qPCR, we observed significantly decreased expression of the *AMMECR1* gene in lymphocytes of this individual. Immunofluorescence showed that the protein was expressed at a lower level in patient-derived lymphoblastoid cells (LCLs) and resided in the nucleus. These results implicate mRNA decay and loss-of function as the underlying disease mechanism.Disease causing variants in *AMMECR1* have been recently described in three independent reports to be associated with growth deficit, hearing impairment, elliptocytosis, and nephrocalcinosis. Even though, no additional symptoms beside short stature were observed in the case reported here, we provide supporting evidence for *AMMECR1* as a novel candidate gene for idiopathic short stature.

Session Title: Mendelian Phenotypes Poster Session III

PB4721 Description New Cases of *SOX17* Related Pulmonary Arterial Hypertension and Review of the Literature

Authors:

J. Nevado[1], N. Gallego-Zazo[1], L. Miranda-Alcaraz[2], A. Cruz-Utrilla[3], M. Del Cerro[4], M. Alvarez-Fuente[4], M. Rodriguez-Vazquez[5], I. Guillén[6], N. Ochoa[3], A. Parra[1], P. Pascual[1], M. Cazalla[1], C. Silvan[1], P. Arias[1], D. Valverde[7], V. De Jesús-Perez[8], P. Lapunzina[1], P. Escribano-Subias[3], J. Tenorio[9]; [1]INGEMM-HULP-CIBERER, Madrid, Spain, [2]INGEMM-HULP, Madrid, Spain, [3]Pulmonary Hypertension Unit, Dept. of Cardiology, Hosp. Univ.rio 12 de Octubre, 28041 Madrid, Spain. | Ctr. for Networked BioMed. Res. in Cardiovascular Diseases, Inst. de Salud Carlos III (CIBERCV), 28029 Madrid, Spain., Madrid, Spain, [4]Pulmonary Hypertension Unit, Hosp. Univ.rio Ramón y Cajal, 28034 Madrid, Spain, Madrid, Spain, [5]Pediatric Cardiology Unit, Hosp. Univ.rio Virgen de las Nieves, 18014 Granada, Spain, Granada, Spain, [6]Pediatric Cardiology Unit, Hosp. Univ.rio Virgen del Rocío, 41013 Sevilla, Spain., Seville, Spain, [7]CINBIO, Univ.e de Vigo, Vigo, Spain. | Rare Diseases and Pediatric Med., Galicia Sur Hlth.Res. Inst. (IIS Galicia Sur), SERGAS-UVIGO, Vigo, Spain., Vigo, Spain, [8]Div. of Pulmonary, Allergy & Critical Care Med., Stanford Univ. Sch. of Med., Stanford, CA, United States., Stanford, CA, [9]INGEMM-HULP, ADNTRO Genetics and BITGENETIC, Madrid, Spain

Pulmonary arterial hypertension (PAH) is an infrequent and progressive disorder characterized by high blood pressure in the pulmonary arteries. It may lead to premature death of the patient or the requirement for lung and/or heart transplantation. PAH can be classified into different groups, with the most common form being idiopathic PAH (IPAH). Other forms of PAH can be associated with primary conditions such as connective tissue diseases (PAH-CTD), congenital heart defects (PAH-CHD), or drug and toxin exposure (PAH-TX). Genetics plays an important role in this early diagnosis and targeted therapy of PAH. The increased use of NGS has led to an increased frequency of reporting new potential risk genes. In this work, we will focus on the relationship between *SOX17* variants and the development of PAH. SOX17 plays a critical role in embryonic development, is involved in the regulation of various developmental processes and plays a crucial role in the development of the cardiovascular system and the remodeling of blood vessels after birth. Thereby, dysregulation in *SOX17* expression or activity can contribute to endothelial dysfunction, which may contribute to the development or progression of PAH. Here, we report six additional patients with variants in *SOX17* and a review of previously described patients in the literature. In general, all patients described in this study suffered for other comorbidities including large septal defects as described by other groups. Collectively, 25 candidate variants were reported, including missense and predicted loss-of-function variants, in patients with H/IPAH. We add four new IPAH cases, two of them showing the same variant previously detected in multiple patients. In total, seven PAH patients have been detected as having the same variant, suggesting a possible hot spot. In addition, two of them were diagnosed with PAH-CHD and the rest as IPAH. Moreover, research conducted on a group of PAH patients with congenital heart disease indicated that *SOX17* variations are particularly prevalent in this subgroup, which experiences an early onset of the disease. Here we report a pediatric patient carrying a pathogenic variant not described previously and in which the same corkscrew tortuous arteries, confirming it relationship.The majority of nonsense and frameshift mutations found in PAH cases occur in the last exon, which encoded for the conserved β-catenin-binding domain. Meanwhile, the majority of missense variants are located in exon 1, in the HMG Box domain. Further research is still needed to clarify the precise mechanisms and extent of the association between SOX17 and the development of PAH

Session Title: Mendelian Phenotypes Poster Session I

PB4722 Detection of intronic variants as the genetic causes of inherited retinal degenerations.

Authors:

Y-S. Huang[1,2], J. Hsu[3], C-H. Yang[2,4], T-C. Chen[2,5], P-L. Chen[1,3,6]; [1]Graduate Inst. of Clinical Med., Coll. of Med., Natl. Taiwan Univ., Taipei City, Taiwan, [2]Dept. of Ophthalmology, Natl. Taiwan Univ. Hosp., Taipei City, Taiwan, [3]Graduate Inst. of Med. Genomics and Proteomics, Coll. of Med., Natl. Taiwan Univ., Taipei City, Taiwan, [4]Dept. of Ophthalmology, Coll. of Med., Natl. Taiwan Univ., Taipei City, Taiwan, [5]Ctr. of Frontier Med., Natl. Taiwan Univ. Hosp., Taipei City, Taiwan, [6]Dept. of Med. Genetics, Natl. Taiwan Univ. Hosp., Taipei City, Taiwan

Inherited retinal degenerations (IRDs) encompass a group of monogenic disorders that lead to the death of retinal cells and eventual vision loss in patients. There have been nearly 270 genes associated with IRDs. In a previous study, we conducted panel sequencing on 556 Taiwanese families and successfully diagnosed approximately 58% of the patients. However, more than forty percent of the patients remain undiagnosed. We hypothesized that disease-causing variants might reside within the intronic regions, which are typically overlooked during routine analysis. To predict variants that affect splicing, we utilized the spliceAI and dbscSNV_ADA tools. We excluded variants with a gnomAD East Asian allele frequency higher than 0.05 and those identified as benign or likely benign according to ACMG guidelines. The remaining variants were filtered based on the following criteria: a spliceAI score of > 0.2 or a dbscSNV_ADA score of > 0.6. As a result, we identified a total of 265 variants as splice-affected variants, with 90 of them located in introns. By evaluating the patients' family history and clinical symptoms, we further identified 18 potential disease-causing variants. Among these, 8 variants were found in deep intronic regions, and 10 variants were located in non-canonical splice sites. The variants *CNGB1*:c.1122-3344G>A, *USH2A*:c.5573-276G>A, *EYS*:c.5644+5G>A, and *TULP1*:c.499+5G>C were the most frequently observed intronic variants in our cohort. The use of splice prediction tools is crucial in the analysis of IRDs as it helps to narrow down the list of candidate variants, which can then be confirmed through minigene assays. Intronic variants should not be overlooked when definitive genetic etiology has not been identified, particularly in cases where only one variant is found in an autosomal recessive inheritance pattern. Furthermore, understanding the expected outcomes of intronic variants contributes to the development of potential therapies for IRDs.

Session Title: Mendelian Phenotypes Poster Session II

PB4723 Developing a morbidity gene panel for evaluation of kidney transplant recipients.

Authors:

B. Ma, N. T. Elefant, M. Tedesco, K. Bogyo, N. Vena, S. A. Bheda, S. K. Murthy, H. Milo Rasouly, A. G. Gharavi; Columbia Univ. Med. Ctr., New York, NY

Background: Significant resources are invested in kidney transplantation because it offers better survival than dialysis for patients with end stage kidney disease. However, morbidity and mortality in kidney transplantation remains high, mostly due to cardiovascular disease, infection and malignancy. Identification of monogenic conditions that contribute to these common complications can enable personalized management plans for kidney transplant recipients (KTRs) starting from the peri-operative period.
Methods: We developed a transplant morbidity panel of 355 genes associated with four common complications of transplantation: 1) cardiovascular diseases and risk factors such as hypertension, diabetes mellitus, dyslipidemia, 2) adult onset immunodeficiency, 3) post-transplant malignancy and hereditary cancer syndromes, 4) thrombophilia. We also incorporated the ACMG recommended genes for secondary findings. The gene list was evaluated in 1590 KTRs who had undergone whole exome sequencing. Diagnostic analysis was performed for genes in the panel per ACMG/AMP and ClinGen guidelines. Reverse phenotyping was performed to look for compatible clinical features and family history. Clinical implications of diagnostic variants were studied case-by-case.
Results: Among the 1590 KTRs, 257 individuals (16.2%) had a monogenic form of kidney disease. The transplant morbidity panel detected 58 different monogenic disorders in 146 individuals (9.2% diagnostic rate). 83 KTRs (5.2%) have diagnostic variants in cardiovascular disease category, 34 (2.1%) in malignancy category, 29 (1.8%) in immunodeficiency category and 3 (0.2%) in thrombophilia category. 8 patients (0.5%) have dual genetic diagnoses. 74 patients (50.7%) had supporting clinical features or family history of the monogenic disorders, while the rest have either absence or no documentation of any supporting features. Among KTRs with diagnostic variants, identification of these monogenic disorders and risk factors would allow physicians to set specific risk factor targets in 11 individuals (7.5%), arrange intensive surveillance in 142 (97.3%), utilize preventive measures in 19 (13.0%), guide disease-specific therapy in 94 (64.4%), initiate specialty referral in 132 (90.4%) and alter immunosuppressive regimens in 84 (57.5%).
Conclusion: Clinically important monogenic disorders, unrelated to the primary cause of kidney disease, were identified in 9.2% of KTRs, with multiple implications for clinical management. Incorporating genetic diagnostics into the transplant evaluation protocol would enable personalized management plans and help reduce complications.

Session Title: Mendelian Phenotypes Poster Session III

PB4724 Development and characterization of a novel TAFAZZIN-deficient iPSC derived skeletal muscle model of Barth Syndrome: a new tool to study muscle phenotypes in BTHS.

Authors:

K. Lee, O. Sniezek, H. Vernon; Johns Hopkins Sch. of Med., Baltimore, MD

Barth syndrome (BTHS) is a rare, X-linked disorder caused by pathogenic variants in the gene *TAFFAZIN* (TAZ), which results in abnormal remodeling of cardiolipin (CL) in mitochondria and a clinical phenotype including cardiomyopathy, skeletal myopathy, and neutropenia. Cardiolipin is responsible for key mitochondrial functions, and the failure to fulfill these roles properly leads to not only mitochondrial dysfunction, but also to broader cellular dysfunction. We generated wild type and TAZ knockout (TAZ-KO) cardiomyocytes (CM) and neural progenitor cells (NPCs) from induced pluripotent stem cells (iPSCs) to compare the mitochondrial and cellular pathology in cell types affected and unaffected, respectively, in BTHS. Interestingly, we observed abnormal cell maturation in TAZ-KO CMs but not NPCs, and dysregulated WNT signaling in the TAZ-KO undifferentiated iPSCs. As BTHS patients often have left ventricular noncompaction, a developmental phenotype, we hypothesized that differentiation defects in cardiomyocytes may underlie the early cardiac phenotype. We further hypothesized that skeletal myopathy in BTHS may also be due to a developmental muscle defect. To characterize the muscle mitochondrial and developmental phenotype in BTHS, we developed WT and TAZ-KO iPSC-derived skeletal muscle cells (SKM). Both WT and TAZ-KO SKMs achieved the appropriate intermediate and terminal stages of differentiation: mesoderm, SKM precursors, myoblasts, myotubes, and mature/fused myotubes. We validated these stages with qRT PCR markers appropriate for each stage and performed immunofluorescence staining with the antibodies for terminally differentiated myoblasts and myotubes. Since CMs and SKMs have similar initial stages of differentiation, and then diverge, we are currently detailing at which stage TAZ-deficiency has the largest impact. By investigating both muscle cell types simultaneously, we will ascertain how TAZ-KO CM and SKM cells differ in their differentiation timeline, progression, and efficacy of reaching each stage of differentiation and gain an understanding of how TAZ-deficiency contributes to the developmental tissue pathology in BTHS.

Session Title: Mendelian Phenotypes Poster Session I

PB4725 Diagnostic whole exome sequencing in presumably autosomal recessively inherited retinal dystrophies in an Iranian population

Authors:

L. Hoefsloot[1], P. Heutinck[2,3], M. van Tienhoven[1], S. Kheradmand Kia[4], C. Klaver[5,3,6,7], D. D. Farhud[8,9], M. Meester-Smoor[2,3], A. A. H. Thiadens[2], V. J. M. Verhoeven[1,2]; [1]Dept Clinical Genetics, Erasmus MC, Rotterdam, Netherlands, [2]Dept Opththalmology, Erasmus MC, Rotterdam, Netherlands, [3]Dept. of Epidemiology, Erasmus MC, Rotterdam, Netherlands, [4]. Lab. for Red Blood Cell Diagnostics, Sanquin, Amsterdam, Netherlands, [5]Dept Opththalmology, Erasmus MC, R0tterdam, Netherlands, [6]Dept Ophthalmology, Radboud UMC, Nijmegen, Netherlands, [7]Inst. of Molecular and Clinical Ophthalmology, Basel, Switzerland, [8]Farhud Genetics Clinic, Tehran, Iran, Islamic Republic of, [9]Sch. of Publ. Hlth., Tehran Univ. of Med. Sci.,, Tehran, Iran, Islamic Republic of

Purpose: Genetic testing in inherited retinal dystrophies (IRD) allows for accurate identification of inheritance pattern, thereby improving genetic counselling and potential (gene) therapy options for affected individuals and their families. IRDs can show various inheritance patters, e.g. autosomal dominant, autosomal-recessive, and X-linked inheritance. In this study, we aimed to uncover the genetic cause of presumably autosomal recessively inherited IRDs in a cohort of patients from Iranian descent. **Methods:** IRD patients were recruited by ophthalmologists of the Tehran University of Medical Sciences. Whole exome sequencing (WES) was performed using a gene panel including ~500 genes involved in inherited eye disorders. Variants were classified according the ACMG guidelines using Alamut Visual, the CADD scoring website and the online Franklin tool. Unsolved cases are currently being taken forward for analyses of relevant homozygous or other variants outside of the gene panel. **Results:** 111 patients (58 females, 53 males) with inherited retinal dystrophies [VV1] were recruited and successfully underwent WES. Mean age at recruitment was 37.2 years old (+/- 11.2 yrs SDS; age range 15-68). ⋯% of the cohort originated from a consanguineous background, mostly due to first cousin marriages. [VV2] In 59% of the cases, we found a causative (homozygous) variant; in another 31% of the cases, we found a potentially causative (homozygous) variant. The most frequently involved genes were CERKL (15%), EYS (11%), and RPE65 (9%). **Conclusion:** Using WES with a filter for known IRD genes as a first-tier genetic test in this consanguineous IRD Iranian cohort, we were able to identify a (potential) causative (homozygous) variant in almost 90% of the cases. Expanding the WES analyses outside of the panel will probably even increase the diagnostic yield.

Session Title: Mendelian Phenotypes Poster Session II

PB4726 Differences in cardiac mechanics among genetically at-risk first-degree relatives of probands with dilated cardiomyopathy indicate that variants of uncertain significance are clinically relevant: The DCM Precision Medicine Study.

Authors:

D. D. Kinnamon[1], J. E. Wilcox[2], L. Beussink-Nelson[2], J. Cao[1], R. Kumar[2], E. Jordan[1], H. Ni[1], S. J. Shah[2], R. E. Hershberger[1]; [1]The Ohio State Univ., Columbus, OH, [2]Northwestern Univ. Feinberg Sch. of Med., Chicago, IL

Among genetically at-risk first-degree relatives (FDRs) of probands with dilated cardiomyopathy (DCM), the ability to detect early changes in left ventricular (LV) mechanics with normal LV size and ejection fraction (LVEF) remains incompletely explored, particularly among FDRs harboring only their proband's variants of uncertain significance (VUSs). We sought to define a pre-DCM phenotype among at-risk FDRs, including those with only VUSs, using LV global longitudinal strain (GLS), an early measure of abnormal cardiac mechanics. LV structure and function, including speckle-tracking analysis for LV GLS, were evaluated in 124 FDRs (65% female; median age 44.9 [IQR: 30.6-60.3] years) of 66 DCM probands of European ancestry sequenced for rare variants in 35 DCM genes. All FDRs had normal LV size and LVEF. Probands' variants were adjudicated using published American College of Medical Genetics (ACMG) and ClinGen-based criteria tailored to DCM and assigned to an ACMG category. Pathogenic (P), likely pathogenic (LP), and VUS variants were confirmed in the proband and cascade-tested in FDRs by Sanger sequencing. Negative FDRs of probands with P/LP variants (n=28) were a reference group to which negative FDRs of probands without P/LP variants (n=30), FDRs with only VUSs (n=27), and FDRs with P/LP variants (n=39) were compared. In an analysis accounting for sex, height, weight, image quality, and the age-dependent penetrance of DCM, FDRs below the sample median age showed minimal differences in LV GLS across groups, consistent with reduced penetrance at younger ages. However, those above the sample median age with P/LP variants or only VUSs had lower absolute LV GLS values than the reference group (-3.9 [95% CI: -5.7, -2.1] or -3.1 [-4.8, -1.4] %-units) and negative FDRs of probands without P/LP variants (-2.6 [-4.0, -1.2] or -1.8 [-3.1, -0.6]). As lower absolute LV GLS values are indicative of worse cardiac mechanics, these findings indicate that both P/LP and some DCM-related VUSs are clinically relevant. In particular, the finding that FDRs harboring only VUSs had lower absolute LV GLS values than another group also considered to have uncertain genetic risk, negative FDRs of probands without P/LP variants, suggests that FDRs currently classified as having uncertain genetic risk could be further stratified by whether they harbor a DCM-relevant VUS. Taken together, these findings suggest that LV GLS may have utility for defining a pre-DCM phenotype in FDRs across the genetic risk spectrum.

Session Title: Mendelian Phenotypes Poster Session III

PB4727 Different variants of cardiac remodeling and genetic variant spectrum in patients with left ventricular noncompaction.

Authors:

M. Polyak[1], S. L. Dzemeshkevich[1], O. V. Blagova[2], Y. A. Lutokhina[2], E. V. Zaklyazminskaya[1]; [1]Petrovsky Natl. Res. Ctr. of Surgery, Moscow, Russian Federation, [2]Sechenov First Moscow State Med. Univ., Moscow, Russian Federation

Abstract. Left ventricular noncompaction (LVNC) is a well-described phenomenon of two-layer myocardium. The genetic findings in LVNC patients overlap with cardiomyopathies, but genetic landscape remains debatable. It is unclear if LVNC represents a distinct entity or is a common sign or stage of cardiomyopathy. We deeply investigated cardiac remodeling in LVNC patients and the diagnostics yield of screening of the 13 genes with the most proven causative role in cardiomyopathies. Methods. We performed clinical and genetic testing for 67 probands diagnosed with LVNC according to Jenni criteria (2001). Custom-designed target genes panel (*MYH7*, *MYBPC3*, *TPM1*, *TNNT2*, *TNNI3*, *TAZ*, *MYL2*, *MYL3*, *ACTC1*, *LDB3*, *LMNA*, *DTNA*, *DES*) was sequenced by IonTorrent PGM. For statistical analysis, we used the SciPy and StatsModels software; for all tests, the differences were considered significant at $p<0.05$.Results. Complete pedigrees were available in 44 families. Sporadic origin of LVNC was predominant (26/44) in probands with reliable family history. Isolated LVNC was observed in 15 out of 67 probands (22%). In 78% of probands, the LVNC was accompanied by another cardiomyopathy. The most common variant of cardiac remodeling was dilated cardiomyopathy (35 out of 67 probands). Hypertrophy was found in 15 probands (22%), and 10 out of these 15 probands had also chambers dilatation. This subgroup of patients might represent dilation phase as an advanced stage of HCM. Two LVNC probands met the Marcus criteria (2010) for ARVC. Thirteen pathogenic (P) and likely pathogenic (LP) variants were found in 15 probands; none of them was detected in patients with an isolated LVNC. Diagnostic yield was the highest in patients with LVNC and cardiac hypertrophy. Patients carrying P/LP variants had an earlier age of manifestation and a higher rate of SCD in the family ($p < 0.05$). We identified 20 genetic variants of uncertain significance (VUS) in 17 unrelated probands. Most of the variants were detected in *MYH7* and *MYPBC3* genes. Other findings included variants in *TNNT2*, *LMNA*, and *DES*. Remarkably, non-carriers of P/LP variants had a higher ratio of non-compact to compact myocardial layers ($p < 0.05$).Conclusion. Diagnostic yield for the 13 genes panel was 22% in the whole group, and 33% in patients with LVNC and HCM]. Young patients with a combination of LVNC and cardiac hypertrophy had the highest benefit from the testing. Genotype-positive patients had earlier manifestation and higher risk of SCD. We hypothesize that lone LVNC phenomena might represent developmental failure, and the spectrum of causative genes differs from primary cardiomyopathies and might be close to one for congenital heart defects.

Session Title: Mendelian Phenotypes Poster Session I

PB4728 † Differential alternative splicing analysis links variation in ZRSR2 to a novel oral-facial-digital syndrome

Authors:

L. Hannes[1], M. Atzori[1], A. Goldenberg[2], J. Argente[3], T. Attie-Bitach[4], C. Attanasio[1], A-L. Bruel[5], M. Castanet[2], C. Dubourg[6], a. jacobs[7], S. Lyonnet[8], N. pezzella[9], E. Pelgrims[10], M. Aerden[11], M. Bauters[12], A. Rochtus[11], A. Swillen[11], A. Sifrim[1], R. Tammaro[13], F. Tran Mau-Them[14], S. Odent[15], C. Thauvin-Robinet[16], B. Franco[17], J. Breckpot[1]; [1]KU Leuven, Leuven, Belgium, [2]CHU Rouen, Rouen, France, [3]Hosp. Infantil Univ.rio Niño Jesús, Madrid, Spain, [4]Hôpital Necker, Université Paris Descartes, Paris, France, [5]Université de Bourgogne, Dijon, France, [6]Rennes Univ. Hosp., Rennes, France, [7]Univ. Hosp. Leuven, Leuven, Belgium, [8]Imagine Inst. Univ. Paris Descartes, Paris, France, [9]TIGEM, Naples, Italy, [10]kuleuven, leuven, Belgium, [11]UZ Leuven, Leuven, Belgium, [12]UZLeuven, Leuven, Belgium, [13]TIGEM, naples, Italy, [14]CHU Dijon Bourgogne, Dijon, France, [15]CHU de Rennes Hosp SUD, Rennes Cdx 2, France, [16]FHU-TRANSLAD, Dijon, Dijon, France, [17]Scuola Superiore Meridionale, Naples, Italy

Orofaciodigital (OFD) syndromes represent a group of rare genetically heterogeneous developmental disorders. OFDs are due to pathogenic variants in genes that are involved in primary cilia formation and/or function. We identified the same germline variant c.1207_1208del (p.Arg403Glyfs*24) in the last exon of ZRSR2, either occurring de novo or inherited in an X-linked recessive pattern, in 6 affected males from 4 unrelated families with OFD syndrome in association with structural brain abnormalities, ranging from alobar holoprosencephaly to pituitary abnormalities. ZRSR2, located on the X chromosome, is part of the minor spliceosome complex which recognizes minor (U12-type) introns, representing 0.35% of human introns. Although somatic pathogenic variants in ZRSR2 were associated with the development of myelodysplastic syndrome (MDS), this is the first association of germline variation in ZRSR2 with a human developmental disorder. Alternative splicing analysis of minor spliceosome gene targets in lymphoblastoid and fibroblast cell lines from an affected patient harboring the ZRSR2 variant and from unrelated controls, showed significant enrichment of minor intron retention. Among the differentially spliced targets are ciliopathy-related genes, such as TMEM107 and CIBAR1. Primary fibroblasts containing this ZRSR2 variant had abnormally elongated cilia, confirming an association between defective U12-type intron splicing and abnormal primary cilia morphogenesis.

Session Title: Mendelian Phenotypes Poster Session II

PB4729 Disease causing variants in several genes causing Bardet Biedl Syndrome

Authors:

F. RATHER[1], M. Umair[2]; [1]KING SAUD BIN ABDUL AZIZ Univ. FRO HEALTH Sci.S, RIYADH, Saudi Arabia, [2]KAIMRC, Riyadh, Saudi Arabia

Background Bardet Biedl syndrome is a rare heterogeneous autosomal recessive multi systemic condition with twenty two identified genes. Six distinct clinical and diagnostic features include rod cone dystrophy, learning difficulties, renal abnormalities, male hypogonadism, postaxial polydactyly, and obesity. Methods Here, we provide data from nine consanguineous families and one nonconsanguineous family, in which multiple members were affected and displayed characteristic BBS clinical symptoms. Ten BBS Pakistani families were involved in the current investigation, and whole exome sequencing identified novel and recurrent gene variations. Results WES revealed four novel biallelic nonsense variants in four families, *IFT27, BBIP1, WDPCP, LZTFL1* genes, likewise six previously reported variants in five different genes *BBS5, BBS1, BBS1, MKKS, BBS5.* Conclusion Our findings confirm the significance of these genes in the emergence of multi systemic human genetic illness and broaden the mutational and phenotypic spectrum of four distinct forms of ciliopathies that cause BBS.

Session Title: Mendelian Phenotypes Poster Session III

PB4730 DNA Repair-related Heritable Photosensitivity Syndromes: Mutation Landscape in a Multiethnic Cohort of 17 Multigenerational Families with High Degree of Consanguinity

Authors:

A. Hozhabrpour[1], M. Mojbafan[1], S. Talebi[1], F. vahidnezhad[2], F. Palizban[3], M. Amani[4], M. Garshasbi[5], A. Naghavi[6], R. Khalesi[7], P. Mansouri[8], S. Sotoudeh[9], H. Mahmoudi[10], A. Varghaei[11], M. Daneshpazhooh[12], F. Karimi[13], S. Zeinali[14], D. Ebrahimi[15], J. Uitto[16], L. Youssefian[17], H. Vahidnezhad[18]; [1]Dept. of Med. Genetics and Molecular Biology, Iran Univ. of Med. Sci., Tehran, Iran, Islamic Republic of, [2]UCSC Silicon Valley Extension, Univ. of California, Santa Cruz, United States of America, california, CA, [3]bInst. of Biochemistry and Biophysics, Univ. of Tehran, Tehran, Iran., Tehran, Iran, Islamic Republic of, [4]Clinical Res. Dev. Unit of Allameh Bohlool Gonabadi Hosp., Gonabad Univ. of Med. Sci., Gonabad, Iran., Gonbad, Iran, Islamic Republic of, [5]Dept. of Med. Genetics, Faculty of Med. Sci., Tarbiat Modares Univ., Tehran, Iran., Tehran, Iran, Islamic Republic of, [6]Cellular and Molecular Res. Ctr., Zahedan Univ. of Med. Sci., Zahedan, Iran., Zahedan, Iran, Islamic Republic of, [7]Dept. of Med. Genetics, DeNA Lab., Tehran, Iran, Tehran, Iran, Islamic Republic of, [8]Skin and Stem Cell Res. Ctr., Tehran Univ. of Med. Sci., Tehran, Iran, Tehran, Iran, Islamic Republic of, [9]Dept. of Dermatology, Children's Med. Ctr., Ctr. of Excellence, Tehran Univ. of Med. Sci., Tehran, Iran., Tehran, Iran, Islamic Republic of, [10]Dept. of Dermatology, Razi Hosp., Tehran Univ. of Med. Sci., Tehran, Iran., Tehran, Iran, Islamic Republic of, [11]Dept. of Dermatology, Sch. of Med., Urmia Univ. of Med. Sci., Urmia, Iran., Urmia, Iran, Islamic Republic of, [12]Dept. of Dermatology, Autoimmune Bullous Diseases Res. Ctr., Tehran Univ. of Med. Sci., Tehran, Iran, Tehran, Iran, Islamic Republic of, [13]Shiraz Univ. of Med. Sci., Shiraz, Iran, Islamic Republic of, [14]Kawsar Human Genetics Res. Ctr., Tehran, Iran., Tehran, Iran, Islamic Republic of, [15]Texas BioMed. Res. Inst., San Antonio, Texas, USA, Texas, TX, [16]Dept. of Dermatology and Cutaneous Biology, Sidney Kimmel Med. Coll., Thomas Jefferson Univ., Philadelphia, Pennsylvania, USA., Pennsylvania, PA, [17]Dept. of Pathology and Lab. Med., UCLA Clinical Genomics Ctr., Univ. of California, Los Angeles, David Geffen Sch. of Med., Los Angeles, California, USA, California, CA, [18]Dept. of Pediatrics, The Univ. of Pennsylvania Sch. of Med., Philadelphia, PA 19104-4318, USA., Pennsylvania, PA

Inherited photosensitivity syndromes are a heterogeneous group of genetic skin disorders with tremendous phenotypic variability, characterized by photosensitivity and defective DNA repair, especially nucleotide excision repair. A cohort of 17 Iranian families with heritable photosensitivity syndromes was evaluated to identify their genetic defect. The patients' DNA was analyzed with whole-exome sequencing. The interpretations of the genomic results were guided by genome-wide homozygosity mapping. Haplotype analysis was performed for cases with recurrent mutations. RNA sequencing (RNA-Seq) was also utilized to determine germline and somatic variants for exploring the pathogenicity and cancer mutational signature analysis, respectively. Thirteen sequence variants, including six previously unreported pathogenic variants, were disclosed in 17 Iranian cohort, with *XPC* as the most common mutated gene in 10 families (59%). In one patient, RNA-Seq, as a first-tier diagnostic approach, revealed a non-canonical homozygous germline variant: *XPC*:c.413-9T>A. The Sashimi plot showed skipping of exon 4 with dramatic XPC down-expression. RNA-Seq-based somatic variant analysis including the somatic variant annotation and mutation signature profiling disclosed driver mutations in cancer-related genes such as *PTCH1*. This mutational signature was deconvoluted into component signatures of 1, 5, 12, and 20 of the COSMIC signatures. Haplotype analysis of *XPC*:c.2251-1G>C and *XPC*:1243C>T in four families showed common haplotypes of 1.7 Mb and 2.6 Mb, respectively, denoting a founder effect. Lastly, two extremely rare cases were presented in this report: a homozygous *UVSSA*:c.1990C>T was disclosed in a family in which homozygous and heterozygous patients had severe and minimal presentations, respectively. A family with *ERCC2*-related COFS with an early childhood death was reported. A direct comparison of our data with the results of previously reported cohorts demonstrates the international mutation landscape of DNA repair-related photosensitivity disorders; although, population-specific differences were observed.

Session Title: Mendelian Phenotypes Poster Session I

PB4731 Dominant *OGDH* mutations cause peripheral neuropathy with ataxia and optical atrophy

Authors:

L. Van de Vondel[1], J. De Winter[1], A. Sandoval[2], J. F. Ramirez[2], R. Horvath[3], J. De Bleecker[4], S. Zuchner[5], W. Yoon[2], J. Baets[6]; [1]Univ. of Antwerp, Wilrijk, Antwerpen, Belgium, [2]Oklahoma Med. Res. Fndn., Oklahoma City, OK, [3]Univ. of Cambridge, Cambridge, United Kingdom, [4]Univ. of Ghent, Ghent, Belgium, [5]Univ. of Miami Miller Sch. of Med., Miami, FL, [6]Univ. of Antwerp, Wilrijk, Antwerp, Belgium

2-oxyglutrate dehydrogenase (OGDH) is an E1 component of α-ketoglutarate dehydrogenase complex (α-KGDH) that plays a crucial role in the Krebs cycle metabolism. Biallelic variants in *OGDH* have been reported to cause OGDH deficiency (OGDHD; OMIM: # 203740), a condition that leads to an early-onset neurodevelopmental and mitochondrial disorder. However, whether monoallelic *OGDH* variants could also exhibit dominant effects in humans had not been known. In this study, we identified a *de novo OGDH* c. 1909C>T (p.Arg637Trp) variant by whole-exome sequencing (WES) in a patient manifesting late-onset neurological phenotypes, including peripheral neuropathy, cerebellar ataxia and bilateral optic atrophy. Blood analysis revealed ketogenosis. In patient lymphoblasts, *OGDH* levels do not seem altered in the whole cell lysate nor specifically in the mitochondrial fraction compared to familial controls. In contrast, these cells displayed decreased α-KGDH activity. These findings strongly suggest that the p.Arg673Trp variant acts as a dominant-negative mutation. To test this hypothesis, we generated *Drosophila* model carrying *UAS-dOgdh (p.Arg639Trp)* transgene, harboring the homologous mutation to the human variant. Using this transgenic fly, we will determine whether neuronal expression of the mutant dOgdh (p.Arg639Trp) leads to developmental and degenerative defects in neurons, as well as impairment of Krebs cycle metabolism.Additionally, through genetic matchmaking, we identified two additional families with patients harboring heterozygous *OGDH*variants, all presenting late-onset neurological phenotypes. Collectively, our data suggest that the monoallelic *OGDH* variants disrupt α-KGDH function, thereby leading to late-onset neurological diseases in humans.

Session Title: Mendelian Phenotypes Poster Session II

PB4732 *Drosophila* model of *de novo PHACTR1* variant highlights the role of exon 5 duplication in novel neurodevelopmental disorder

Authors:

J. Andrews[1,2], S. Jangam[1,2], L. Briere[3], R. Townsley[1,2], Undiagnosed Disease Network, S. Yamamoto[1,2], D. Sweetser[3], M. Wangler[1,2]; [1]Baylor Coll. of Med., Houston, TX, [2]Jan and Dan Duncan Neurological Res. Inst., Houston, TX, [3]Massachusetts Gen Hosp, Boston, MA

Phosphatase and Actin Regulator 1 (PHACTR1) is a member of a family of Protein Phosphatase 1 (PP1) binding proteins which defines the substrate specificity of the Phactr1/PP1 holoenzyme. PHACTR1 is broadly expressed within the central nervous system and is believed to play an important role in the regulation of actin cytoskeletal dynamics, actin stress fiber formation, cell motility, and cell survival. Multiple *PHACTR1* variants have been previously associated with different pathologies, including developmental and epileptic encephalopathy, West syndrome, myocardial infarction, melanoma growth, coronary artery disease, and cervical artery dissection. Here we report a proband with a *de novo* duplication of exon 5 and parts of the surrounding introns who presented with global developmental delay, hypotonia, absent speech, seizures, behavioral problems, mild prominence of ventricles and sulci, and slight dysmorphism of the left hippocampus. We have generated a *Drosophila* model utilizing the UAS/GAL4 system to drive the expression of *PHACTR1* human cDNA within the fly. Overexpression of either the reference or variant human cDNA via a Nubbin-Gal4 driver produced no changes within the wing tissues, and overexpression within the eye by GMR-Gal4 produced only minor changes in eye morphology. However, the ubiquitous overexpression of the putatively pathogenic variant using either an Actin or Tubulin-Gal4 driver was sufficient to induce forking or abnormal bends in bristles within the head, notum, and sternopleurum. These changes within the actin-rich bristle structures were not observed in animals overexpressing the reference human cDNA. Likewise, the ubiquitous overexpression of *CG32264*, the *Drosophila* ortholog to *PHACTR1*, showed no alterations in bristle morphology. Conversely, ubiquitous depletion of *CG32264* via two different RNAi lines ($CG32264^{JF02975}$ and $CG32264^{HMC05840}$) failed to cause any changes in bristle morphology. However, the expression of $CG32264^{JF02975}$, but not $CG32264^{HMC05840}$, in eye tissues via GMR-Gal4 produced a rough-eye phenotype. Our initial assessment demonstrates that *Drosophila* is a suitable model for further exploration of *PHACTR1* variants and identifies that duplications of the 5th exon are putatively deleterious and may underly a novel disorder separate from those previously associated with *PHACTR1*.

Session Title: Mendelian Phenotypes Poster Session III

PB4733 Dysfunction of the centrosomal protein, CEP76, is associated with syndromic ciliopathies.

Authors:

K. Khan[1,2], K. Bishara[2], A. Ozanturk[3], S. Frangakis[3], F. Ullah[2], T. Arbogast[2], M. A. Al Khateeb[4], L. Damaj[5], B. Bocquet[6], I. Meunier[6], T. Bartolomaeus[7], L. Qebibo[8], F. Ramond[9], M. Rydzanicz[10], R. A. Jamra[7], L. Burglen[11,8], E. Davis[2,12]; [1]Northwestern Univ., Chicago, IL, [2]Stanley Manne Children's Res. Inst., Ann & Robert H. Lurie Children's Hosp. of Chicago, Chicago, IL, [3]Ctr. for Human Disease Modeling, Duke Univ. Med. Ctr., Durham, NC, [4]Women's Wellness Res. Ctr. and Neonatal Intensive Care Unit, Hamad Med. Corp.; and Weill Cornell Med., Doha, Qatar, [5]Dept. of Pediatrics, Competence Ctr. of Inherited Metabolic Disorders, Rennes Hosp., Rennes, France, [6]Inst. for NeuroSci.s of Montpellier (INM), Univ. Montpellier, INSERM, Montpellier, France, [7]Inst. of Human Genetics, Univ. of Leipzig Med. Ctr., Leipzig, Germany, [8]Ctr. de Référence Malformations et Maladies Congénitales du Cervelet, Département de génétique, AP-HP. Sorbonne Université, Hôpital Trousseau, Paris, France, [9]Dept. of Genetics, Ctr. Hosp.ier Univ.ire de Saint-Étienne, Saint-etienne, France, [10]Dept. of Med. Genetics, Med. Univ. of Warsaw, Warsaw, Poland, [11]Dev.al Brain Disorders Lab., Imagine Inst., INSERM UMR 1163, Paris, France, [12]Dept. of Pediatrics and Dept. of Cell and Dev.al Biology, Feinberg Sch. of Med., Northwestern Univ., Chicago, IL

Widespread use of clinical exome and genome sequencing in pediatric cohorts with suspected ciliopathies has led to an accelerated improvement of diagnostic rate. Yet, a subset of individuals who display hallmark ciliopathy features harbor hitherto unknown causal loci offering an opportunity for disease gene discovery. Here, we report the identification of *CEP76*, encoding centrosomal protein 76, as a causal ciliopathy locus. Data sharing platforms connected five independent groups who had performed either targeted sequencing or proband-parent trio whole exome/genome sequencing on unrelated pedigrees with suspected genetic disorders. Five unrelated affected children with partially overlapping phenotypes harbor rare biallelic variants in *CEP76* that segregate with disease and are predicted to impact protein function. One male proband with biallelic nonsense variants was diagnosed with Bardet-Biedl syndrome; a second affected male with an mRNA splice altering variant in trans with a rare missense change fulfilled diagnostic criteria of Joubert syndrome; a third male with a rare homozygous missense change displayed syndromic intellectual disability; a fourth affected female harboring biallelic missense variants exhibited Joubert syndrome-like features; and a fifth affected female carrying homozygous missense variant presented with retinitis pigmentosa. CEP76 has been shown previously to associate with CP110, a CDK2 substrate, and CEP97, a calcium-binding protein, which localize to the basal body and daughter centriole of primary cilia. Importantly, our zebrafish stable mutants with a homozygous intragenic deletion of *cep76* display photoreceptor deficits as evidenced by significantly altered response to light-dark stimuli and histological phenotypes including rhodopsin mislocalization and significantly reduced ciliary length compared to controls. Furthermore, RNA silencing of endogenous *CEP76* and/or chemical perturbation induced centriolar proliferation phenotypes *in vitro* in U2OS cell lines. These phenotypes were rescued by introduction of wild type *CEP76*, however, introduction of a subset of patient variants did not rescue the aberrant centriolar phenotypes. Together, our human genetic and functional data provide evidence to support causality of *CEP76* in a clinically heterogenous human disorder. Our ongoing studies aim to explain phenotypic differences among affected individuals and pursue hypotheses of differing *CEP76* variant strength and physical interactions with other centrosomal and ciliary loci.

Session Title: Mendelian Phenotypes Poster Session I

PB4734 Dysmyelination or demyelination: Investigating the role of *SLC17A5* in myelination in a murine model of Free Sialic Acid Storage Disorder.

Authors:

M. Hossain[1], M. Sabir[1,2], M. Hackbarth[3,1], P. Leoyklang[1], H. Dorward[1], M. Huizing[1], W. Gahl[1], M. C. V. Malicdan[1]; [1]NIH, Bethesda, MD, [2]Univ. of Oxford, Oxford, United Kingdom, [3]Univ. of Maryland Sch. of Med., Baltimore, MD

Background: Free Sialic Acid Storage Disorder (FSASD) is a rare lysosomal storage disorder resulting from bi-allelic variants in *SLC17A5*. It is characterized by a multisystemic phenotype, including progressive neurodegeneration and significant hypomyelination. Although SLC17A5, a lysosomal membrane protein, is known to transport sialic acid from lysosomes to the cytosol, the pathophysiology of myelination defects and neurodegeneration in FSASD and the relationship between lysosomal SLC17A5 function and myelin production, composition, and maintenance remain enigmatic. To investigate SLC17A5 function, we generated an $Slc17a5^{R39C}$ knock-in mouse model with the prevalent p.Arg39Cys (R39C) mutation. These mice exhibit features of FSASD, including progressive ataxia, neurodegeneration, and hypomyelination.
Methods: To understand the extent by which SLC17A5 dysfunction leads to myelination problems and neurodegeneration, we further characterized $Slc17a5^{R39C}$ mice. Using histopathological techniques, we examined expression of proteins related to myelin or CNS regulation in the cortex, cerebellum, and cervical spinal cord of $Slc17a5^{R39C}$ mice and control littermates at various stages of postnatal development. Immunofluorescence staining was performed using antibodies specific for myelin basic protein (MBP), Purkinje cells (PCP4) and astrocytes (GFAP).
Results: $Slc17a5^{R39C}$ mice had decreased MBP expression, significant white matter astrogliosis, and decreased PCP4 expression over time, indicating progressive Purkinje cell loss throughout development. Luxol Fast Blue staining, which highlights myelinated regions, revealed prominent hypomyelination in the corpus callosum, cerebellar white matter, and all major spinal white matter tracts as early as P14 in $Slc17a5^{R39C}$ mice. Surprisingly, Nissl staining did not show differences in neuronal body numbers, suggesting myelin density changes are independent of neuronal loss.
Conclusion: Our findings suggest dysmyelination or hypomyelination has occurred as early as P14, within the biological window for myelination, resulting in myelination deficits in the $Slc17a5^{R39C}$ mouse model. Progressive neurodegeneration occurs alongside myelination abnormalities. These results align with clinical presentation in FSASD patients and provide insights into the pathophysiology of FSASD and the impact of SLC17A5 on myelin production, composition, and maintenance. The findings could have implications for the development of personalized therapies for patients with FSASD or other disorders characterized by myelin defects or neurodegeneration.

Session Title: Mendelian Phenotypes Poster Session II

PB4735 E-cadherin gene deletion and complete anodontia: A case report.

Authors:

S. Beiraghi[1], O. Klein[2], C. Billington, Jr.[1]; [1]Univ Minnesota, Minneapolis, MN, [2]Cedars-Sinai, Los Angeles, CA

The CDH1 gene codes for the Cadherin-1, or E-Cadherin, protein which is important for the epithelial tissue growth and function. In-frame gene deletions of *CDH1* have been linked to tooth anomalies such as complete anodontia. This case report presents the first incidence of a female toddler with in-frame deletion of the *CDH1* gene and complete anodontia. Dental x-rays revealed no primary teeth and no evidence of tooth buds. Exome sequencing was performed and it revealed heterozygous change in *CDH1*. This patient has an in-frame deletion of Asn458 (NM_004360 c.1371_1371del). It is *de novo* and rare in the population. This case report reveals that the *CDH1* gene might impact tooth development and the first case of a patient with complete anodontia and *CDH1* gene deletion.

Session Title: Mendelian Phenotypes Poster Session III

PB4736 Effect of Polymorphisms in *BCL11A, APOL1* genes and Expression of Foetal Haemoglobin on Sickle Cell Disease Phenotypes

Authors:

M-L. Sakyi[1,2,3], N. A. Karikari Agyemang[4,2], M. E. Gakpey[4,2,3], D. Adu[2,3], L. N. Amenga-Etego[5]; [1]Univ. of Ghana, Accra, Ghana, [2]Human Hereditary Hlth.Africa Kidney Disease Res. Network, Genomics Lab., Accra, Ghana, [3]West Africa Genetic Med. Ctr., Accra, Ghana, [4]Noguchi Mem. Inst. For Med. Res., Accra, Ghana, [5]West Africa Ctr. for Cell Biology of Infectious Pathogens, Accra, Ghana

Renal complication is a leading cause of early mortality in sickle cell disease (SCD). Foetal haemoglobin (HbF), a major genetic modulator in SCD ameliorates related complications. Mutations in BCL11A gene also induce increased expression of Hb F transcription. However, APOL1 genetic variants (G1 and G2) are linked with development of chronic kidney disease (CKD). This study sought to understand the interplay of APOL1 mutations in SCD patients with CKD. Cross-sectional random sampling was used to recruit 384 SCD patients visiting Komfo Anokye and Korle Bu Teaching Hospitals in Ghana after ethical approval. Hematological and biochemical tests were conducted to determine SCD phenotypes followed by haemoglobin fractionation. Genomic DNA was isolated from whole blood followed by PCR-Ligase detection reaction and TaqMan SNP genotyping assay to genotype APOL1 and BCL11A SNPs. Pearson's correlation was used to test for correlation of two variables (r). Chi Square test and logistic regression was used to test for association of genotype and phenotypes. In all participants, age had a negative correlation (p=0.014) with HbF. Age was significantly (p=0.014) associated with CKD. No significant correlation was observed between HbF and CKD (r=0.02, p<0.71), whereas weak negative correlation was seen between age and HbF in participants without CKD (r= -0.25, P<0.001). HbS and F showed a significant increased risk of developing CKD for having either one or both variants (HbS (OR=1.02-1.07, p=0.001), HbF (OR = 0.95; 95%CI [0.90-0.94], p<0.001). APOL1 showed significant difference between disease status using the combined risk alleles (p=0.021). No significant association was observed between APOL1 risk alleles and SCD-CKD. HbF may modulate SCD complication such as CKD for an extended life expectancy by persisting in adult SCD patients. Risk and development of SCD associated CKD may not be solely attributed to presence of the APOL1 risk variants. However, BCL11A may be used as genetic modulators for SCD clinical management. More research using a larger study population is required to confirm these significant findings.

Session Title: Mendelian Phenotypes Poster Session I

PB4737 Elucidating the genetic etiology underlying septo optic dysplasia (SOD).

Authors:

S. T. Beheshti[1], A. Jolly[1], J. R. Lupski[1,2,3,4], R. A. Lewis[1,5,3,4], J. E. Posey[6]; [1]Dept. of Molecular and Human Genetics, Baylor Coll. of Med., Houston, TX, [2]Human Genome Sequencing Ctr., Baylor Coll. of Med., Houston, TX, [3]Dept. of Pediatrics, Baylor Coll. of Med., Houston, TX, [4]Texas Children's Hosp., Houston, TX, [5]Dept. of Ophthalmology, Baylor Coll. of Med., Houston, TX, [6]Baylor Coll. of Med., Houston, TX

Septo-optic dysplasia (SOD) is a rare congenital disorder with an estimated incidence of 1 in 10,000 newborns. SOD is a cerebral midline anomaly, and its classic triad of features includes optic nerve hypoplasia (ONH), agenesis of the septum pellucidum and corpus callosum, and hypoplasia of the hypothalamo-pituitary axis. The presence of two or more components of the classic triad is "diagnostic" for SOD. Despite this narrowly defined clinical definition, only 40-47% of individuals who receive a diagnosis of SOD will have the full triad of SOD features. Moreover, additional brain abnormalities have been observed in SOD individuals, including hydrocephalus, polymicrogyria, and grey matter heterotopia, among others. The full molecular etiology of SOD is not yet known; however, it is widely believed that a combination of genetic predisposition and prenatal environmental factors contribute substantially to its development. Several genes, such as *HESX1, SOX2/SOX3*, *OTX2*, *GLI2*, *PAX6*, *PROP1*, and *TAX1BP3* have been identified to play crucial roles in the embryonic development of the eyes, optic nerves, and pituitary gland, thereby influencing the expression of SOD. Despite these advancements in the underlying molecular mechanisms of SOD, a substantial number of SOD families still lack a molecular diagnosis. Through the Baylor Hopkins Center for Mendelian Genomics (BHCMG) and Baylor College of Medicine Genomic Research to Elucidate the Genetics of Rare (BCM-GREGoR) databases, we have identified 12 unrelated SOD probands and their families for exome sequencing (ES) data. Using a rare variant, family-based analysis approach leveraging existing genomic knowledge sources, we have identified putative candidate, rare variants (MAF ≤ 0.1%) predicted to impact protein function in four candidate disease genes that demonstrate biological roles relevant to the SOD phenotype (*KIF26A*, *NFIA*, *TUBB2B*, and *SLCO1A2*). These molecular findings in rare disease may inform septo-optic dysplasia molecular diagnosis and clinical presentation(s).

Session Title: Mendelian Phenotypes Poster Session II

PB4738 Engineering and analysis of new mouse mutants carrying the novel triplications of human chromosome 21 syntenic regions to reveal the unique genotype/phenotype relationship.

Authors:

A. Douglas, X. Jiang, C. Liu, Z. Xing, A. Pao, Y. Yu; Roswell Park Comprehensive Cancer Ctr., Buffalo, NY

Down syndrome (DS), associated with trisomy 21, is one of the most important chromosomal alterations because it is both common and connected with a constellation of medically important co-occurring conditions. DS occurs in approximately 1 in 691 live births and is present in approximately 1 in 1,499 inhabitants in the US; notably, DS is a leading genetic cause of developmental delays and intellectual disabilities, which currently lack effective treatments. The mouse has been used extensively to study DS based on the evolutionary conservation between mouse genomic regions and human chromosome 21. Mouse mutants carrying three copies of human chromosome 21 syntenic regions have been generated and analyzed to understand the impacts of gene dosage increase on DS-associated phenotypes. In this project, we have engineered three new duplication mutants spanning three human chromosome 21 syntenic regions on mouse chromosome 16 by using Cre/loxP-mediated chromosome engineering. These duplications were first engineered in mouse embryonic stem (ES) cells, and these mutant ES cells were then used to generate chimeric mice. Germ-line transmissions led to the establishment of the duplication mutant mouse lines constitutionally. These duplications span the regions between *Setd4* and *Kcnj6* as well as between *Kcnj15* and *Mx2*. A duplication mouse mutant line was also engineered for Dyrk1a and Kcnj6 alone. The behavioral tests were performed for these mutants, including open field test, T maze test, novel objective recognition test, and fear conditioning test. These efforts identified a novel genomic region associated with cognitive deficits and revealed the unexpected effects of simultaneous triplications of *Dyrk1a* and *Kcnj6*. These results will have significant implications for understanding of disease mechanisms of DS and for development of novel and effective treatments for this genetic condition.

Session Title: Mendelian Phenotypes Poster Session III

PB4739 Establishment the transgenic pig model and exploration the molecularmechanisms of rare disease white sponge nevus

Authors:

W. Cai[1,2]; [1]Tongji Univ. Sch. of Med., Shanghai, China, [2]Tenth People's Hosp. Affiliated of Tongji Univesity, Shanghai, China

White sponge nevus (WSN) is an autosomal dominant inherited rare disease and associated with the mutation of Keratin 4 (KRT4) and Keratin 13 (KRT13) gene. Till now, almost all WSN study are focus on case report or mutation report. The mechanism of WSN is still unclear. In this project, we recruited WSN Chinese family, and oral lesions biopsy with hematoxylin and eosin staining showed that patients had significantly pathological characteristics. The mutations of KRT4 and KRT13 gene were detected by PCR and direct sequencing. Sequencing analysis shows that the mutation of KRT13 was T332C, and amino acid change was Leu111Pro. Preliminary molecular research showed that the abnormal degradation of KRT13 protein of WSN probably associate with abnormal ubiquitination process. However, the molecular mechanism of this process in vivo is not comprehensive. In this project, using the CRISPR/Cas9 to build the KRT13 T332C WSN transgenic mini pig model and phenotyping, RNA-seq sequencing expression profiles provide important signaling pathways important to focus on gene transgenic mini pig cell transcriptome and genome comparative analysis to explore the pathogenesis of WSN. The platform of phenotype reversal comprehensive treatment and drug screening to interpretation and validation of molecular pathogenesis mechanism WSN in vivo. It provides an important reference for further explore the molecular mechanism of WSN, clinical treatment and drug screening to interpretation and validation of molecular pathogenesis mechanism WSN in vivo. It provides an important reference for further explore the molecular mechanism of WSN, clinical treatment and drug screening.

Session Title: Mendelian Phenotypes Poster Session I

PB4740 Estimation of carrier frequencies in Finnish population: utilizing the gnomAD database for ACMG recommended carrier screening genes and Finnish disease heritage disorders

Authors:

E. Salminen, M. Kandolin, M. Pöyhönen; Univ. of Helsinki, Helsinki, Finland

Background: ACMG recently published practice resource proposing that Tier 3 carrier screening for conditions with a combined carrier frequency ≥1/200 should be offered for all pregnant patients and those planning a pregnancy. Altogether 96 genes causing autosomal recessive disorders and 18 X-chromosomal genes were considered as appropriate for pan-ethnic carrier screening. In Finland, a relatively small number of founder individuals and strong genetic isolation over centuries have shaped the genetic landscape leading to certain disorders (referred as Finnish disease heritage) being more common than elsewhere. In these disorders, a single founder variant explains majority of the Finnish cases. **Objective:** To study the carrier frequencies in Finnish population for the ACMG suggested carrier screening genes and estimate the carrier frequencies in Finnish disease heritage disorders utilizing gnomAD v2.1 database. **Methods**: gnomAD v.2.1.1 sequence variant data was extracted for the genes causing autosomal recessive and X-linked Finnish disease heritage disorders and for the genes suggested for carrier screening by ACMG. Variants reported as pathogenic or likely pathogenic (P/LP) in ClinVar, loss-of-function (lof) variants in gnomAD (stop_gain, frameshift_variant, splice_acceptor_variant, splice_donor_variant), as well as Finnish founder variants were included. Gene carrier rate (GCR), cumulative carrier rate (CCR) and at-risk carrier couple rate (ACR) were estimated. In addition, estimated number of affected children born annually was estimated for the autosomal recessive disorders. Analysis was performed for Finnish (FIN), non-Finnish European (NFE) and Ashkenazi Jewish (ASJ) subpopulations in gnomAD. Some of the suggested genes were not able to be analyzed due to known variant type (eg single exon deletion, repeat expansion). **Results:** The cumulative carrier rate (CCR) for the Finnish founder variants for Finnish disease heritage disorders was 28,0 % and the Finnish founder variants covered 93,4 % of the CCR for Finnish disease heritage disorders in Finnish population. There were 47 genes having GCR over 0,5 % in Finnish population. The CCRs in the FIN, non-Finnish European, and Ashkenazi Jewish populations were 51,3 %, 46,7 % and 56,9 %, respectively. The CCR of Finnish heritage diseases in each population was 30,0 %, 9,1 %, and 5,8 %, respectively. At-risk couple rates in AR conditions for the Finnish, non-Finnish European and Ashkenazi Jewish populations were 1,6 %, 1,3 % and 2,6 %, respectively. In Finland, 133 children affected with an AR condition studied are estimated to be born annually.

Session Title: Mendelian Phenotypes Poster Session II

PB4741 Existing and novel tumor associations in NF1: the DiscovEHR cohort experience.

Authors:

A. Pemov[1], J. Kim[2], J. S. Haley[3], U. Mirshahi[4], H. Rao[5], DiscovEHR Collaboration, D. J. Carey[4], D. R. Stewart[2]; [1]NIH, NCI, Rockville, MD, [2]Natl. Cancer Inst., Rockville, MD, [3]Geisinger, Danville, PA, [4]Geisinger Clinic, Danville, PA, [5]Geisinger Hlth.System, Danville, PA

Background. Large-scale exome sequencing linked to electronic health records (EHR) and broad ascertainment of participants' clinical phenotypes provides a unique opportunity to investigate genotype-phenotype correlations, improve the estimates of prevalence and penetrance of Mendelian disorders, and ultimately, advance disease prognosis, risk stratification and care for patients. **Methods**. First, we identified persons with heterozygous pathogenic (P) or likely pathogenic (LP) variants as well as large deletions (CNV) in *NF1*, the causative gene in neurofibromatosis type 1 (NF1). We then examined EHR (C, D an Z icd10 codes for neoplasms) and tumor registry records for the participants carrying these variants, combined similar neoplasms in 20 larger groups, and performed a series of association analyses of carriers (participants with P/LP or CNV in *NF1*, n=130) vs. non-carriers (participants without any rare (<1% population frequency in gnomAD) *NF1* variants, n=55,000) for these 20 tumor phenotypes. Age, sex, BMI, and smoking status were used as co-variates. **Results.** Out of 170,053 participants, we identified 123 with P/LP variants, and 7 with *NF1* CNV (prevalence 1:1,308). The association analysis identified the following significant (p-value<0.0025, after the Bonferroni correction) tumor phenotypes that were previously shown to be involved in NF1: adrenal/neuroendocrine, GIST, peripheral nervous system, central nervous system, and soft and connective tissue tumors. In addition, we identified several novel significant associations such as bone, head/face/neck, leukemia/lymphoma, ovary, and thorax/mediastinum tumors. There were several suggestively significant associations as well (nominal p-value<0.05), including breast (p-value=0.0046), melanocytic nevi/melanoma (p-value=0.0034), salivary (p-value=0.0092), and uterine (p-value=0.023) tumors. **Conclusions.** By utilizing a genome-first approach in the DiscovEHR cohort, we confirmed existing tumor associations with NF1 and identified potential novel neoplasm risks in NF1 patients. One important distinction between this and previous association studies is that deleterious *NF1* variant carriers rather than clinically affected NF1 individuals were the focus of the present investigation. Given the wide range of expressivity of the NF1 phenotype, previous studies might have underestimated prevalence of some tumors in the NF1 population. A validation of these findings in an independent cohort is underway.

Session Title: Mendelian Phenotypes Poster Session III

PB4742 † *EXOC6B* associated with spondyloepimetaphysealdysplasia with joint laxity type 3 plays an essential role in primary ciliogenesis and chondrogenic differentiation *in vitro*

Authors:

V. Guleria, N. Quadri, P. Upadhyai; Kasturba Med. Coll., Manipal, Manipal, India

Spondyloepimetaphyseal dysplasia with joint laxity type 3 (SEMDJL3) is a rare skeletal dysplasia characterized by profound skeletal abnormalities and joint laxity. It is associated with *EXOC6B*, a component of the exocyst complex, involved in vesicle tethering and exocytosis at the plasma membrane. So far, the underlying molecular mechanisms driving the pathogenesis of SEMD-JL3 remain largely unknown. In ATDC5 prechondrocytes, intracellular localization of Exoc6b was determined by immunofluorescence. It appeared predominantly perinuclear and localized at the base of primary cilia. Immunofluorescence and western blot analysis revealed that *Exoc6b* shRNA lentiviral knockdown (KD) in prechondrocytes significantly impaired primary ciliogenesis and expression of Ift88, an essential ciliary protein. Chondrogenic differentiation was evaluated by extracellular matrix (ECM) mineralization and estimation of transcript levels of chondrogenic markers using Alizarin Red staining and reverse transcription quantitative real-time PCR. In *Exoc6b* KD prechondrocytes *Col2a1* and *Ihh*, markers of proliferative chondrocytes, were upregulated following 7 and 14 days of differentiation. At 14 days of differentiation, the chondrocyte hypertrophy marker, *Col10a1* mRNA levels were increased. Congruently, at day 14 of differentiation, *Axin2*, a canonical Wnt pathway modulator that inhibits chondrocyte hypertrophy, was suppressed. The expression of *Mmp13* and *Adamts4* that are terminal chondrocyte hypertrophy markers involved in ECM remodeling were downregulated at 7 and 14 days of chondrogenesis. Finally, at 14 days of chondrocyte differentiation, *Bglap*, which encodes for the most prevalent non-collagenous bone matrix component and ECM calcification, was suppressed. This study provides initial molecular insights into the pathogenesis of SEMD-JL3 by elucidating the cellular distribution and role of Exoc6b in primary ciliogenesis and effects on chondrogenic differentiation.

Session Title: Mendelian Phenotypes Poster Session I

PB4743 Exome sequencing for children with encephalopathy and severe COVID-19 infection

Authors:

N-C. Lee, T-Y. Yen, H-H. Yu, Y-H. Chien, W-L. Hwu, W-T. Lee, L-Y. Chang, COVID study group; Natl. Taiwan Univ. Hosp., Taipei, Taiwan

Background Host genetic factors are one of major factors in the consequence after COVID-19 infection. A variety of genes related to immunodeficiency or immune dysregulation, such as TLR3 pathway and RNA metabolism, innate RNA POL III signaling had been to this condition. In Taiwan, a subgroup of children was noted to have rapidly progress encephalopathy during COVID-19 pandemic. Therefore, we would like to decipher relationship between host genetic susceptibility and COVID infection in our cohort. **Materials and Methods** During May 2022 to May 2023, children with encephalopathy and severe COVID-19 infection were enrolled. Genomic DNA from peripheral blood was collected. Exome sequencing was performed with KAPA HyperExome kit (Roche) and Illumina NovaSeq 6000 system. Sequencing result was aligned to human reference genome (GRCh38) using Burrow-Wheeler Aligner for short-read alignment (BWA-MEM2) and variant calling will be performed using Genome Analysis Tool Kit (GATK V3.5, Broad institute) and further annotated using ANNOVAR. Monogenic disorders or susceptible genes were filtered with maximum minor allele frequency less than 0.01 and ACMG interpretation criteria. **Results** A total 33 children with 21 boys and 12 girls were enrolled. The median age was 6.5 y (range 0.2-18 y). Among them, 10 were noted to have rare variants related to single gene disorders, including DNM1_related mitochondrial disease, mandibulofacial dysostosis, Guion-Almeida type (EFTUD2), intellectual developmental disorder with macrocephaly, seizures, and speech delay (PAK1), Brugada syndrome (SCN5A), Dent disease 2 (OCRL), POLG_related mitochondrial disorders, GDF1_related heart disorders, LIG4 syndrome, chr4 duplication, and DiGeorge syndrome. For inborn error of immunity, 23 susceptible variants on 19 genes were also identified. Among them, PLB1, TYK2, RANBP2, and TNFSF12 were recurrent. **Conclusion** Children with single gene condition tended to have encephalopathy or severe conditions after COVID-19 infection. A promptly early identification of these conditions may prevent further decline or perceive these at-risk groups earlier.

Session Title: Mendelian Phenotypes Poster Session II

PB4744 Exome sequencing leads, in a consanguineous family, to the identification of a novel gene, SLC22A24, for an autosomal-recessive non-syndromic hearing disorder

Authors:

R. Birkenhager, S. Weis, S. Hollander, S. Arndt, A. Aschendorff, A. Knopf; Univ. Med. Ctr., Dept. of Otorhinolaryngologie, Freiburg, Germany

Hearing impairment is the most common sensorineural disorder in humans. Approximately one of every 1000 newborns is born profoundly deaf and one of every 300 newborns has a permanent mild-to-profound congenital hearing impairment. Hearing disorders can be caused by environmental factors or viral infections, strong sources of noise, ototoxic substances and genetic causes. About ~ 60% of all prelingual hearing disorders are genetic. Inherited hearing disorders are divided into syndromal (SHL) or nonsyndromal (NSHL). Almost 70% of cases of inherited hearing disorders are non-syndromic and mainly due to sensorineural causes. About 80% of this cases follow an autosomal recessive (DFNB) and 18% an autosomal dominant (DFNA) inheritance, about 2% are x-chromosomal (DFNX) or mitochondrial (MT) linked. A total of 196 gene loci have been described to date, from which 131 genes have so far been identified, 65 genes at least are still unknown.Autosomal recessive non-syndromic hearing loss (ARNSHL) is a genetically hetero-geneous sensorineural disorder, with prelingual hearing loss and absence of other clinical manifestations. Based on the clinical diagnosis it is not possible to recognize in which gene mutations are present. This is only feasible in exceptional cases (7/131). The aim of this study is to identify the pathogenic gene in different families. Hearing test BERA/Electrocochleography and a radiological high-resolution CT scan was performed. Mutational analysis of two affected family members, using direct sequencing of the coding exon and intron transitions of the genes GJB2 and GJB6, including deletion analysis, was performed. For investigation of autosomal recessive nonsyndromic hearing loss genes, whole exome sequencing was performed, with the “INVIEW HUMAN EXOME” platform; array Agilent Genomics SureSelectXT All Exon V5. In the DNFB1 gene locus, containing the genes GJB2 and GJB6, no mutations could be identified. A further targeted analysis of other genes was not possible; therefore complete exome sequencing took place. All known genes for hearing impairment were analyzed, in different families. A homozygous splice site mutation IVS1+1G>A was identified in the gene SLC22A24 in a consanguineous family. Sanger sequencing confirmed this mutation. This mutation prevents the expression of the gene SLC22A24, detected by quantitative real-time PCR. SLC22A24 belongs to a large family of transmembrane proteins that function as transport of organic ions across cell membranes. Here, for the first time, the gene SLC22A24 (MIM 611698, Gene ID: 283238) is described that leads to an autosomal recessive hearing disorder.

Session Title: Mendelian Phenotypes Poster Session III

PB4745 Exome Sequencing Links the SUMO Protease SENP7 with Fatal Arthrogryposis Multiplex Congenita, Early Respiratory Failure, and Neutropenia.

Authors:

K. Weiss[1], N. Samra[2], N. S. Jansen[3], I. Morani[2], T. Paperna[1], R. Wevers[4], H. Mandel[5], A. C. Vertegaal[6]; [1]Genetics Inst. Rambam Hlth.Care Campus, Haifa, Israel, [2]Dept. of Genetics Ziv Med. center, Safed, Israel, [3]Dept. of Cell and Chemical Biology, Leiden, Netherlands, [4]Translational Metabolic Lab., Dept. Lab. Med., Radboud Univ. Med. Ctr., Nijmegen, The Netherlands, Nijmegen, Netherlands, [5]Metabolic Unit Ziv Med. center, Safed, Israel, [6]Dept. of Cell and Chemical Biology, Leiden Univ. Med. Ctr., Leiden, Netherlands

Background: SUMOylation involves the attachment of Small Ubiquitin-like Modifier (SUMO) proteins to specific lysine residues on thousands of substrates with target-specific effects on protein function. Sentrin-specific proteases (SENPs) are proteins involved in the maturation and de-conjugation of SUMO. Specifically, SENP7 is responsible for processing polySUMO chains on targeted substrates including the heterochromatin protein HP1α. Methods: we performed exome sequencing and segregation studies in a family with several infants presenting with an unidentified syndrome. RNA and protein expression studies were performed in fibroblasts available from one subject. Results: We identified a kindred with four affected subjects presenting with a spectrum of findings including congenital arthrogryposis, no achievement of developmental milestones, early respiratory failure, neutropenia, and recurrent infections. All died within four months after birth. Exome sequencing identified a homozygous stop gain variant in SENP7 c.1474C>T; p.(Gln492*) as the probable etiology. The proband's fibroblasts demonstrated decreased mRNA expression. Protein expression studies showed significant protein dysregulation in total cell lysates and in the chromatin fraction. We found that HP1α levels as well as different histones and H3K9me3 were reduced in patient fibroblasts. These results support previous studies showing interaction between SENP7 and HP1α, and suggest loss of SENP7 leads to reduced heterochromatin condensation and subsequent aberrant gene expression. Conclusion: Our results suggest a critical role for SENP7 in nervous system development, hematopoiesis, and immune function in humans.

Session Title: Mendelian Phenotypes Poster Session I

PB4746 Exome sequencing of a large Indian cohort of inherited diseases identifies actionable mutations

Authors:

S. Seshagiri[1], S. N[1], S. Thenral[1], I. verma[2], T. Familial Disease Genetics Group[1], R. Gupta[3], V. Ramprasad[1]; [1]MedGenome Labs, Bangalore, India, [2]Inst. of Med. Genetics and Genomics, New Delhi, India, [3]MedGenome Labs Ltd., Bangalore, India

India is the world's most populous country with over 1.4 billion people. Its contemporary population is made of ~4000 anthropological groups that use over 100 languages. Its extraordinary population diversity, in part due to the historical and extant endogamous cultural practices and the ~26 million annual live births is an unprecedented opportunity for genetic discovery, therapeutic development and management of inherited disease burden.
To understand the prevalence of causal variants in patients with a history of inherited genetic disorders from India we created a familial genetic disease study cohort. It included 4697 samples from 1446 families with probands with clinical symptoms in areas of neurology, otolaryngology, ophthalmology, metabolic disorders, nephrology, immunology, skeletal dysplasia, cardiology, gastroenterology, oncology, and others rare disease.
We performed whole exome sequencing of the samples and analyzed them for causal variants using the pedigree and inheritance data. About 68% of the familial samples sequenced were trios. We identified a total 857,403 unique variants, 26% (222,942) of which were not previously reported. Among these, as per ACMG, we prioritized 1,208 (974 SNVs, 193 Indels and 41 CNVs) variants in 1,006 families. This included 596 variants involving 328 genes that were novel and not reported previously. One or more pathogenic variants were identified in 44% of the families. The highest diagnosis rate was obtained for trios that included additional family members (~63%) as compared to singletons where only the proband was sequenced. The highest diagnosis rate of 51% was in families with metabolic disorder. Among the 513 genes prioritized in our study, DYSF, CAPN3, ABCA4, GJB2, GNE, LAMA2, CLCN1, DMD, SGCB and MYO15A were the top 10 genes with highest number of prioritized variants. Among all genes, GJB2, GNE, DMD, SGCB carried the most number of pathogenic variants. Nonsense variant in GJB2 (p.Trp24Ter) was the most common prioritized variant, found in families with otolaryngology symptoms. Hemizygous deletion in the DMD was the most common hotspot CNV in this study. Previously not reported loss-of-function variants in PRKG2 in a skeletal case of dysplasia, USP53 in a case of cholestasis, and GOLGA2 in a case of neuromuscular phenotype were identified. In addition, an inherited missense variant in G6PD linked to Aspirin toxicity symptom was identified and functionally characterized to validate its relevance as a pharmacogenetic marker for pharmacovigilance.
Overall our study has identified genetic variants that confirm and clarify the clinical symptoms while uncovering previously not identified causal variants.

Session Title: Mendelian Phenotypes Poster Session II

PB4747 Expanded Clinical and Neuroradiological Phenotype of *RARS2*-Related Mitochondrial Disorder

Authors:

A. Walimbe[1,2], K. Machol[3,2], S. F. Kralik[4], E. A. Mizerik[2,3], Y. Gofin[5,6], R. Bekheirnia[3,2], C. Gijavanekar[3], S. H. Elsea[3], L. T. Emrick[1,2,3], F. Scaglia[2,3,7]; [1]Div. of Pediatric Neurology and Dev.al NeuroSci., Dept. of Pediatrics, Baylor Coll. of Med., Houston, TX, [2]Texas Children's Hosp., Houston, TX, [3]Dept. of Molecular and Human Genetics, Baylor Coll. Med., Houston, TX, [4]Dept. of Radiology, Baylor Coll. of Med., Houston, TX, [5]Sackler Sch. of Med., Tel Aviv Univ., Tel Aviv, Israel, [6]Genetics Inst., Meir Med. Ctr., Kfar Saba, Israel, [7]BCM-CUHK Ctr. of Med. Genetics, Prince of Wales Hosp., Hong Kong, Singapore

RARS2-related mitochondrial disorder is an autosomal recessive mitochondrial encephalopathy caused by biallelic pathogenic variants in the gene encoding the mitochondrial arginyl-transfer RNA synthetase (*RARS2*, MIM *611524, NM_020320.5). RARS2 catalyzes the transfer of L-arginine to its cognate tRNA during the translation of mitochondrially-encoded proteins. The classical presentation of *RARS2*-related mitochondrial disorder includes pontocerebellar hypoplasia (PCH), progressive microcephaly, profound developmental delay, feeding difficulties, and hypotonia. Most patients also develop epilepsy, which presents within the first three months of life with focal or generalized seizures that frequently become pharmacoresistant and lead to a developmental and epileptic encephalopathy (DEE). Here, we describe a 4-year-old boy with developmental delay, hypotonia, and failure to thrive who developed an epileptic encephalopathy consistent with Lennox-Gastaut Syndrome (LGS), previously unreported in this disorder. He was noted to have dysmorphic features including bilateral macrotia, overriding second toes, a depressed nasal bridge, downslanting palpebral fissures, and retrognathia. His exam did not reveal microcephaly. Whole genome sequencing identified two *in trans* variants in *RARS2*, c.36+1G>T, a previously unreported variant that is predicted to affect splicing, and is, therefore, likely pathogenic, and c.419T>G (p.Phe140Cys), a known pathogenic variant. Unlike most patients with *RARS2*-related mitochondrial disorder, our patient did not demonstrate PCH on brain MRI. Severe volume loss of the cerebral hemispheres and thinning of the corpus callosum were noted. Treatment with a ketogenic diet (KD) reduced seizure frequency and enabled him to make developmental progress. Our study highlights the importance of appropriate seizure phenotyping in *RARS2*-related mitochondrial disorder and indicates that patients can develop LGS, for which a KD may be a viable therapeutic option. This work further reports that patients can also present with dysmorphic features and an absence of progressive microcephaly. In addition, this condition may demonstrate neuroradiological features other than PCH on brain MRI. The case reported herein expands the known phenotypic spectrum associated with *RARS2*-related mitochondrial disorder and may help guide the diagnosis and management of patients with this condition.

Session Title: Mendelian Phenotypes Poster Session III

PB4748 Expanding the clinical spectrum of *COPB2*-related disorder

Authors:

R. Marom[1], K. Boycott[2], D. S. Brightman[3], J. Cohen[4], A. Cortes Fernandez[5], T. Hershkovitz[6], M. Jain[7], B. Keren[8], L. Lambert[9], A. Parida[10], F. Rauch[11], B. N. Simpson[12], F. Mau-Them[13], N. I. Wolf[14], M. Zoetman[15], B. Lee[1]; [1]Baylor Coll. of Med., Houston, TX, [2]CHEO Res. Inst., Ottawa, ON, Canada, [3]Cincinnati Children's Hosp., Cincinnati, OH, [4]Kennedy Krieger Inst, Baltimore, MD, [5]Children's Hosp. Colorado, Aurora, CO, [6]Genetics Inst., Rambam Hlth.Care Campus, Haifa, Israel, [7]Kennedy Krieger Inst, West Friendship, MD, [8]AP-HP.Sorbonne Université, Pitié-Salpêtrière Univ. Hosp., Paris, France, [9]Service de Génétique Clinique, CHRU Nancy, Nancy, France, [10]Birmingham Children's Hosp., Birmingham, United Kingdom, [11]Shriners Hosp Children, Montreal, QC, Canada, [12]Cincinnati Children s Hosp. Med. Ctr., Cincinnati, OH, [13]Pitié-Salpêtrière Hosp., Paris, France, [14]Amsterdam UMC, locatie VUmc, Amsterdam, Netherlands, [15]Mary Bridge Children’s Hlth.Ctr., Tacoma, WY

“Coatopathies”, or disorders of the coatomer complexes, are a group of childhood-onset, multisystem disorders that are growing in number and present with diverse phenotypes including immunodeficiency, intellectual and developmental disabilities (IDD), and skeletal abnormalities. Heterozygous, loss-of-function variants in *COPB2*, encoding a subunit of the coatomer complex I (COPI), were previously reported in individuals presenting with a clinical spectrum of osteopenia, recurrent fractures, and IDD. Here, we describe the molecular and phenotypic features in 5 individuals that were found to have heterozygous variants in *COPB2* via exome sequencing. We identified *de novo* variants leading to either frameshift (p.Lys386Glufs*30, p.Ile414Tyrfs*31), missense (p.Trp396Arg), or splice (c.2484+2dupT, c.894+1G>T) mutations. Clinically, all probands have IDD, including 2 individuals with intellectual disability and 4 individuals with variable degree of motor and speech delays. Spasticity was noted in all of the probands, which is more frequent than reported in the previously published *COPB2* patient cohort. Microcephaly was reported in 1/6 of the probands. Additional neurological features noted include high pain tolerance in 1/6 and neuropathy in 1/6 probands. With regards to the skeletal phenotype, participants were either not known to have osteopenia and fractures (2/6) or were evaluated and confirmed to have a normal bone phenotype (3/6). Interestingly, other connective tissue abnormalities, including joint hypermobility, pes planus and dental anomalies, were noted in one of the participants. In conclusion, we have identified 5 additional individuals presenting with heterozygous variants in *COPB2* and a clinical spectrum of developmental delay, spasticity, and connective tissue abnormalities. Combined with our previously published cohort, developmental delay (in 11/11 individuals) and spasticity (in 9/11 individuals) remain prominent clinical features, while osteoporosis and fractures (in 5/11) are variable. More research is required to further delineate the phenotypic spectrum and long-term outcome in patients with *COPB2*-related disorder.

Session Title: Mendelian Phenotypes Poster Session I

PB4749 Expanding the mutational spectrum in elastin-driven disease through a gene-first approach.

Authors:

T. Luperchio[1], N. Raja[1], E. Hasanzadeh[1], M. Levin[1], M. Chen[1], L. Huryn[2], M. A. Kelly[3], A. E. Justice[4], G. M. Urban[3], K. D. Yu[3], N. A. Walton[5], M. S. Williams[6,3], B. A. Kozel[1]; [1]Natl. Heart, Lung, and Blood Inst., NIH, Bethesda, MD, [2]Natl. Eye Inst., NIH, Bethesda, MD, [3]Dept. of Genomic Hlth., Geisinger, Danville, PA, [4]Geisinger, Danville, PA, [5]Intermountain Precision Genomics, Salt Lake City, UT, [6]Geisinger, Holmen, WI

Elastin is essential in the extracellular matrix and well known to be associated with vascular, skin, and pulmonary conditions. Reduction of elastin dosage (mediated by deletion or stopgain variation of gene *ELN*) is the most common mechanism of disease resulting in supravalvar aortic stenosis (SVAS). Frameshift variants in the 3' end of the molecule are associated with autosomal dominant cutis laxa (ADCL). Despite these genotype-phenotype correlations and the sufficient evidence for haploinsufficency, *ELN* does not demonstrate loss-of-function constraint (pLI=0, o/e=0.64, LEOUF=0.87). In population sequencing databases, missense variation in *ELN* is prevalent though effects on individuals' health are unknown. To understand this phenomenon, we undertook a gene-first approach to understanding the effect of *ELN* missense and splice variation on human health. Using the MyCode Community Health Initiative data, our partners at Geisinger conducted a PheWAS and medical chart review of individuals with rare missense and splice variants of uncertain significance (VUS) in *ELN* and age/sex matched controls (PheWAS only) that revealed vascular disease that was distinct from the phenotypes currently associated with elastin haploinsufficency. Individuals with relevant *ELN* variants were then enrolled in a deep phenotyping study at the NIH. To date, those individuals have displayed near complete penetrance for vascular features, with much of their pathology not captured in the existing medical record. In the first 8 deeply phenotyped individuals, relevant vascular findings include defects of arterial size in 75% (both dilation and narrowing), dissections and aneurysm (12.5%), tortuosity (37.5%) and anatomical variants of the vasculature (50%), suggesting that these VUS may impact both developmental patterning and long-term vascular integrity. While ongoing, this study has already shown that *ELN* variation has a wider mutational spectrum than previously thought, and even missense variants have the potential to confer risk for vascular outcomes. Continued study is needed to further quantify this association.

Session Title: Mendelian Phenotypes Poster Session II

PB4750 Expanding the neurodevelopmental phenotype of HIVEP2-related disorders

Authors:

S. Finlayson[1], K. Park[1], T. Wenger[2]; [1]Univ. of Washington, Seattle Children's Hosp., Seattle, WA, [2]Univ. of Washington, Seattle Children's Hosp., Seattle, WA

HIVEP2 encodes a zinc finger containing transcription factor that is involved in several neuronal processes. The phenotype of *HIVEP2* related disorders is broad and is characterized by intellectual disability, developmental delay, hypotonia, mild dysmorphic features, hyperphagia, hirsutism, microcephaly, gastrointestinal issues, and behavioral issues including autism spectrum disorder (ASD), hyperactivity, oppositional defiance and anxiety. A recent case series by Mo et. al (2021) sought to characterize the neurodevelopmental profile of *HIVEP2*-related disorders. They reported 12 patients with *HIVEP2* pathogenic variants and used standardized assessments to evaluate levels of adaptive functioning, autism symptomatology, and emotional and behavioral characteristics. Notably, these patients were ascertained from a database of patients with pathogenic variants in genes associated with ASD and intellectual disability. Five patients in their cohort had a formal diagnosis of ASD. Their entire cohort showed impairments in all domains of adaptive behavior, as well as significant impairment in social behavior and emotional/behavioral symptoms. They concluded that individuals with *HIVEP2*-related disorder have impairments in adaptive and social-related behaviors as well as difficulties in emotional and behavioral symptoms. We report a 5-year-old patient initially referred due to macroglossia and concern for Beckwith-Wiedemann. She did not fit clinically with BWS aside from her macroglossia and was ultimately found on exome sequencing to have a frameshift pathogenic variant causing truncation of *HIVEP2*. She received early intervention services for gross motor and language delays but graduated out before age three and was not a candidate for developmental preschool. At age five she attends typical kindergarten class and can write her name. Speech language pathology evaluations had no concerns on pragmatics of speech, social reciprocity, or eye contact. She is described as very friendly and can easily make friends with peers and adults. There have been no concerns regarding her adaptive functioning. Current speech concerns are articulation likely secondary to her macroglossia as well as issues with sentence structure with telegraphic speech. Our patient suggests that there is likely a much broader neurodevelopmental phenotype of *HIVEP2*-related disorders than has been reported in the literature thus far. Particularly we may see pathogenic variants in patients who are referred for reasons other than ASD or intellectual disability, as in our patient with macroglossia as a chief complaint.

Session Title: Mendelian Phenotypes Poster Session III

PB4751 Expanding the phenotype of distal deletions in 22q11.2 deletion syndrome to include hemifacial microsomia: a case report.

Authors:

D. Mahoney[1], J. Black[2], A. B. Boura'a[3], R. Lebel[2], S. A. Tatum[4]; [1]Norton Coll. of Med., SUNY Upstate Med. Univ., Syracuse, NY, [2]Ctr. for Dev., Behavior and Genetics, SUNY Upstate Med. Univ., Syracuse, NY, [3]Crouse Hosp. Neonatal Intensive Care Nursery, Syracuse, NY, [4]Dept. of Otolaryngology, SUNY Upstate Med. Univ., Syracuse, NY

22q11.2 deletion syndrome (22q11.2 DS) is the most frequently diagnosed chromosomal microdeletion and it is highly pleomorphic. It is also referred to by many names, often assigned on the basis of certain phenotypic characteristics: velocardiofacial syndrome (OMIM #192430), DiGeorge syndrome (OMIM #188400), conotruncal anomaly face syndrome (OMIM #217095). This high degree of variability is due to haploinsufficiency of many coding genes, particularly T-box Transcription Factor 1 (TBX1, OMIM #602054), DiGeorge Critical Regions (DGCRs) such as DGCR8 (OMIM #609030), and others. About 85% of cases are caused by a recurrent ~2.54 Mb deletion of low copy number repeats (LCRs) A-D, with the remaining 15% being atypical and often smaller deletions. We saw a 13-month-old female with hemifacial microsomia, aural atresia, cleft lip, unilateral renal agenesis, nystagmus, frequent infections, and short stature. The variant identified in this patient is a 432 kbp distal deletion at 22q11.2, involving clusters C-D. The complexity and variability in 22q11.2 DS has led to the discovery of many associations with deletion size and location which could have significant implications for developmental and clinical management. Specifically, there are several reports of hemifacial microsomia, also known as oculo-auriculo-vertebral spectrum (OAVS, OMIM #164210) or Goldenhar syndrome, in 22q11.2 DS, but only a few reports have described patients with this presentation and atypical distal deletions of 22q11.2, and all have included cluster D. These deletions may in fact be a previously unrecognized genetic etiology of hemifacial microsomia, and it can potentially be asserted that it is no longer appropriate to describe such persons as having “Goldenhar syndrome” since we now understand that hemifacial microsomia is oftentimes a component of the complex phenotype associated with 22q11.2 deletion syndrome. Conversely, this indicates that Goldenhar syndrome should not be accepted as a final diagnosis unless microarray has been employed to exclude 22q11.2 DS.

Session Title: Mendelian Phenotypes Poster Session I

PB4752 Expanding the Phenotype of *ZMIZ1* Syndromic Neurodevelopmental Disorder

Authors:

P. Gupta, L. King, L. Fried; St. Joseph's Univ. Med. Ctr., Paterson, NJ

Introduction: Neurodevelopmental disorder with dysmorphic facies and distal skeletal anomalies (NEDDFSA) is a global neurodevelopmental disorder with highly variable features. Patients often show poor feeding, poor overall growth, and hypotonia from early infancy, followed by mildly delayed motor development, poor language acquisition, and behavioral abnormalities. Intellectual development varies from severe with absent speech to mild with the ability to attend special schools. Common features include dysmorphic facial features with notable eye anomalies, joint hypermobility, and mild skeletal anomalies of the hands and feet. Case Presentation: The patient was born to a 26 year old mother G1P0 at 40 weeks gestation, via normal spontaneous vaginal delivery. Pregnancy was complicated by gestational diabetes and polyhydramnios. APGARS were 9 and 9 at 1 and 5 minutes respectively. Birth weight was 7lbs 11 oz. She was admitted in the NICU for apnea, fever, poor feeding and maternal chorioamnionitis. She was noted to have left eye ptosis, bilateral hemivertebrae at T11and a PFO. Head ultrasound revealed a small 5 mm left choroidal cyst. A SNP microarray was normal and revealed a high density of short runs of allele homozygosity (ROH) observed throughout the genome. She was seen for a follow up evaluation at 3 years of age and noted to have global developmental delay, sloping, small forehead, bitemporal narrowing, prominent metopic ridge, left eye ptosis, bilateral clinodactyly; large toes, bilateral 2/3 toe syndactyly, and persistent mongolian spots on the sacrum. Whole genome sequencing on the patient and both parents was performed at a CLIA certified lab which identified a de novo, heterozygous, pathogenic variant in the *ZMIZ1* gene (c.2416_2423delinsCTTCAAACT; p.Ile806LeufsTer7). Patient has a diagnosis of autosomal dominant neurodevelopmental disorder with dysmorphic facies and distal skeletal anomalies (NEDDFSA). This variant introduces a premature termination codon on exon 22 out of 25. It is expected to result in loss of function which is an expected disease mechanism for *ZMIZ1* in this disorder. It is absent from control population and has not been previously reported in individuals with *ZMIZ1* disorders. Discussion: Patient has a diagnosis of autosomal dominant neurodevelopmental disorder with dysmorphic facies and distal skeletal anomalies (NEDDFSA). Only 33 patients have been previously identified. In addition to the previously described phenotype our patient also presented with hemivertebrae, expanding the known phenotypic spectrum for *ZMIZ1* disorders.

Session Title: Mendelian Phenotypes Poster Session II

PB4753 Expanding the phenotypic spectrum of *RAB11B*-related disorder with a case report of the oldest reported patient.

Authors:

A. Poulos[1], R. Kodvawala[1], J. Owens[2], J. Strader[2], K. Widmeyer[2], A. Shillington[2]; [1]Univ. of Cincinnati Coll. of Med., Cincinnati, OH, [2]Cincinnati Children's Hosp., Cincinnati, OH

RAB11B encodes part of the RAS GTPase subfamily involved in regulating vesicular trafficking, exocytotic/endocytotic pathways, and endosome recycling. Five previously reported patients have had pathogenic c.64G>A (p.Val22Met) or c.202G>A (p.Ala68Thr) variants that interfere with the GDP/GTP binding pocket and enhance affinity to SH3BP5. We report the sixth and oldest known patient with *RAB11B*-related disorder. The patient is a 19-year-old female with a history of cortical visual impairment, global developmental delay, intellectual disability, spastic quadriplegic cerebral palsy, microcephaly, eosinophilic esophagitis, and bilateral contractures. At 4 months, she had poor feeding, bilateral clenched fists, and hand wringing that led to an MRI revealing moderate parenchymal loss, gliosis, and prominent extra-axial spaces. She was found to have cortical visual impairment, primary optic atrophy, amblyopia, and esotropia, though she maintains functional vision. By age 5, she had limited mobility due to neuromuscular scoliosis, hypertonicity, and hip and finger contractures. Baclofen, surgery, and physical therapy allowed her to be mobile with a wheeled stander. She stood independently at age 12 but could no longer do so by age 16. Workup for ongoing feeding difficulties at 11 led to a diagnosis of eosinophilic esophagitis. At 18 she developed altered mental status with atypical laughter to no clear stimuli, decreased sleep, and no longer used the 10-20 words she spoke previously. Initial genetics workup at 14 for atypical Rett syndrome was non-diagnostic, but further workup at 18 demonstrated a de novo pathogenic c.64G>A (p.Val22Met) *RAB11B* variant. All six individuals with *RAB11B*-related disorder had intellectual disability, global developmental delay, ataxic and broad-based gait, white matter volume loss, ventriculomegaly, thinning of the corpus callosum, brainstem abnormalities, and vermian hypoplasia. Most individuals also had epilepsy, ophthalmologic abnormalities, microcephaly, and hip dysplasia. It is unclear if her loss of language / mobility and acute change in mental status at age 18 represent neurodegeneration related to *RAB11B* dysfunction, catatonia, or a currently unrecognized etiology. Similarly, it is unclear if her eosinophilic esophagitis could be related to *RAB11B*-related epithelial dysfunction or is unrelated. More research is needed to better understand the neuropsychiatric complications and multi-systemic symptoms associated with *RAB11B*-related disorder.

Session Title: Mendelian Phenotypes Poster Session III

PB4754 Expanding the phonotypic spectrum of WDFY3-related disorder: first case with cerebellar hypoplasia.

Authors:

S. Otsuji[1,2], K. Kato[1], S. Mizuno[2], S. Saitoh[1]; [1]Nagoya City Univ. Graduate Sch. of Med. Sci., Nagoya, Japan, [2]Aichi Dev.al Disability Ctr., Kasugai, Japan

Introduction: WDFY3 is a large multidomain scaffolding protein implicated in the selective degradation of ubiquitinated protein aggregates by autophagy. PH domain variants cause microcephaly, while variants putatively resulting in haploinsufficiency lead to macrocephaly. Cerebellar malformation has not been reported in a patient with a *WDFY3* variant, and thus a developmental role of WDFY3 in cerebellum is unknown. Here we report on a case with cerebellar hypoplasia, in addition to microcephaly caused by a novel heterozygous missense variant in the PH-domain of WDFY3. **Case report:** The case was 6-year-old Japanese female, who was referred to our department due to developmental delay. She was born to unrelated, healthy Japanese parents at 38 weeks' gestation with a birth weight of 2462g (-1.5SD), a height of 47cm (-0.9SD), and a head circumference of 30.5cm (-2.0SD). She had no family history of neurological disorders. Developmental delay was profound with no meaningful words, and she had a friendly personality. She showed severe microcephaly (-4.0 SD), craniofacial features, and small 5th fingers. She also showed postnatal growth retardation (−2.5 SD). Brain MRI revealed hypoplasia of the posterior cerebellar vermis and caudal medial cerebellar hemisphere but no cerebral or white matter lesions. Whole exome sequencing identified a de novo heterozygous missense variant in the PH-domain of WDFY3 (NM_014991.6:c.7821C>G;p.Ser2607Arg), which was evaluated as likely pathogenic according to the ACMG guideline. **Discussion:** This case report provides the first patient with cerebellar hypoplasia along with microcephaly associated with a de novo heterozygous missense variant in *WDFY3*. WDFY3 is known to regulate Wnt signaling in brain development. In cerebellar development Wnt signaling plays a critical role at embryonic rhombic lip (RL). Therefore, WDFY3 mutant may perturb timely attenuation of Wnt signaling at the embryonic RL. Cerebellar hypoplasia might have been missed in a patient with a WDFY3-related disorder, and serial MRI examination would be warranted.

Session Title: Mendelian Phenotypes Poster Session I

PB4755 Expanding the range of WNT signaling syndromes: a promoter variant in *WNT9B* is a candidate in a case with Femoral Facial syndrome.

Authors:

C. Marvin[1], J. X. Chong[1], K. J. Buckingham[1], K. M. Shively[1], A. F. Rope[2], University of Washington Center for Rare Disease Research, M. J. Bamshad[1,3,4]; [1]Dept. of Pediatrics, Univ. of Washington, Seattle, WA, [2]Genome Med., South San Francisco, CA, [3]Dept. of Genome Sci., Univ. of Washington, Seattle, WA, [4]Seattle Children's Hosp., Seattle, WA

The WNT family of genes encode highly conserved secreted signaling proteins expressed in embryonic and adult tissues. During embryonic development these genes regulate cell growth and differentiation through a canonical WNT/β-catenin pathway and non-canonical pathways without β-catenin. Pathogenic germline variants in WNT genes underlie a range of congenital disorders affecting the skeleton, limbs, genitourinary system, integument, dentition, and face. We report a child with clinical features consistent with Femoral Facial syndrome (FFS), also known as Femoral Hypoplasia-Unusual Facies syndrome, including bilateral femoral aplasia, cleft palate, micrognathia, talipes equinovarus, ventriculomegaly, absent septum pellucidum, and genitourinary abnormalities requiring right orchiectomy and left orchiopexy. While maternal diabetes is a known risk factor for FFS, the etiology of the condition remains unknown. Short read genome sequencing of the proband identified a non-maternally inherited heterozygous c.-2C>G variant in the 5'UTR promoter of *WNT9B*. This variant is predicted to be damaging (CADD 18) and is not present in gnomADv2.1.1, gnomADv3.1.2, or TOPMed freeze 8. *WNT9B* is expressed in mouse craniofacial and hindlimb bud tissue. Homozygous variants in *WNT9B* have been reported in two consanguineous families to underlie renal agenesis and hypoplasia with reproductive tract abnormalities but to date no condition is known to result from heterozygous pathogenic variants in *WNT9B*. Expression studies are planned to confirm the pathogenicity of this variant in *WNT9B* and the mechanism by which it may result in FFS.

Session Title: Mendelian Phenotypes Poster Session III

PB4757 Familial Thoracic Aortic Aneurysm & Dissection (FTAAD) candidate genes detection using Whole Exome Sequencing (WES) in a large UK Caucasian family.

Authors:

V. Khurana[1], J. A. Aragon-Martin[1,2,3], S. Babu[1], S. La Penna[1], D. Morris-rosendahl[4], C. Nienaber[4], N. Naqvi[4], F. Dijk[4], N. Ghali[4], A. Reed[4], J. A. R. Gaer[4], J. Pepper[4], U. Rosendahl[4], J. Sneddon[5], E. Fisher[6], J. Dean[7], A. Saggar[8], M. Jahangiri[8], E. Behr[9], M. Simpson[10], A. Smith[10], M. Y. B. Salmasi[2,4], A. Child[2,11], M. Moosajee[3,12], T. Athanasiou[2,13], P. Deloukas[1]; [1]Queen Mary Univ. of London, London, United Kingdom, [2]Imperial Coll., London, United Kingdom, [3]Univ. Coll. London, London, United Kingdom, [4]Royal Brompton Hosp., Guy's and St. Thomas' NHS Fndn. Trust, London, UK, London, United Kingdom, [5]East Surrey Hosp., Redhill, United Kingdom, [6]The Park Surgery, Horsham, United Kingdom, [7]Clinical Genetics, NHS Grampian, Aberdeen, United Kingdom, [8]St. George's Univ. Hosp. NHS Fndn. Trust, London, United Kingdom, [9]St. George's Univ. Hosp., NHS Fndn. Trust, London, United Kingdom, [10]King's Coll. London, London, United Kingdom, [11]Marfan Trust, London, United Kingdom, [12]Moorfiels Eye Hosp. NHS Fndn. Trust, London, United Kingdom, [13]Imperial Coll. Hlth.care NHS Trust, London, United Kingdom

Introduction

A thoracic aortic aneurysm (TAA) is characterised by dilatation of the aorta, which may eventually lead to a dissection (TAAD) or a subsequent fatal rupture. They tend to be asymptomatic and are often not detected till complications occur. TAAD seems to have a monogenic autosomal-dominant mode of inheritance. It can be familial and syndromic, or non-syndromic (~20% familial). Post *ACTA2* (commonest TAAD causative gene), it has become harder to detect other implicated genes due to their low incidence in the TAAD population.From a lab database of >100 TAAD probands (53 with positive family history, 38 sporadic), one of the probands, who had aortic dissection at 48 years was included in the study. The family of the proband was also affected with 2 of his sisters and a daughter having upper borderline sinus of Valsalva measurements (4, 4.2 and 3.9cm respectively). On examination, a complete profile suggestive of MFS, LDS or EDS was not found and were subsequently excluded by testing for *ACTA2*, *FBN1*, *PRKG1*, *TGFBR1/2* and *SMAD3* genes.

Methodology

WES was performed on 3 most distantly related affected individuals from the family pedigree. No known TAAD gene was found to carry a pathogenic variant in all 3 affected individuals. In-silico filtering and analyses of the data was done with an in-house protocol to shortlist candidate genes that were common to all 3 patients by removing heterozygous variants that did not segregate with the disease, had CADD scores <20, and MAF >0.001 in gnomAD and ExAC, and all homozygous variants. The MAF of the most common pathogenic variant in *ACTA2* (R258C, MAF=0.00032) was further utilised to filter the list.

Results

Out of 21,310 variants (original raw VCF file), 7 candidate genes were shortlisted: 6/7 were non-synonymous mutations *(FAT1*, *RNF213*, *KIR3DL2*, *ZNF606*, *PLA2G1B* and *TLR3*), and 1/7 was a non-frameshift deletion (*AMDHD2)*. Out of the 7 candidate genes, 2 were associated with artery wall physiology (*FAT1* & *RNF213*); 2 with immunology; 1 with oxidative injury prevention; 1 with tumour suppression and 1 with carbohydrate and amino-sugar metabolism. The expression profile (>92% in ascending aorta and smooth muscle tissue - Bgee) of *FAT1* & *RNF213* gave further supportive evidence for them as possible causative making them a priority for family segregation analyses, even though *PLA2G1B* and *TLR3* were ranked the highest by the Phenolyser. Exomiser prioritised *NOTCH3*, *RNF213* and *TLR3*. *NOTCH3* was filtered out of the candidate gene list when the *ACTA2* MAF filter was applied.

Conclusions

Family segregation analyses are being performed for these 7 TAAD candidate genes to confirm their association with TAAD phenotype.

Session Title: Mendelian Phenotypes Poster Session I

PB4758 *FGFR3* missense variant in a family with short stature

Authors:

M. Muurinen[1,2,3], O. Mäkitie[1,2,3]; [1]Res. Program for Clinical and Molecular Metabolism, Univ. of Helsinki, Helsinki, Finland, [2]Children's Hosp., Univ. of Helsinki and Helsinki Univ. Hosp., Helsinki, Finland, [3]Folkhälsan Res. Ctr., Helsinki, Finland

Pathogenic variants in the *FGFR3* gene, encoding fibroblast growth factor receptor 3, are known to cause dominantly inherited skeletal dysplasias ranging from lethal thanatophoric dysplasia to achondroplasia (ACH) and hypochondroplasia (HCH). ACH, and the less severe HCH, cause disproportionate short stature with typical findings such as macrocephaly and radiologic abnormalities. The clinical spectrum of these *FGFR3*-related skeletal dysplasias is wide, and in a few cases *FGFR3* variants have been reported to cause isolated short stature. We report a young girl with short stature, who was referred to pediatric care during infancy. Her length was -2.4 SD at birth, -4.0 SD at four months, and -4.0 at 2,5 years. Her father was of normal height and mother had short stature (139 cm) of unknown etiology. Investigations for underlying chronic illness and molecular karyotype of the child were normal. Next, a gene panel study was undertaken to find out the etiology. The gene panel, which included more than 200 genes associated with skeletal dysplasias, identified a novel heterozygous missense variant c.937G>T, p.(Gly313Cys) in the *FGFR3* gene. The variant was not present in large reference populations, and it was predicted to be pathogenic by multiple prediction algorithms. The patient was disproportionate especially in the lower limbs but did not present with the typical clinical picture of ACH or HCH, including macrocephaly and typical facial features. Radiographs showed only very mild abnormalities, including mild metaphyseal changes and slightly short femoral necks. Analysis of parental samples revealed the same variant in the child's mother, who also had short stature but lacked major features typically seen in *FGFR3*-related skeletal dysplasias. We describe a novel missense *FGFR3* variant c.937G>T, p.(Gly313Cys) in a family with significant short stature without typical ACH/HCH. Phenotype relating to this variant has not been published previously. The finding expands the phenotypic range of FGFR3 variants at the milder end of the disease spectrum. New therapies are emerging for FGFR3-related skeletal disorders, and it remains to be established whether these could be used to promote growth even in the milder forms of the disorders.

Session Title: Mendelian Phenotypes Poster Session II

PB4759 Four and a Half LIM Domains 1 variants: only a male matter?

Authors:

M. Cason[1], C. Calore[1], M. Losi[2], L. Bello[1], F. Borrelli[2], G. Canciello[2], F. Manganelli[2], S. Tozza[2], S. Chen[3], M. Bueno Marinas[1], R. Celeghin[1], M. Perazzolo Marra[1], G. Todde[2], G. Frisso[2], M. Taylor[3], R. Lombardi[2], K. Pilichou[1]; [1]Univ. of Padua, Padua, Italy, [2]Univ. of Naples, Naples, Italy, [3]Univ. of Colorado, Denver, CO

Background. Four and a Half LIM Domains 1 (*FHL1* - MIM*300163) is located on the chromosome Xq26.3 and alternative splicing of this gene is leading to three different major isoforms: *FHL1A*, *FHL1B* and *FHL1C*, with tissue specific expression patterns. FHL1A is a zinc-finger protein with multiple roles in biological processes in heart and skeletal muscle. *FHL1* rare genetic variants, inherited as dominant traits, have been associated with Scapuloperoneal myopathy (MIM# 300695) and rare cases of Hypertrophic Cardiomyopathy (HCM). *FHL1*-related cardiomyopathy is poorly characterized due to the few and isolated case series. This study aims to establish the prevalence and clinical characterization in HCM patients with *FHL1* variants.

Materials and Methods. 723 unrelated European/US HCM patients referred from 3 Referral Centers for hereditary cardiomyopathies, underwent whole exome sequencing. Data including personal and family history, physical examination, 12-lead ECG, echocardiography (ECHO) and Cardiac Magnetic Resonance (CMR) were collected for genotype-phenotype correlation. Neuro-muscular examination occurred upon request. Genetic variants were classified according to the current ACMG criteria. Cascade genetic screening was carried out.

Results. Eight of the 723 HCM index cases resulted carrying 5 unique *FHL1* genetic variants: 2 Pathogenic (P) and 3 Likely pathogenic (LP). No P or LP variants in sarcomeric genes were identified in these 8 patients. Interestingly, genetic variants identified affected all three *FHL1* trancripts.

All but one *FHL1* carriers exhibited non-obstructive HCM associated with scapulo-peroneal myopathy, documented by ECG with marked ventricular repolarization abnormalities and high rate of atrial fibrillation, ECHO and CMR showed diastolic ventricular dysfunction, atrial enlargement and variable non-ischemic patterns of late gadolinium enhancement. Cerebral palsy was observed in the *FHL1* carrier with non-obstructive HCM without scapulo-peroneal myopathy.

Four of the 8 families underwent clinical and genetic evaluation. Hemizygous carriers (males) presented frequent events of heart failure often related to arrhythmic episodes and Sudden Cardiac Death (SCD), while female heterozygous carriers showed mild cardiac phenotypes.

Conclusion. Heterozygous and hemizygous *FHL1* carriers display different degrees of a systemic disease characterized by non-obstructive HCM and scapulo-peroneal myopathy. Adverse prognosis is related to male carriers with early onset of HCM, severe diastolic dysfunction, higher risk of SCD risk and/or increased evolution toward heart failure requiring heart transplantation.

Session Title: Mendelian Phenotypes Poster Session III

PB4760 POLD3 deficiency is associated with syndromic severe combined immunodeficiency including neurodevelopmental delay and hearing impairment.

Authors:

C. Mehawej[1], E. Chouery[1], S. Azar-Atallah[1], W. Shebaby[1], V. Delague[2], I. Mansour[3], M. Mustapha[4], G. Lefranc[5], A. Megarbane[1,6]; [1]Lebanese American Univ., Byblos, Lebanon, [2]INSERM UMR_S 910, Marseille, France, [3]Faculty of Hlth.Sci., American Univ. of Sci. and Technology, Beirut, Lebanon, [4]Sheffield Univ., Sheffield, United Kingdom, [5]Université de Montpellier, Montpellier, France, [6]Inst. Jérôme Lejeune, Paris, France

Combined immunodeficiency diseases (CID) represent the most severe forms of inborn errors of immunity. Defective T cell development and/or function, leading to an impairment in adaptive immunity are responsible for these diseases. Genome duplication and maintenance, regulated by the DNA polymerase δ complex has been recently shown to contribute to CID. Biallelic mutations in *POLD1*, encoding the catalytic subunit of this complex or in *POLD2*, encoding an accessory subunit of the DNA polymerase δ have been recently linked to a syndromic CID with or without intellectual deficiency and sensorineural hearing loss. Here we report a homozygous *POLD3* variant (NM_006591.3; p.Ile10Thr) in a Lebanese patient, the product of a consanguineous family, presenting with a syndromic severe combined immune deficiency (SCID) with neurodevelopmental delay and hearing loss. POLD3 is the accessory subunit 3 of the DNA polymerase δ that stabilizes the complex. The homozygous POLD3^{Ile10Thr} variant abolishes POLD3 as well as POLD1 and POLD2 expression. Our findings implicate POLD3 deficiency as a novel cause of syndromic SCID.

Session Title: Mendelian Phenotypes Poster Session I

PB4761 Functional analysis of *ALG1* variants of uncertain significance

Authors:

E. Gillesse[1], P. Au[1,2], F. Bernier[1,2], V. Li[3], J. Parboosingh[1,2], Care4Rare Canada Consortium, A. Innes[1,2], V. Zaremberg[3], R. Lamont[1,2]; [1]Dept. of Med. Genetics, Univ. of Calgary, Calgary, AB, Canada, [2]Alberta Children's Hosp. Res. Inst., Calgary, AB, Canada, [3]Dept. of Biological Sci., Univ. of Calgary, Calgary, AB, Canada

A common issue in the diagnosis of rare genetic diseases is identification of a single heterozygous pathogenic variant without detection of a second compounding variant in a gene associated with an autosomal recessive disorder. If a second variant cannot be identified, a molecular diagnosis cannot be confirmed. This issue may be in part due to technical limitations of first-line genetic tests, such as exome sequencing, which is primarily limited to the detection of small coding variants. In comparison, whole-genome sequencing (WGS) has better coverage across the entire genome and allows detection of all types of genetic variants, including intronic and structural variants. Here, we present two deceased siblings with severe epileptic encephalopathy in association with pachygyria and cerebellar hypoplasia, for whom a known pathogenic, paternally inherited heterozygous variant in *ALG1* was detected by initial exome sequencing. *ALG1* is associated with autosomal recessive type 1 congenital disorder of glycosylation, and was felt to be a good clinical fit for the siblings' phenotype. Using WGS, both siblings were found to carry a maternally inherited duplication of exon 10 of *ALG1*. The is the first report of a copy number variant in this gene. Further, this duplication maintains the transcript reading frame, therefore impact on protein function is not clear. Variant pathogenicity was assessed using a conditional yeast assay, as ALG1 is an essential protein in *Saccharomyces cerevisiae*. We tested our duplication of interest alongside 11 other variants of uncertain significance (VUSs) reported in the ClinVar database. Two known pathogenic and benign variants were also tested to establish assay validity. Expression clones containing the human *ALG1* sequence with the variant of interest were transformed into strains with a haploid *ALG1* knockout and the wild-type yeast *ALG1* sequence under control of an inducible *GAL1* promoter. The pathogenicity of variants was assessed based on growth of transformed yeast in nonselective media against the wild-type *ALG1* allele (glucose). Western blots identified the presence of downstream glycosylation products of ALG1. In summary, we apply a new assay design for testing variants in *ALG1* to clarify the pathogenicity of several reported VUSs, including the first reported copy number variant in this gene. These results will help confirm a molecular diagnosis for the described family, and potentially contribute to the diagnosis of other affected individuals in the future.

Session Title: Mendelian Phenotypes Poster Session II

PB4762 Functional analysis of pathogenic *LAMB1* variants in the nervous system of *Drosophila melanogaster*.

Authors:

R. Yasuda[1,2,3], H. Hashimoto[1,2], M. Oka[1,2], J-W. Mok[1,2], O. Kanca[1,2], T. Mizuno[3], S. Yamamoto[1,2]; [1]Dept. of Molecular & Human Genetics, Baylor Coll. of Med., Houston, TX, [2]Jan & Duncan Neurological Res. Inst., Texas Children's Hosp., Houston, TX, [3]Dept. of Neurology, Graduate Sch. of Med. Sci., Kyoto Prefectural Univ. of Med., Kyoto, Japan

Laminins are made from alpha, beta, and gamma subunits, which are encoded by five *LAMA*, four *LAMB*, and three *LAMC* genes in humans, respectively. Homozygous nonsense variants in *LAMB1* were originally reported in patients with congenital leukoencephalopathy and severe developmental delay. Recently, we reported the first adulthood-onset leukoencephalopathy patient with a rare homozygous missense variant. More recently, heterozygous late-truncation alleles of *LAMB1* were reported in middle-aged to elderly patients with leukoencephalopathy and hippocampal memory defects. Although nonsense, missense, and late-truncating variants in *LAMB1* are associated with a diverse severity of symptoms, the functional consequences of these variants are largely unknown. In addition, while the function of *LAMB1* orthologs has been studied during development in various model organisms, the role of this gene in the mature nervous system has been ill-defined. We have been developing genetic reagents in *Drosophila melanogaster* to study the function of human *LAMB1* and its role in the post-developmental nervous system. *Drosophila LanB1* is orthologous to all four human *LAMB* family genes. First, we found that *LanB1* is expressed in a subset of glial cells in developing and mature fly brains. Second, we found that biallelic knockout of *LanB1* causes lethality, consistent with earlier reports that this is an essential gene in flies. Third, while ubiquitous knockdown of *LanB1* also caused lethality, neuron and glia-specific knockdown failed to induce obvious phenotypes. Fourth, we found that human LAMB1 was not able to rescue the lethality of fly *LanB1* knockout mutants. These data suggest that while fly *LanB1* is an essential gene that is expressed in the nervous system during development and post-development, human LAMB1 may lack the ability to replace the function of the fly gene and additional studies are required to assess the function of this gene in the nervous system. To further assess whether *Drosophila* could be used to study the functional consequences of disease-associated variants, we designed additional transgenic lines that allow expression of human LAMB1 carrying 8 pathogenic variants. While reference and variant LAMB1 fails to induce any scorable phenotypes when overexpressed in various tissues in *Drosophila*, our preliminary data suggests that there are some differences in the subcellular location pattern of variant LAMB1 compared to the reference. In conclusion, while *Drosophila* can be used to extract biological information that is relevant to understand the mechanisms of human *LAMB1*-related disorders, there are some limitations that need to be taken into considerations.

Session Title: Mendelian Phenotypes Poster Session III

PB4763 Functional assessment of human *GLRA2*-variants associated with neurodevelopmental disorders in *Drosophila melanogaster*.

Authors:

P. Bhadane[1,2], R. Marom[3], P. Liu[1], J. Andrews[1,2], K. Jay[1,2], S. Jangam[1,2], J. A. Madden[4], M. Willemsen[5], C. Muss[6], C. Curry[7], S. Yamamoto[1,2], M. Wangler[1,2]; [1]Baylor Coll. of Med., Houston, TX, [2]Jan and Dan Duncan Neurological Res. Inst., Houston, TX, [3]Baylor Coll. Med., Houston, TX, [4]Boston Children's Hosp., Boston, MA, [5]Radboud Univ. Med. Ctr., Nijmegen, Netherlands, [6]Nemours/A.I duPont Hosp. for Childern, Wilmington, DE, [7]Univ California, San Francisco/ Fresno, Carmel, CA

Pathogenic variants in the *Glycine Receptor Subunit Alpha-2 (GLRA2)*, are associated with intellectual and developmental disabilities (IDD), also known as Intellectual developmental disorder, X-linked syndromic, Pilorge type (MIM 301076). We had previously established a *Drosophila* model to study the mechanism of *GLRA2* variants. In this model, we identified that loss-of-function variants lead to *GLRA2*-related IDD in males, and gain-of-function variants result in unique symptoms in female patients. We used proband variant cDNA overexpression and rescue approaches to determine gain- and loss-of-function mechanisms in flies. Here, we describe the molecular and phenotypic characteristics in five additional probands with variants in *GLRA2*. We identified missense variants in *GLRA2* (p. Gly83Arg, p. Lys310Thr and p. Arg352Gln) in four males, including two siblings, from three unrelated families. Two of the variants were maternally inherited, while segregation was not possible in one family. All probands present with variable degree of developmental delay and behavioral abnormalities, including autism spectrum disorder and attention-deficit and hyperactivity disorder. In addition, we identified a *de novo* missense variant (p. Gln220Pro) in a female proband with epileptic encephalopathy, profound intellectual disability, microcephaly, and brain malformations (abnormal cortical gyration and lissencephaly). Ocular findings included myopia in two of the male probands, and retinal dystrophy with nystagmus in the female proband. In addition, two of the probands presented with growth retardation and two of the probands were noted to have craniofacial dysmorphism (features that were not reported in our previously published cohort). We are generating the fly model for these new variants. Functional validation of the variants will be performed using overexpression and rescue experiments to compare with the previously established *GLRA2*-fly model. In summary, we report five novel *GLRA2* variants in probands with a phenotypic spectrum that fits the previously described *GLRA2*-related IDD.

Session Title: Mendelian Phenotypes Poster Session I

PB4764 Functional characterization of the developmental delay-associated 16p12.1 deletion using iPSCs.

Authors:

J. Sun[1,2,3], M. Das[1,2,3], S. Noss[1,2,3], D. Banerjee[1,2,4], S. Girirajan[1,2,5]; [1]The Pennsylvania State Univ., University Park, PA, [2]Dept. of Biochemistry and Molecular Biology, University Park, PA, [3]Molecular, Cellular, and Integrative BioSci.s Graduate Program, The Huck Inst.s of Life Sci., University Park, PA, [4]Bioinformatics and Genomics Graduate Program, University Park, PA, [5]Dept. of Anthropology, University Park, PA

Copy number variants (CNVs) are associated with neurodevelopmental disorders with heterogeneous clinical phenotypes. 16p12.1 deletion (520-kb) is a CNV that is strongly associated with developmental delay, autism, intellectual disability, and craniofacial abnormalities. Previous studies have shown that the 16p12.1 deletion interacts with variants in the genetic background to modulate clinical manifestation, suggesting a "two-hit" model for the phenotypic variability. In this study, we generated the induced pluripotent stem cell (iPSC) model with 16p12.1 deletion to characterize the deletion and its genetic interactions in the context of human genome background. We found that the deletion led to a significant increase in the number of differentially expressed genes (DEGs) as iPSCs converted to neural lineage, suggesting cell type-specific effects of 16p12.1 genes. Among the DEGs, genes such as *DLX1*, *DLX2*, and *DLX5*, which are associated with interneuron differentiation and migration, were expressed at significantly higher levels in the CRISPR/Cas9-mediated deletion line. In agreement with the RNA-seq results, we observed an increased ratio of NKX2.1-positive cells to PAX6-positive cells in neural progenitor cell (NPC) populations by immunostaining, indicating altered dorsal-ventral pattern formation caused by the deletion. Further, we performed RNA sequencing on iPSCs derived from 13 individuals (8 carriers and 5 noncarriers) from three families, as well as NPCs differentiated from the iPSCs. We identified shared and distinct DEGs among families between carrier NPC lines and noncarrier NPC lines, including known risk genes such as *FOXG1*, *CNTN6*, and *KCNQ5*. Gene ontology (GO) analyses of DEGs revealed significant enrichment related to neuronal function, such as "axon guidance", "axonogenesis", and "synapse organization", and signaling pathways, such as "Notch signaling pathway" and "Wnt signaling pathway". In addition, using EdU and TUNEL assays, we also found distinct rates of cell proliferation and apoptosis among families. Overall, our results show the effects of 16p12.1 deletion on brain development, as well as family-specific alterations on cellular phenotypes and transcriptome profiles.

Session Title: Mendelian Phenotypes Poster Session II

PB4765 † Functional filter for whole genome sequencing data identifies hereditary haemorrhagic telangiectasia non-coding *SMAD4* polyadenylation site variants 5kb from coding DNA

Authors:

C. Shovlin[1], Z. Kai[2], D. Murphy[3], D. Li[1], D. Patel[1], A. Bielowka[1], M. E. Bernabeu-Herrero[1], A. Abdulmogith[1], S. McCarley[1], A. D. Mumford[4], S. Westbury[4], M. A. Aldred[5], N. Vargesson[6], M. Caulfield[7], Genomics England Research Consortium, S. Xiao[8]; [1]Imperial Coll., London, United Kingdom, [2]Topgen Biopharm Technology Co. Ltd, Shanghai, China, [3]Imperial Coll. Hlth.care NHS Trust, London, United Kingdom, [4]Univ. of Bristol, Bristol, United Kingdom, [5]Indiana Univ., Indianapolis, IN, [6]Univ. of Aberdeen, Aberdeen, United Kingdom, [7]Queen Mary Univ. of London, Queen Mary University of London, United Kingdom, [8]Univ. of Oxford, Oxford, United Kingdom

Despite whole genome sequencing (WGS), many single gene disorder cases remain unsolved. Since early data analytic steps prioritize protein-coding sequences, we developed GROFFFY which prioritized 44.4% of the human genome based on experimental evidence of functionality, and combined with additional filters (maximum allele frequency 0.0001; minimum combined annotation dependent depletion (CADD) score 10), for WGS analyses in the UK 100,000 Genomes Project (100KGP). Data from patients with hereditary hemorrhagic telangiectasia (HHT), an autosomal dominant trait, were used for validation and variant discovery. At our reference centre, for ~1 in 7 (74/486) clinically-confirmed HHT probands, no pathogenic or likely pathogenic variants were identified by panel tests for known HHT genes (ENG, ACVRL1, SMAD4, GDF2), and recruitment to 100KGP was allowed without prior genetic testing, providing valuable datasets. First, at the time of GROFFFY application, 34 cases had HHT-causative variants already identified by 100KGP clinical pipelines. These provided a validation dataset where GROFFFY-based filtration reduced mean variants per DNA from 4,867,167 to 21,486, retaining intergenic proportions (one-third), and all HHT-causal variants. Second, since a proportion were expected to have conventional loss-of-function variants in known genes, in the "Discovery" dataset of 98 DNAs, GROFFFY was further validated by identifying clusters of variants in HHT gene coding sequences. In 3 DNAs (2 from an affected duo), unique variants were identified in the 3' untranslated region (UTR) of SMAD4. Sited 5.4kb distal to coding DNA, the deletions did not modify or generate microRNA binding sites, but instead disrupted the sequence context of the final cleavage and polyadenylation site necessary for protein production: By iFoldRNA, an AAUAAA-adjacent 16 nucleotide deletion brought the cleavage site into inaccessible neighboring secondary structures, while a 4-nucleotide deletion unfolded the downstream RNA polymerase II roadblock. Clinical correlation supported HHT-SMAD4 phenotypes (SMAD4 germline loss-of-function alleles cause a combined HHT and polyposis/aortopathy syndrome). Monocyte SMAD4 RNA expression differed between patients and controls in resting and cycloheximide-stressed states. Patterns predicted the mutational site for an unrelated clinically diagnosed SMAD4 case, where a complex insertion was subsequently identified. In conclusion, a new type of functional rare variant is described, exposing novel regulatory systems based on polyadenylation. Extension of coding sequence-focused gene panels is required to capture these variants.

Session Title: Mendelian Phenotypes Poster Session III

PB4766 Further validation of association of bi-allelic variants in *SLC25A10* with developmental and epileptic encephalopathy.

Authors:

M. Yeole[1], P. Majethia[1], M. Jain[1], S. Siddiqui[2], S. Chakrabarty[3], G. Katta Mohan[1,4], A. Shukla[1]; [1]Kasturba Med. Coll., MAHE, Manipal, Manipal, India, [2]Dept. of Neuroimaging and interventional radiology, STAR Inst. of NeuroSci.s, STAR Hosp., Hyderabad, India, [3]Dept. of Cell and Molecular biology, Manipal school of life Sci.s,, MAHE, Manipal, Manipal, India, [4]Coll. of medicine and Hlth. Sci.s, Sultan Qaboos Univ., Muscat, Oman

Background/Objectives: Mitochondrial depletion syndrome 19 (MTDPS19; MIM #618972) is a recently described severe mitochondrial neurodegenerative disorder characterized by developmental delay, refractory seizures, neuroregression, and severe encephalopathy. Currently, one individual has been reported carrying bi-allelic missense variant in *SLC25A10* (MIM #606794).**Clinical details:** We ascertained two individuals from India, proband 1 (P1) and proband 2 (P2). P1, born from a consanguineous union presented with global developmental delay, seizure disorder with progressive encephalopathy features, regression of attained milestones and contractures at all major joints. P2, born from a non-consanguineous union presented with seizure disorder and regression of attained milestones. Singleton exome sequencing (ES) followed by Sanger validation and segregation was performed for the families.On singleton ES, we identified homozygous missense variant, c.449G>A p.(Arg150His) in P1, and c.507C>G p.(Ser169Arg) in P2 in *SLC25A10* (NM_012140.5). Sanger validation and segregation showed the variant in heterozygous state in parents of both the individuals (P1 and P2). This variant is absent in population databases like gnomAD, Singapore genome database, GenomeAsia and our inhouse database of 2849 exomes. *In silico* analysis tools predicted the variant as damaging to the protein function.Previously reported one individual with bi-allelic variants in *SLC25A10* had clinical findings of progressive form of epileptic encephalopathy with hypotonia which were similar to the phenotypes observed in P1 and P2.**Conclusion:** This report of two additional individuals provides further evidence for a novel autosomal recessive mitochondrial depletion disorder caused by pathogenic bi-allelic variants in *SLC25A10*.**Grant References:** 1R01HD093570-01A1(NIH,USA)Indian Alliance/DBT Wellcome Trust funded project, "Centre for rare disease diagnosis, research and training" (IA/CRC/20/1/600002)

Session Title: Mendelian Phenotypes Poster Session I

PB4767 † GAIN OF FUNCTION PATHOGENIC VARIANTS IN*RIT1*CAUSE CENTRAL CONDUCTING LYMPHATIC ANOMALY

Authors:

S. Sheppard[1], S. Paulissen[1], C. McInerney[2], M. March[2], D. Shukla[1], Y. Dori[2], H. Hakonarson[2], D. Li[2]; [1]NICHD/NIH, Bethesda, MD, [2]Children's Hosp. of Philadelphia, Philadelphia, PA

Central conducting lymphatic anomaly (CCLA) is a devastating lymphatic disorder and only 40% of individuals receive a genetic diagnosis. We identified six individuals with CCLA due to pathogenic variants in *RIT1*, including one novel mosaic variant. RIT1 is a RAS GTPase lacking a CAAX domain and is involved in mitogen activated kinase signaling. Germline pathogenic variants cause Noonan syndrome, and mosaic variants have not previously been reported in association with complex lymphatic anomalies. Five out of six individuals, all with germline pathogenic variants, have typical features of Noonan syndrome. All six individuals had CCLA manifesting as pulmonary lymphangiectasia, protein losing enteropathy, pleural effusions, chylopericardium, plastic bronchitis, pleural effusions, and ascites. Dynamic contrast magnetic resonance lymphangiography demonstrated abnormalities of the thoracic duct, dermal backflow, retrograde mesenteric flow, and hepatopulmonary connections.
To evaluate the effect of the RIT1 p.Met90Ile on lymphatic development we conducted *in vitro* and *in vivo* modeling. Spheroid sprouting assay demonstrated significantly increased number of sprouts and cumulative sprout length compared to RIT1 wild type control ($n = 3$, $p < 0.05$). Using the *mrc1a* promoter, transient zebrafish transgenic larvae with mosaic expression of RIT1 p.Met90Ile in the venous and lymphatic endothelium resulted in pericardial edema in 69% (68/98) of larvae, compared to 6% (8/139) of uninjected controls ($p < 0.0001$). Additionally, about 20% (20/98) cystic malformations of the caudal plexus and 67% (66/98) had disorganized vasculature compared to 0% (0/139) and 2% (3/139) of uninjected controls, respectively ($p < 0.0001$; $p < 0.0001$).
In summary, we present the lymphatic phenotype in six individuals due to *RIT1*, including a novel mosaic variant. Our functional models will allow for further investigation of the cellular and molecular mechanisms of lymphatic dysplasia due to *RIT1* activation and facilitate therapeutic development.

Session Title: Mendelian Phenotypes Poster Session II

PB4768 Gain-of-function *MYCN* causes a megalencephaly-polydactyly syndrome manifesting mirror phenotypes of Feingold syndrome

Authors:

Y. Nishio[1,2], K. Kato[1,2], F. Tran Mau-Them[3,4], H. Futagawa[5], C. Quélin[6], S. Masuda[7], A. Vitobello[3,4], S. Otsuji[1], H. H. Shawki[8], H. Oishi[8], C. Thauvin-Robinet[3,4,9], T. Takenouchi[10], K. Kosaki[11], Y. Takahashi[2], S. Saitoh[1]; [1]Dept. of Pediatrics and Neonatology, Nagoya City Univ. Graduate Sch. of Med. Sci., Nagoya, Japan, [2]Dept. of Pediatrics, Nagoya Univ. Graduate Sch. of Med., Nagoya, Japan, [3]Unité Fonctionnelle 6254 d'Innovation en Diagnostique Génomique des Maladies Rares, Pôle de Biologie, CHU Dijon Bourgogne, Dijon, France, [4]INSERM UMR1231 GAD, Dijon, France, [5]Dept. of Clinical Genetics, Tokyo Metropolitan Children Med. Ctr., Fuchu, Japan, [6]Service de Génétique Clinique, CLAD Ouest, CHU Rennes, Hôpital Sud, Rennes, France, [7]Dept. of Hematology and Oncology, Tokyo Metropolitan Children's Med. Ctr., Fuchu, Japan, [8]Dept. of Comparative and Experimental Med., Nagoya City Univ. Graduate Sch. of Med. Sci., Nagoya, Japan, [9]Ctr. de Référence Maladies Rares "Anomalies du développement et syndromes malformatifs", Ctr. de Génétique, FHU TRANSLAD et Inst. GIMI, CHU Dijon Bourgogne, Dijon, France, [10]Dept. of Pediatrics, Keio Univ. Sch. of Med., Tokyo, Japan, [11]Ctr. for Med. Genetics, Keio Univ. Sch. of Med., Tokyo, Japan

Background: *MYCN*, a member of the MYC proto-oncogene family, regulates cell growth and proliferation. Somatic mutations of *MYCN* are identified in various tumors, and germline loss-of-function variants are responsible for Feingold syndrome, characterized by microcephaly. In contrast, only one megalencephalic patient with a gain-of-function variant in *MYCN*, p.Thr58Met, has been reported, and additional patients and pathophysiological analysis are required to establish the disease entity. **Results:** We identified two unrelated megalencephalic patients with polydactyly harboring *MYCN* variants of p.Pro60Leu and Thr58Met. Functional analyses for MYCN-Pro60Leu and MYCN-Thr58Met revealed decreased phosphorylation at Thr58 which reduced protein degradation mediated by FBXW7 ubiquitin ligase. The gain-of-function mouse model recapitulated the human phenotypes of megalencephaly and polydactyly, while brain analyses revealed excess proliferation of intermediate neural precursors during neurogenesis, which we determined to be the pathomechanism underlying megalencephaly. Interestingly, the kidney and female reproductive tract exhibited overt morphological anomalies, possibly as a result of excess proliferation during organogenesis. **Conclusions:** We established a *MYCN* gain-of-function-induced megalencephaly-polydactyly syndrome, which showed a mirror phenotype of Feingold syndrome, and revealed that *MYCN* plays a crucial proliferative role, not only in the context of tumorigenesis, but also organogenesis.

Session Title: Mendelian Phenotypes Poster Session III

PB4769 Gene expression changes in the ossicles and long bones of the osteogenesis imperfecta mouse (*oim/oim)*: an RNAseq pilot study

Authors:

E. Carter, K. Seide, K. Brookler, C. Raggio, U. Ayturk; Hosp. for Special Surgery, New York, NY

Osteogenesis imperfecta (OI) is a group of genetic connective tissue disorders commonly caused by dominant type 1 collagen gene variants. Common findings include fractures, bone fragility, and hearing loss. Hearing loss is reported in 50-92% of people with OI, but the mechanism is not well understood. This pilot study evaluated mice with OI (*oim/oim*) for gene expression changes that might be linked to the pathogenesis of hearing loss. Identification of cell types and transcripts associated with hearing loss may suggest novel targets for hearing loss therapies in OI. This IACUC-approved study used RNA-seq on ossicles and femora isolated from 12- and 25-week-old *oim/oim, oim*/+, and wild type (WT) mice. Genotypes were detected with PCR and Sanger sequencing, later confirmed using RNA-seq data. Differential gene expression analysis was performed to detect the effects of age and genotype in each tissue type. Principal component analysis identified clear separation by tissue type, but not genotype or age. Yet, multiple age-related changes in pathways previously associated with bone formation and OI pathogenesis were identified, including those associated with *Runx2*-mediated osteoblast differentiation (FDR<0.1, except for female ossicles (FDR=0.22)) and collagen processing (males: FDR<0.1; females FDR=0.12). Genotype-related changes included elevation of osteoblast-associated transcripts (e.g. *Bglap, Col1a1*) in OI compared to wild-type controls. Male OI femurs contained more osteoblast-associated transcripts than WT femurs at 12 weeks; however, these differences diminished with age. *Runx2*-associated genes were significantly altered in female femurs (FDR<0.1), and ossicles from male and female *oim/oim* mice showed altered transcripts associated with endoplasmic reticulum and golgi activity (FDR<0.1). Transcriptional changes associated with *TFAP2A*-signaling due to genotype (FDR<0.1) were identified in both sexes. Inner ear and long bones from *oim* mice exhibit different transcriptomes, likely representing the differences in their cellular content, and age had a significant effect on the expression of osteogenic transcripts on all genotypes with the exception of male heterozygotes. This finding underscores the importance of evaluating for sexual dimorphism and aging in translational studies. RNA-seq confirmed the significance of previously identified biological processes in OI pathogenesis, and suggests a novel pathway, *TFAP2A*-signaling, that has previously been associated with temporal bone anomalies in patients with craniofacial disorders and specification of inner ear neurons.

Session Title: Mendelian Phenotypes Poster Session I

PB4770 † Generation and characterization of a novel mouse model for Free Sialic Acid Storage Disorder.

Authors:

M. Hackbarth[1,2], J. D. Burke[2], G. Elliott[2], L. Garrett[2], P. Leoyklang[2], P. Zerfas[2], D. Springer[2], M. Sabir[2], L. Pollard[3], M. Huizing[2], W. Gahl[2], M. Malicdan[2]; [1]Univ. of Maryland Sch. of Med., Baltimore, MD, [2]NIH, Bethesda, MD, [3]Greenwood Genetic Ctr., Greenwood, SC

Background: Free Sialic Acid Storage Disorder (FSASD) is an autosomal recessive lysosomal storage disorder due to biallelic mutations in *SLC17A5* and resulting in sialin transporter deficiency and subsequent lysosomal accumulation of free sialic acid in tissues and biological samples. FSASD manifests as a spectrum of clinical severity, ranging from mild developmental delay with progressive neurodegeneration, to severe infantile disease with fatal outcomes. Given the lack of approved therapies, the development of a reliable animal model is crucial for advancing preclinical translational studies.
Methods: Utilizing CRISPR/Cpf1 genome editing, we generated the *Slc17a5*^{R39C} mouse model carrying the most prevalent *SLC17A5* variant observed in FSASD patients. Extensive phenotyping was performed through biochemical, molecular, behavioral, and histological assays.
Results: *Slc17a5*^{R39C} mice displayed Mendelian distribution at birth, but homozygous animals exhibited increased perinatal lethality, with 25% dying within four weeks. Progressive neurodegeneration was evident, characterized by gait abnormalities, tremors, and discoordination. Neurological evaluations revealed significant ataxia and impaired motor performance compared to wildtype littermates. A significant proportion of mutant mice exhibited seizures. Additionally, urinary analysis demonstrated a five-fold increase in free sialic acid levels, establishing that the *Slc17a5*^{R39C} mouse is a bona fide model of the disease. Histological examination revealed hypomyelination in the corpus callosum, cerebellum, and spinal cord tracts, along with a reduction in Purkinje neurons in the cerebellum. Ongoing investigations aim to elucidate the spatiotemporal progression of these findings.
Conclusion: The *Slc17a5*^{R39C} mouse model faithfully replicates the biochemical and neurological phenotypes observed in FSASD. This model represents a valuable tool for investigating the underlying molecular mechanisms and exploring potential therapeutic interventions, particularly regarding neurodevelopmental aspects. With its relevance to human FSASD, the *Slc17a5*^{R39C} mouse provides a reliable mammalian model for future translational studies. This model holds promise for identifying longitudinal biomarkers and determining optimal treatment windows, thus enabling the development of molecular and pharmacological interventions to slow or halt disease progression.

Session Title: Mendelian Phenotypes Poster Session II

PB4771 † Generation and characterization of novel *Clec16a* knockout (*Clec16a*$^{\Delta UBC}$mCAT$^{+/+}$) mouse model where catalase overexpression is targeted to mitochondria (mCAT)

Authors:

M. Bakay[1], B. P. Strenkowski[1], R. Morrow[2], R. Pandey[1], A. Angelin[2], D. Wallace[2,3], H. Hakonarson[1,3]; [1]The Ctr. for Applied Genomics, The Children's Hosp. of Philadelphia, Philadelphia, PA, [2]The Ctr. for Mitochondrial and Epigenomic Med., The Children's Hosp. of Philadelphia, Philadelphia, PA, [3]Dept. of Pediatrics, the Perelman Sch. of Med., Univ. of Pennsylvania, Philadelphia, PA

GWAS have shown consistent association of the *CLEC16A* SNPs with several autoimmune diseases and most recently Parkinson disease. While the exact role of CLEC16A in health and disease is still being elucidated the protein plays a critical role in the regulation of autophagy, mitophagy, endocytosis, and intracellular trafficking. To investigate the CLEC16A role on a whole organism level we have created a *Clec16a* whole-body, inducible-knockout mouse model (*Clec16a*$^{\Delta UBC}$). These global *Clec16a* KO mice exhibit dysregulated mitophagy/autophagy and have a complex phenotype with marked immune dysfunction, rapidly progressing sensory neurodegeneration and severe lipodystrophy. Herein, considering 1) the role of CLEC16A in regulating mitochondrial health through mitophagy and 2) several positive effects in mCAT mice as a result of enhancing antioxidant activity of mitochondria, we investigated if genetic interventions rescue the *Clec16a*$^{\Delta UBC}$ phenotype by improving mitochondrial function. We generated a novel stain by crossing *UBC-Cre-ER*^{T2}*-Clec16a*$^{loxP/loxP}$ with mCAT mice to continue exploration of *Clec16a*$^{\Delta UBC}$ for loss-of-function studies. Ten-week-old mice were treated with tamoxifen for four consecutive days to create experimental *Clec16a*$^{\Delta UBC}$ ± mCAT$^{+/+}$ groups. We established a novel mouse strain and confirmed the *Clec16a* deletion and mCAT overexpression on DNA, RNA, and protein levels. A pilot experiment using 48 mice of both sexes revealed that catalase overexpression in mitochondria (mCAT) partially rescued the lipodystrophic phenotype by significantly slowing the progressive weight loss. Molecular studies are under way to evaluate if ubiquitous catalase overexpression targeted to mitochondria has positive effects in *Clec16a*$^{\Delta UBC}$mCAT$^{+/+}$ mice on neuronal and autoimmune phenotypes and how it attenuates alterations previously observed in mitophagy-related proteins, endoplasmic reticulum (ER) stress, and oxidative phosphorylation (OXPHOS). mCAT partially rescued the lipodystrophic phenotype in the novel mouse strain by significantly slowing the progressive weight loss. We propose to explore the use of mCAT overexpression to correct the phenotype arising from CLEC16A loss-of-function. Results of this study are likely to inform our understanding of how to improve mitochondrial health withstands dysregulated mitophagy. We hypothesize that using the novel *Clec16a*$^{\Delta UBC}$mCAT$^{+/+}$ model for testing drugs with modulatory effects on mitophagy, aiming at compensating for the attenuated CLEC16A activity, will be more successful and will present feasible candidates for future targeted interventions in individuals with *CLEC16A* risk variants.

Session Title: Mendelian Phenotypes Poster Session III

PB4772 Genetic analysis of Bardet-Biedl Syndrome: unraveling novel mutations and insights from consanguineous families

Authors:

N. Kakar[1], A. Gul[1], F. Rehman[2], N. Wasif[3]; [1]Dept. of Biotechnology, BUITEMS, Quetta, Pakistan, [2]Dept. of Pathology, Bolan Med. Coll., Quetta , Pakistan, Quetta, Pakistan, [3]Inst. of Human Genetics, Ulm Univ., Ulm, Germany, ulm, Germany

Bardet-Biedl Syndrome (BBS) is a rare monogenic disorder characterized by a spectrum of clinical features, including retinitis pigmentosa, polydactyly, obesity, renal anomalies, and intellectual disabilities. To date, mutations in 26 genes are known to cause BBS Syndrome. The aim of this study was to identify the underlying genetic cause and clinical manifestations of BBS syndrome in four consanguineous families segregating homozygous mutation. we performed exome sequencing in four highly consanguineous families from Pakistan affected with BBS syndrome. We were able to identify novel mutations in different BBS genes including a homozygous 1bp deletion (c.775delA) in MKKS (BBS6), resulting in a frameshift variant (p.Thr259Leufs*21), a duplication mutation in BBS5 was identified as c.159_180dup, causing a frameshift variant (p.Ser61Lysfs*32). Furthermore, a novel missense mutation, c.2014G>A (p.A672T), was detected in the BBS12 gene in a consanguineous family. Lastly, we report a homozygous mutation, c.183_184delTT, in the BBS5 gene. Clinical examination of the affected individuals of all four consanguineous families reveals phenotypic variability, highlighting the complex nature of BBS syndrome. We have thus revealed mutation across BBS genes including BBS5, BBS6, and BBS12.In summary, our study provides comprehensive insights into the clinical spectrum of BBS, and novel mutations associated with Bardet-Biedl Syndrome contribute to understanding the pathophysiology of BBS.

Session Title: Mendelian Phenotypes Poster Session I

PB4773 Genetic Analysis of Segawa Disease Patients in Japan

Authors:

H. Fujioka[1,2], H. Shintaku[2], N. Nakano[2], T. Sakaguchi[2], T. Hamazaki[2]; [1]Kansai Univ. of Welfare Sci., Kashiwara, Japan, [2]Dept. of Pediatrics, Graduate Sch. of Med., Osaka Metropolitan Univ., Osaka, Japan

Background: Segawa disease is a rare autosomal dominant inherited disorder characterized by dopa-responsive dystonia (DRD) with diurnal fluctuation. Patients exhibit mutations in the GCH1 gene, which codes GTP cyclohydrolase I (GTPCHI). Dopamine production in the central nervous system is diminished due to a partial shortage of tetrahydrobiopterin, a co-factor of amino acid decarboxylase, which catalyzes L-dopa into dopamine. No common mutation has been reported. Some mutations are detected by gene sequences, while some are large exon deletions detected using the MLPA method.Methods: Patients exhibiting symptoms of DRD or other related involuntary movements who asked to join our study underwent measurement of biopterin and neopterin concentration in their cerebral spinal fluid and GCH1 gene analysisfrom January 2008 to December 2022. Results and Discussion: A total of 228 patients participated, of whom 68 patients (29.8%) from 53 families exhibited mutations in their GCH1 gene. Among them, 14 patients had point mutations; 20 patients from 14 families, frame-shift mutations; 14 patients from 11 families, mutations in their introns; 13 patients from 10 families, large exon deletions; 3 patients from 2 families, nonsense mutations; and 3 patients, multiple mutations in their alleles. One patient who was negative for GCH1 analysis exhibited some mutations in the PARK2 gene in a study at another institute. In our study, mutations in the GCH1 gene were detected in 29.8% of participants. Among families with mutations in the GCH1 gene, 25.9% were point mutations. The same number of families had frame-shift mutations. Intron mutations were found in 20.7% of all families. Large deletions were also found in 18.8% of families. Nonsense mutations and compound-hetero-mutations were found in two families each (3%). Conclusion: We analyzed the GCH1 gene and detected mutations in 53 families. No common mutation was detected. Many participants exhibited DRD symptoms at a young age.

Session Title: Mendelian Phenotypes Poster Session II

PB4774 Genetic background of hereditary spastic paraplegias in the Czech Republic

Authors:

A. Meszarosova[1], P. Seeman[2], I. Soldatova[2], M. Bittoova[2], P. Lassuthova[1], D. Safka Brozkova[1]; [1]Neurogenetic laboratory, Dept. of Pediatric Neurology, Second Faculty of Med., Charles Univ. and Univ. Hosp. Motol, Prague, Czech Republic, [2]GENNET, Ctr. of Clinical Genetics, Prague, Czech Republic

Introduction: Hereditary spastic paraplegia (HSP) is a rare highly heterogeneous disabling disorder manifesting with a progressive gait impairment. The disorder is caused by pathogenic variants in almost 100 known genes. We comprehensively mapped the genetic background of HSP in the Czech Republic since, until now, molecular-genetic data related to this European region has not been available. **Patients and Methods:** A total of 647 patients with a suspected diagnosis of HSP were examined using various methods. In 380 patients, only a separate examination of the *SPAST* gene (Sanger sequencing and MLPA) was performed. A total of 270 patients (predominantly pre-excluded SPG4) underwent NGS sequencing. Among these patients, 100 had early disorder onset (< 20 years of age) and the remaining 170 had onset in adulthood. **Results:** Of the total number of 647 examined patients, a genetic diagnosis was found in 171 examined (clarified 26% of examined patients). A total of 98 patients were diagnosed with SPG4 (4 diagnosed by NGS). Another 73 patients were diagnosed using NGS with other types of SPG. As expected, the most common diagnosis was SPG4 (58% of all clarified). Of the other more frequently occurring types, we detected SPG3 (6/3.5%), SPG7 (9/5.3%), SPG10 (7/4.1%), SPG11 (8/4.7%) and SPG31 (7/4.1%). Five patients were diagnosed with X-linked adrenoleukodystrophy that can manifests clinically as spastic paraparesis. We also diagnosed rare and recently described types of HSP (*KPNA3, HPDL*). In seven cases the de novo occurrence was confirmed; in all cases with early onset. Very rare types of complicated HSP (*KCNJ10, KCNA2, KY, NFU, FAR1*) were detected only in the group with early onset. In three cases, different neurological disorder was finally diagnosed (Wilson disease; Galloway-Mowat syndrome; Wieacker-Wolf syndrome); spastic paraparesis was only the primary manifestation. Segregation analysis and follow-up examination of family members genetically confirmed the diagnosis of HSP in another 62 related individuals. **Discussion:** As expected patients with SPG4 make up the largest group. The inconsistency of the methods and the partial overlap of the examined groups introduces an initial bias; the percentage of SPG4 cases is probably slightly overestimated. In patients with early onset, we emphasize the importance of a segregation analysis and a broader assessment of exome data. According to our results, we estimate the prevalence of the HSP in the Czech population to be approximately 2.5:100,000. All authors declare no conflict of interest. The research meet the criteria of Declaration of Helsinki. Supported by Ministry of Health of the Czech Republic number NU22-04-00097.

Session Title: Mendelian Phenotypes Poster Session III

PB4775 Genetic Blueprint of Congenital Muscular Dystrophies with Brain Malformations

Authors:

K. Flannery[1], S. Safwat[1,2], E. S. Cauley[3], A. Horvath[4], E. P. Hoffman[5], C. Walsh[6], M. A. Salih[7], M. Mokhtar[8], A. El Beheiry[2], E. Abdalla[2], M. Manzini[1]; [1]Dept. of NeuroSci. & Cell Biology, Child Hlth.Inst. of New Jersey, Rutgers-Robert Wood Johnson Med. Sch., New Brunswick, NJ, [2]Dept. of Human Genetics, Med. Res. Inst., Alexandria Univ., Alexandria, Egypt, [3]Dept. of Pharmacology & Physiology, The George Washington Univ., Washington, DC, [4]Dept. of Biochemistry & Molecular Med., The George Washington Univ., Washington, DC, [5]Sch. of Pharmacy & Pharmaceutical Sci., Binghamton Univ., Binghamton, NY, [6]Div. of Genetics & Genomics, Manton Ctr. for Orphan Disease Res., Howard Hughes Med. Inst., Boston Children's Hosp., Boston, MA, [7]Div. of Pediatric Neurology, Dept. of Pediatrics, King Saud Univ. Coll. of Med., Riyadh, Saudi Arabia, [8]Dept. of Radiodiagnosis & Interventional Radiology, Faculty of Med., Alexandria Univ., Alexandria, Egypt

Congenital muscular dystrophies (CMDs) are characterized by infantile hypotonia, motor delay, and progressive muscle atrophy. Severe brain involvement is also observed in subtypes such as the dystroglycanopathies Muscle Eye Brain disease (MEB) and Walker Warburg syndrome (WWS), which have 18 known disease genes, and merosin-deficient CMD (MDC1A), which is caused by mutations in *LAMA2*. Whole exome sequencing (WES) is a remarkable tool for identifying genetic variants that cause CMDs, yet many individuals remain without a genetic diagnosis. In addition, rare variants that do not directly impact protein sequence including intronic, synonymous, and noncoding regulatory variants are often overlooked. Here, we present a cohort of patients from ethnically diverse populations in three groups: one group with WWS (52 individuals from 48 families), one group with MEB (18 individuals from 16 families), and one group with MDC1A (21 individuals from 21 families). Standard analysis of variants that impact protein sequence yielded a genetic diagnosis rate of 72.9% (35/48) for WWS, 50% (8/16) for MEB, and 85.7% (18/21) for MDC1A. Additional analysis of frequently overlooked variant classes yielded several variants of interest that were previously undetected: two in the WWS group, two in the MEB group, and four in the MDC1A group. These variants require experimental validation but could improve genetic diagnosis rates to 77.0% (37/48) for WWS, 62.5% (10/16) for MEB, and 100% (21/21) for MDC1A. To expand these analyses to additional populations, we recruited a second cohort of patients diagnosed with severe CMDs with brain malformations from Egypt (12 patients from 11 families) and achieved a genetic diagnosis rate of 85.7% (6/7) for dystroglycanopathies and 100% (4/4) for MDC1A. Additionally, a regulatory variant was predicted in the remaining unsolved dystroglycanopathy case, resulting in a potential genetic diagnosis for all patients. Our findings indicate that disease-causing noncoding variants can often be detected through WES, which may be sufficient on its own to genetically diagnose all MDC1A cases. However, other strategies such as whole genome sequencing and RNA-seq are likely necessary to provide a genetic diagnosis for all dystroglycanopathy cases.

Session Title: Mendelian Phenotypes Poster Session I

PB4776 Genetic diagnosis of different dyslipidemia in Iran.

Authors:

B. Rabbani[1], N. Mahdieh[1,2], M. Agli-Moghadam[2], S. Esmaeli[1], S. Nematolahi[1], L. Hejazi[2]; [1]Growth and Dev. Res. Ctr., Tehran, Iran, Islamic Republic of, [2]Rajaie cardiovascular research center, Tehran, Iran, Islamic Republic of

Dyslipidemias affecting about 1 in 600 worldwide are defined by abnormal lipid profile. Affected individuals may show different phenotypes from being asymptomatic to milky plasma, vague abdominal pain, lipemia retinalis, nausea, vomiting, eruptive xanthomas, hepatosplenomegaly and pancreatitis. In addition to lifestyle and eating habits, genetic factors play a leading role for developing hyperlipidemias. These genetic agents are still unknown in Iran as the crossroad of the Middle East. In this study, a genetic investigation is described to unravel the involving genes in this heterogeneous population.Clinical features of the studied patients were documented. Family history and pedigrees were taken. In the first step, the following genes were sequenced *APOA5, APOC2, LDLR, APOA1, APOE, LDLRAP1 (ARH)* and *GPIHBP1*. Then, negative patients were subjected to WES to identify their causing variants.Clinical features, family pedigrees and genetic data of 63 patients were analyzed in this study. *LDLR* variant was the most common cause accounting for 8 patients. Two novel variants were identified in *LDLRAP1* and *APOA5* genes. The e2 and e4 alleles of *APOE* gene were found in 7 patients. We are analyzing the data of WES sequencing.Due to the large number of affected individuals in Iran, there should be a national planning for genetic study throughout the country since this study is the starting point. Genetic studies help for family planning and management of lifestyle and genetic counseling could decrease the burden of dyslipidemia in families.

Session Title: Mendelian Phenotypes Poster Session II

PB4777 Genetic disorders associated with dental problems in a pediatric orthodontic population

Authors:

E. Severin; Carol Davila Univ Med Pharm, Bucharest, Romania

Background: Dental anomalies associated with genetic disorders are conditions characterized by deviations from normal development and eruption of teeth and adjacent tissues, causing both morphological and functional or aesthetic changes, in both primary and permanent dentition. These anomalies can have a significant impact on the oral and general health of children, hence early evaluation and proper treatment are essential in preventing complications and additional costs.**Study aim:** The study based on our experience and a systematic review aimed to study the impact of genetic makeup on oral health in general and pediatric orthodontic population.**Materials and Methods:** The study population consisted of 526 pediatric patients between 6 and 14 years of age who presented with dental anomalies and were referred to the dental clinic. Clinical and radiographical assessments were performed, followed by detailed investigation into the personal and family medical history of the orthodontic patient.Patients with dental anomalies suggestive of certain genetic disorders underwent genetic testing, which included karyotyping or targeted gene panel sequencing. Informed consent was obtained from the guardians of the pediatric patients.**Results:** The prevalence of dental anomalies observed in this study is higher than that of the general population. Various oro-dental abnormalities were noted among orthodontic patients. The most common abnormality was hypodontia followed by microdontia, and supernumerary teeth, as well as functional abnormalities such as malocclusion, especially in the permanent dentition. The highest number of dental abnormalities was observed among patients with Down syndrome. Oro-dental anomalies (determined by genetic background) were the cause of oral health problems.**Conclusions:** Evaluating the prevalence of dental anomalies among children and diagnosing them in an early stage can help prevent complications and additional costs by ensuring prompt and adequate treatment. Therefore, it is crucial for pediatricians to anticipate the development of syndromic dental anomalies and refer patients to appropriate dental services. Performing a preventive dental examination in children aged 7 can aid in early detection and treatment of dental anomalies, ultimately improving the overall oral and general health of syndromic patients.

Session Title: Mendelian Phenotypes Poster Session III

PB4778 Genetic landscape of developmental and epileptic encephalopathy in Mali

Authors:

S. Bamba[1], M. Dembele[1], W. Ji[2], S. Diarra[1], A. Yalcouyé[3], L. Cisse[4], A. Maiga[5], L. Jeffries[2], S. Lakhani[2], G. Landoure[6], H3AFRICA CONSORTIUM; [1]USTTB, Bamako, Mali, [2]Yale Univ., New Haven, CT, [3]USTTB, Mali, Bamako, Mali, [4]Teaching Hosp. of Point G, Bamako, Mali, [5]Univ. of Sci., Techniques and Technologie, Bamako, Mali, [6]CHU du Point G, Bamako, Mali

Background Developmental and epileptic encephalopathy (DEE) is a severe neurological disorder characterized by seizures, developmental delay, and progressive cognitive impairment. Its genetic basis remains largely unelucidated. Recent advances in genetic technologies have provided new opportunities for gene identification. **Objectives** We studied three families from Mali sharing phenotypes of DEE, caused by novel variants in the *GRIN1*, *SYNJ1* and *RARS2* genes. **Methods** Patients were examined by neurologists and DNA was obtained from whole exome sequencing (WES). Brain imaging was performed to exclude structural causes and refine the phenotype description. In addition, several bioinformatics tools were used to assess the deleteriousness of the variants, and putative variants were confirmed with Sanger sequencing in all available family members. **Results** Five patients from three unrelated families were enrolled. The mean age at diagnosis was 2.6 years (ranging from 25 days to 7 years) and symptoms started during the neonatal period with convulsive seizures and myoclonus that became refractory to antiepileptic drugs. The first patient was a boy, 25-month-old from a consanguineous family, born from a normal pregnancy and an uneventful delivery was referred for uncontrolled neonatal seizures. Symptoms started on day 4 after birth with myoclonic seizures, spasms in flexion of the upper limbs, and tonic seizures involving the four limbs. Two affected individuals (one male and one female) aged 7 and 2 years, from healthy consanguineous parents were referred for motor acquisition delay. The disease started in the first 10 days after birth with generalized tonic-clonic seizures. Neurological examination showed delayed motor acquisition, hypotonia, and general muscle amyotrophy in the four limbs and a multidirectional nystagmus in the older patient with bilateral Babinski sign. Two other affected siblings from non-consanguineous healthy parents. were seen for psychomotor delay and visual loss. Symptoms started with tonic and clonic seizures. Neurological examination found a psychomotor acquisition delay, visual loss, and clonic seizures in both patients. WES identified novel variants homozygous (p.Leu589Pro) in *GRIN1*, and *(*p.*Arg380Gln) in SYNJ1* and compound heterozygous (p.His141Arg and p.Ile150Thr) in *RARS2* genes. Several predictor tools predict these variants to be pathogenic or likely pathogenic. **Conclusion**
This study revealed novel variants in known genes associated with severe forms of DEE. This is the first report of molecularly diagnosed cases in Mali and contributes to providing further evidence of the implication of these genes.

Session Title: Mendelian Phenotypes Poster Session I

PB4779 Genetic modifiers of *FIG4* pathology

Authors:

G. Lenk[1], x. cao[2], V. Mikusevic[3], J. A. Mindell[3], M. Meisler[4]; [1]Univ. of Michigan, Ann Arbor, MI, [2]Univ. of Michigan Hlth.System, Ann Arbor, MI, [3]NINDS, Bethesda, MD, [4]Univ Michigan, Ann Arbor, MI

Loss-of-function mutations of *FIG4* are responsible for neurological disorders in human and mouse. Loss-of-function of *FIG4* reduces the abundance of the signaling lipid $PI(3,5)P_2$, which regulates the activity of several lysosomal ion channels and transporters We investigated two functionally related genes for their potential to ameliorate *FIG4* pathology: *PIP4K2C*, another $PI(3,5)P_2$ biosynthetic gene, and *CLCN7*, a lysosomal transporter that is inhibited by $PI(3,5)P_2$.

Loss-of-function mutations of the phosphoinositide kinase *PIP4K2C* result in elevated abundance of $PI(3,5)P_2$. This suggested that we might be able to compensate for deficiency of *FIG4* by **reduced** expression of *PIP4K2C*. To test this hypothesis in a whole animal model, we crossed *Fig4*$^{-/-}$ mice with *Pip4k2c*$^{-/-}$ mice. The neonatal lethality of *Fig4* null mice on the C57BL/6J strain background was rescued by reduced expression of *Pip4k2c*. In addition, the enlarged lysosomes characteristic of *Fig4* null cells was reduced by heterozygous loss of *Pip4k2c*.

A gain of function mutation of *CLCN7* encoding ClC-7 results in enlarged lysosomes and phenocopies *Fig4*$^{-/-}$ mice. We therefore tested the ability of **reduced** *CLCN7* expression to compensate for loss of *FIG4*. In *FIG4* null HAP1 cells, knock-out of *CLCN7* corrected lysosomal swelling and partially corrected lysosomal hyperacidification. In the *Fig4* null mouse, reduction of ClC-7 by expression of the dominant negative *CLCN7* variant p.Gly215Arg improved growth and neurological function and increased lifespan by 20%.

These data indicate that reducing expression of *CLCN7* or *PIP4K2C* provides potential targets for treatment of *FIG4* deficiencies such as Charcot-Marie-Tooth Disease Type 4J.

Session Title: Mendelian Phenotypes Poster Session II

PB4780 Genetic profile of early onset Alzheimer's disease and unspecified dementia based on rare variants from pooled whole genome sequencing data

Authors:

S. Karachanak-Yankova[1,2], D. Serbezov[1], L. Balabanski[3], D. Belejanska[1], S. Mehrabian[1], M. Petrova[1], L. Traykov[1], D. Toncheva[1]; [1]Med. Univ.-Sofia, Bulgaria, Sofia, Bulgaria, [2]Sofia Univ., St. Kliment Ohridski, Sofia, Bulgaria, [3]CellGenetics Lab., Sofia, Bulgaria

Alzheimer's disease has been studied in many genetic and genomic analyses. Early onset Alzheimer's disease (EOAD) accounts for less than 5% of Alzheimer's disease cases, but it can be due to mutations in single genes established in a small fraction of families. On the other hand, unspecified dementia has been underrepresented in genetic research. In order to broaden the understanding of the single gene causes of these two forms of dementia, we have used the cost and time effective approach of whole genome sequencing of pooled DNA samples. Since whole genome sequencing data can not be used to infer individual haplotypes we have focused on rare variants which can be highly pathogenic on their own. We have collected blood DNA samples from 49 EOAD and 50 unspecified dementia patients. Equal amounts of DNA were used to construct pooled DNA samples, which were sent for 100 x whole-genome sequencing. Reads were aligned to the reference genome (GRCh37/hg19) and variants were annotated using wANNOVAR. We extracted all single nucleotide variants with rs numbers called by gnomAD. Among these we have chosen only those that have MAF < 0,001 in gnomAD non-Finish Europeans. The EOAD pool contained 5 151 635 variants, whereas the unspecified dementia pool 5 097 536 variants, among which 156 and 195 variants respectively have rs numbers and have MAF < 0,001 in gnomAD. None of these variants is found in EOAD related genes. For most of the detected variants there is insufficient available literature. Still, the interesting finding is the presence of rare variants in genes related with disorders of the brain. Thus, among EOAD patients there are variants in: *ACOT7* -regulating neuronal fatty acid metabolism; *ARSB* - progressive neurodegeneration related; *ATP8A1* -phospholipid transporter; *BCAT1*- major contributor to brain glutamate production; *KCNJ6*- in 'Down syndrome critical region'; *MACROD2*- autism susceptibility; *TNR* -neural extracellular matrix protein. The unspecified dementia pool contained variants in: *AMD1* - involved in polyamine biosynthesis; *ATRX* -chromatin remodeling; *CIAO1* - iron-sulfur cluster assembly; *CNTN4* - Ig cell adhesion molecule;*CSMD1* - complement control protein; *FAM151B* - involved in dendritic branching; *FAR1* - for reduction of fatty acids to fatty alcohol; *FRMPD4* - leading to intellectual developmental disorder; *TRPC5* - from transient receptor family. These findings point to the fact that neurological and psychiatric disorders can show overlapping etiology and that certain genetic variants can present with pleiotropic effect. Acknowledgment: The work is funded by the Bulgarian National Science Fund, project KP-06-N33/5 from 13.12.2019.

Session Title: Mendelian Phenotypes Poster Session III

PB4781 Genetic Underpinnings of Non Syndromic High Myopia in South Korea: Focus on Early Onset High Myopia

Authors:

Y. Seo[1], S. Choi[2], S. Kim[3], J. Han[4]; [1]Dept. of ophthalmology, Yonsei Coll. of medicine, Yongin, Korea, Republic of, [2]Eyereum Eye Clinic, Seoul, Korea, Republic of, [3]Inst. of Vision Res., Dept. of Ophthalmology, Severance Hosp., Yonsei Univ. Coll. of Med., Seoul, Korea, Republic of, [4]Inst. of Vision Res., Dept. of Ophthalmology, Gangnam Severance Hosp., Yonsei Univ. Coll. of Med., Seoul, Korea, Republic of

High myopia significantly impacts patients during their productive years. In South Korea, the prevalence of myopia (96.5%) and high myopia (20.6%) is notably high. Our study investigated the clinical and genetic aspects of non-syndromic high myopia in a South Korean cohort. The study involved 36 individuals from 27 families, with mean age 27.84±21.99 years. Criteria included a refractive error of less than -9.0D or axial length over 28mm, and early onset high myopia (EoHM), defined as a refractive error of less than -6.0D before age 10. Among the probands, 25% were EoHM. Females constituted 63.3% of the affected individuals. Genetic causes were identified in 11% of patients, including genes associated with Stickler syndrome (*COL2A1* c.1772del:p.(Gly591ValfsTer38)), pseudoxanthoma elasticum (*ABCC6* c.2542del: p.(Met848CysfsTer83)), and retinitis pigmentosa (*RP1* c.6181del : p.(Ile2061SerfsTer12)). All genetically solved cases were EoHM, related to syndromic diseases or inherited retinal dystrophies, and none were found in non-EoHM probands. These findings emphasize the role of genetic analysis in EoHM for early detection of systemic diseases or inherited retinal dystrophies, leading to appropriate genetic counseling.

Session Title: Mendelian Phenotypes Poster Session I

PB4782 Genome sequencing from 50 autism trios identified novel candidate genes with mixed inheritance pattern in the Qatari population

Authors:

V. Gupta, A. Ben-Mahmoud, F. Alshaban, H-G. Kim; Neurological Disorders Res. Ctr., Qatar BioMed. Res. Inst., Hamad Bin Khalifa Univ., Qatar Fndn., Doha, Qatar

The advent of next-generation sequencing (NGS) has revolutionized human genetics in rare genetic disorders, improving the diagnostic yield and reducing the turnaround time. While exome sequencing is commonly used for molecular diagnosis, genome sequencing (GS) has emerged as a more powerful alternative due to its extensive coverage. Despite challenges related to deep reads and managing large dataset, GS has successfully identified genetic variants associated with autism spectrum disorder (ASD). To address the underrepresentation of genetic studies on autism in the Middle East, we conducted GS on 50 trios of independent simplex autism families from Qatar with unknown etiology. They included 14 consanguineous and 36 non-consanguineous simplex families, consisting of 41 males and 9 females. Our analysis revealed diverse variants, including *de novo*, homozygous, X-linked, and compound heterozygous variants, totalling 42 missense, 2 frameshift, 2 nonsense, 1 splicing, 1 in-frame 3 nucleotide deletion,1 start loss, and 1 >200 CGG repeats of FMR-1 variants. They were found in 30 known and 20 novel candidate genes, providing an aetiologic diagnosis. The pathogenicity of these variants was substantiated by high CADD scores, their low frequency in gnomAD, reported sporadic variants, and direct physical interaction with known ASD genes. Of the participants, comprising 41 males and 9 females, 40 individuals had syndromic and 10 had non-syndromic ASD. The cohort comprised individuals from various backgrounds, including 33 Qatari (66%), 5 Syrian (10%), 2 Egyptian (4%), 2 Yemeni (4%), 2 Sudanese (4%), 1 Saudi Arabian (2%), 1 Algerian (2%), 1 Jordanian (2%), 1 Indian (2%), 1 Tunisian (2%), and 1 Palestinian (2%).The most common phenotypes observed among the probands included autism, intellectual disability, epilepsy, developmental disorders, and language/speech delay. Our preliminary findings illustrate the mixed inheritance patterns in autism families in Qatar, characterized by a high rate of consanguinity (52%). Additionally, our results highlight the association of ASD with fundamental cellular processes such as ion transport pathway, ubiquitin pathway, neuron migration, and transcription activity, which are crucial for normal cognitive development and function.

Session Title: Mendelian Phenotypes Poster Session II

PB4783 † Genome sequencing identifies second molecular diagnoses in individuals with Phelan-McDermid syndrome.

Authors:

R. Gore[1], R. Duncan[2], S. Byers[1], C. Jodarski[1], N. Reynolds-Lallement[1], M. Kamen[1], J. Yan[1], K. Lewis[1], D. Zeltser[2], B. Seifert[1], R. Ghosh[1], M. Similuk[1], M. Walkiewicz[1], A. Thurm[2]; [1]Natl. Inst. of Allergy and Infectious Diseases, NIH, Bethesda, MD, [2]Natl. Inst. of Mental Hlth., NIH, Bethesda, MD

Introduction: Phelan-McDermid syndrome (PMS) is a genetic condition caused by deletions of chromosome 22q13.3 and pathogenic variants in the *SHANK3* gene. Features typically include intellectual disability, hypotonia, and absent speech, though there is considerable variability even among individuals with the same molecular finding. Our study aimed to determine the extent to which additional Mendelian molecular diagnoses exist, with exploratory questions about contribution to phenotypic variability in a cohort of patients with PMS.
Methods: Individuals with a prior diagnosis of PMS were recruited from a larger natural history study and offered genetic counseling, genome sequencing, and chromosomal microarray. Samples were collected via blood and saliva, and data was analyzed via phenotype informed and agnostic approaches. Results were CLIA-validated with Sanger analysis and returned to participating families by a genetic counselor.
Results: Fifteen individuals underwent comprehensive genetic testing. This group was 53% female, 80% white, and ranged in age from 7-48 years (mean=23). Sixty percent had a known 22q.13.3 deletion including *SHANK3* and 40% had a pathogenic variant in *SHANK3*. Genome sequencing identified a second molecular diagnosis in 7/15 participants (46.7%). Overall, 4/7 of these additional molecular diagnoses were related to neurological phenotypes, including: Charcot-Marie-Tooth disease, type4B3 (*SBF1* NM_002972.4:c.5507C>G; hemizygous due to 22q deletion), spinal muscular atrophy, lower extremity-predominant, 2A, autosomal dominant (SMALED2A) (*BICD2* NM_001003800.2:c.2108C>T), spastic paraplegia 7 (*SPG7* NM_003119.4:c.861dup and c.1045G>A), and 16p11.2 deletion syndrome. We also identified a secondary finding in *TTR* (NM_000371.4:c.424G>A) associated with TTR-related amyloidosis, as well as incidental findings in *ATM* (NM_000051.4:c.3137T>C) linked to susceptibility to breast cancer and related to the patient's family history, and in *MPEG1* (NM_001039396.2:c.181C>T) associated with Immunodeficiency 77 and the patient's personal history of recurrent respiratory infections, asthma, and folliculitis.
Discussion: Unexpectedly, nearly half of the patients in our cohort had a second molecular diagnosis on genome sequencing in addition to their known diagnosis of PMS. Our findings suggest that these "second hits" in genes related to neurological phenotypes may be an underappreciated source of variability among individuals with rare neurodevelopmental disorders and may warrant expanded research in larger patient cohorts.

Session Title: Mendelian Phenotypes Poster Session III

PB4784 Genome sequencing of multiplex families affected by neuropsychiatric disorders

Authors:

U. Gurusamy[1], D. C. Koboldt[1], H. Brickman[2], M. Latsko[2], S. Ramadesikan[1], M. Marhabaie[1], E. Varga[1], A. Daley[1], C. Colwell[1], O. Roof[1], M. Norris[2], M. Cermark[2], K. Stephenson[2], K. Sanford[2], N. Russell[2], K. Heck[2], J. Turner[3], M. Artomov[1], S. E. Hickey[2], J. A. Bridge[2], E. Butter[2], M. Fristad[2], D. Axelson[2], E. R. Mardis[1], R. K. Wilson[1]; [1]The Steve and Cindy Rasmussen Inst. for Genomic Med., The Abigail Wexner Res. Inst. at Nationwide Children's Hosp., Columbus, OH, [2]Nationwide Children's Hosp., Columbus, OH, [3]The Ohio State Univ. Med. Ctr., Columbus, OH

Behavioral health disorders represent a major medical challenge in pediatrics, but the genetic risk factors have been understudied. To date, most of our knowledge of the genetic architecture of psychiatric disease comes from large-scale genome-wide association studies (GWAS). However, there is a scarcity of genomic data from families with multiple individuals affected by neuropsychiatric disease. For the past two years, we have conducted an IRB-approved translational research study in which we applied state-of-the-art-genomics — whole-genome sequencing and RNA-seq — to multiplex families affected by neuropsychiatric disorders, including bipolar and psychotic disorders, depression/suicidality, and autism spectrum disorder (ASD). Until now, we have sequenced 211 individuals from 58 families, comprising more than 30 Tbp of sequencing data. All individuals and available family members were comprehensively phenotyped using standard measures by a psychologist and gave informed consent to participate in the research. We analyzed genomic variants within each family to identify rare variants that might contribute to these complex, multifactorial conditions. So far, we have identified phenotypically relevant findings in 16% of families. For example, we enrolled two males with hemizygous *MECP2* variants, one of whom manifested only the psychiatric features associated with X-linked recessive *MECP2* disease. In addition, we identified *de novo* variants in strong susceptibility genes such as *SCN8A, SHANK*3, and *NRXN10* in 29% of families. We also identified rare coding variants in candidate novel susceptibility genes — *NCAN, ZSWIM6, GRIN2B*, and several *CACNA1x* cluster genes — in an additional 41% of families. To further assess the contribution of common DNA variants to disease risks in our cohort, we performed polygenetic risk score analyses for four phenotypes - major depressive disorder, bipolar disorder, attention-deficit/hyperactivity disorder (ADHD), and ASD. Interestingly, families enrolled in this study had overall higher susceptibility to ADHD compared to the 1000 Genomes population. The observed higher ASD susceptibility was specific to affected probands and was not observed among the family members. Our results suggest that while neuropsychiatric disorders that affect children have complex multifactorial etiologies, family-based genomic studies identify medically relevant variants and compelling new candidate genes.

Session Title: Mendelian Phenotypes Poster Session I

PB4785 Genome-Wide Sequencing Implicates Monogenic Disruption of Vascular Integrity in Childhood-Onset Essential Hypertension

Authors:

E. Wonkam Tingang[1], M. Gupta-Malhotra[2], I. Copeland[3], Y. Han[1], A. Jajoo[1], A. Haldipur[1], S. Brown[1], T. Billawala[1], N. Hall[3,4], D. Scott[3,5,6], N. Lie[1,3,4], E. Banfield[1], J. Posey[3], N. Hanchard[1,3]; [1]Childhood Complex Disease Genomics Section, Natl. Human Genome Res. Inst., NIH, Bethesda, MD, [2]Dept. of Cardiology, Baylor Coll. of Med., San Antonio, TX, [3]Dept. of Molecular and Human Genetics, Baylor Coll. of Med., Houston, TX, [4]USDA ARS Children's Nutrition Res. Ctr., Baylor Coll. of Med., Houston, TX, [5]Texas Children's Hosp., Houston, TX, [6]Dept. of Molecular Physiology and Biophysics, Baylor Coll. of Med., Houston, TX

Childhood-onset essential hypertension (COEH) is characterized by elevated blood pressure of unknown etiology diagnosed before age 18. COEH affects ~2.1% of children in the United States, with a higher prevalence among those of self-reported African ancestries. Unlike adult-onset essential hypertension, the genetic underpinnings of COEH are unknown, but its early onset, high heritability, low correlation with environmental factors, and low prevalence, suggest a role for monogenic variation. We examined evidence for Mendelian inheritance in COEH by performing genome sequencing in 49 probands and their relatives. We identified six candidate genes with strong evidence of their involvement in hypertension in ten families (20.4%; 10/49), including *SYNE1* (three families), *HMCN1* (two), *TNS1* (two), *FBN2* (one), *THBS2* (one), and *PCDH7* (one). We also identified 14 genes with moderate evidence for involvement in hypertension in 14 families (28.5%; 14/49), with 25 families remaining unsolved (51%; 25/49). The probands of the families with identified candidates (strong and moderate, N= 24 families) were generally young (mean age = 13.6 +/- 2.9 years) without known cardiovascular disease comorbidities such as high blood cholesterol, diabetes, or obesity (91.7%; 22/24). Implicated variants in almost all candidate genes (91.7%; 22/24) showed evidence of autosomal recessive inheritance, with the resulting candidate genes (90%; 18/20) estimated to be intolerant to homozygous loss of function and predominantly expressed in vessels (80%; 16/20). All identified variants (42 missense, one stopgain, and one splicing) were predicted to be damaging to the resulting protein by multiple in silico tools and were either novel or exceedingly rare (frequency <1%) in population databases. *SYNE1*, *HMCN1*, and *TNS1* have a presumed role in activating Rho-kinase (ROCK), a protein that regulates myosin chains phosphorylation and contractility in vascular smooth muscle cells. We showed that rare, putatively protein-damaging, recessively inherited variants in *SYNE1* are strongly enriched in our cohort and among individuals of African ancestries. Knockdown of *SYNE1* in vascular smooth muscle cells resulted in increased stiffness and contractility that could be rescued by the vasodilator Fasudil (a ROCK inhibitor), consistent with the postulated role for *SYNE1* in regulating vascular tone. Cell-based assays for the other candidates and functional validations using mouse models constitute our next steps. Our results suggest that genes associated with COEH are mainly involved in vascular tone regulation, offering insights to potential therapeutic precision medicine in the future.

Session Title: Mendelian Phenotypes Poster Session II

PB4786 Genomic analyses of 281 consanguineous kindreds from the Middle East and North Africa facilitate the discovery of novel recessive neurodevelopmental rare disease traits

Authors:

R. Duan[1], A. Saad[1,2], D. Calame[1], Z. Dardas[1], J. Fatih[1], P. Kaur[1], H. Du[1], I. Herman[3], T. Mitani[1], H. Elbendary[4], K. Rafat[5], I. Hegazy[6], M. Issa[4], M. Zaki[4], G. Abdel-Salam[4], N. Saadi[7], R. Alshamari[8], M. Ashkanani[9], M. Arabi[10], M. Djuric[11], D. Sadik[12], H. Alrohaif[12], L. Bastaki[13], E. El-Anany[14], M. Zakarriah[14], H. Mohamed Moussa[15], S. Jhangiani[16], Z. Coban Akdemir[17], R. Gibbs[16], D. Pehlivan[1], J. Posey[1], D. Marafi[18], J. Lupski[1]; [1]Baylor Coll. of Med., Houston, TX, [2]Human Genetics and Genome Res. Inst., Natl. Res. Ctr., Cairo, Egypt, [3]Boys Town Natl. Res. Hosp., Boys Town, NE, [4]Natl. Res. Ctr., Cairo, Egypt, [5]Clinical Genetics Dept., Human Genetics and Genome Res. Div., Natl. Res. Ctr., Cairo, Egypt, [6]Human Genetics and Genome Res. Div., Natl. Res. Ctr., Cairo, Egypt, [7]Univ. of Baghdad, Baghdad, Iraq, [8]Kuwait Med. Genetic Ctr., Sulaibikhat, Kuwait, [9]Amiri Hosp., Ministry of Hlth., Kuwait City, Kuwait, [10]Ibn Sina Hosp., Sabah Medical Area, Kuwait, [11]Farwaniya Hosp., Ministry of Hlth., Sabah Al-Nasser, Kuwait, [12]Al-Sabah Hosp., Ministry of Hlth., Sulaibikhat, Kuwait, [13]Kuwait Med. Genetic Ctr, Khaldiya, Kuwait, [14]Adan Hosp., Ministry of Hlth., Kuwait City, Kuwait, [15]Specialties Hosp. Ctr., Nouakchott, Mauritania, [16]Baylor Coll. Med., Houston, TX, [17]UTHlth.Sch. of Publ. Hlth., Houston, TX, [18]Kuwait Univ., Jabriya, Kuwait

Rare variants contributing to Mendelian disease can occur in different allelic configurations: monoallelic (heterozygous), biallelic (homozygous or compound heterozygous), and multigenic/multiallelic. Variant transmission patterns can vary amongst different population structures, influencing the relative contributions of *de novo* and new variants to recessive or complex traits. To investigate allelic transmission in presumed recessive traits, we established a cohort study of 281 unrelated families from the Middle East and North Africa (MENA), with 241/281 (85.8%) of the families noted to have parental consanguinity from familial-history reports. The MENA cohort, comprising 726 individuals, included 335 affected subjects with neurodevelopmental disorders (NDDs) and thus served as a valuable resource for 'disease gene' discovery. Family-based exome sequencing and rare variant analyses yielded molecular findings in 127 families, including 115/281 (40.9%) with known disease genes/loci and 12/281 (4.3%) with novel candidate disease genes. Overall, 121 non-overlapping NDD genes/genomic loci were uncovered, including 106 (87.5%) distinct genes with single nucleotide variants (SNVs) and small indels and 15 (12.5%) different genomic intervals associated with copy number variants. Most variants (74%) were homozygous due to parental consanguinity; as anticipated, these variant loci were embedded within absence-of-heterozygosity blocks, derived through identity-by-decent. Notably, one case featured a homozygous whole-gene deletion of *SLC13A5* generated by *Alu/Alu*-mediated genomic rearrangement. Other allelic configurations, including *de novo* heterozygosity (11%), X-linked hemizygosity (3.1%), and biallelic compound heterozygosity (1.6%), were observed. Large genomic rearrangements (8.7%) and chromosomal abnormalities (1.6%) were also uncovered. Remarkably, 59.8% of the identified SNVs in known disease genes were novel variants. Proposed novel and candidate NDD-associated genes, including *CAPZA2, CCDC53, CDK9, COG3, GOLGA2, LIMCH1, NAV2, PPP1R3F, STX6, TRAK2, UGGT1, USP19,* are currently undergoing further studies. These findings highlight the power of integrating family-based genomic analysis, comprehensive clinical phenotyping, and international collaborative data sharing to delineate allelic series and expedite the discovery of novel NDD genes/loci, paving the way for a deeper understanding of the underlying neurobiology of disease.

Session Title: Mendelian Phenotypes Poster Session III

PB4787 Hermansky-Pudlak Syndrome: Using AAV vectors to understand the development of HPS1 pulmonary fibrosis

Authors:

M. Behan[1], G. Nieto-Alamilla[2], S. Abudi[2], M. Hossain[2], W. Introne[2], W. Gahl[3], M. Malicdan[4]; [1]NHGRI, Bethesda, MD, [2]NIH, Bethesda, MD, [3]NHGRI (NIH), Bethesda, MD, [4]NIH, Bethesda, MD

Introduction: Hermansky-Pudlak Syndrome type 1 (HPS-1) is a rare autosomal recessive multisystemic disease caused by loss-of-function mutations in *HPS1*. Regardless of the mutation, all HPS-1 patients present with adult-onset pulmonary fibrosis (HPS1-PF) leading to death. The current treatment for HPS pulmonary fibrosis is lung transplantation, which is associated with significant morbidity. Dysfunctional alveolar epithelial type II cells (ATII) have been implicated in the pathogenesis of HPS1-PF. In this study, we investigated the role of ATII cells in HPS1-PF and evaluated the potential for gene complementation to rescue their dysfunction. Methods: We generated Hps1 knockout (HPS1-KO) mice by targeting the promoter region up to intron 3, and humanized knock-in (HPS1-KI) mice with the common human 16bp duplication HPS1 mutation using CRISPR/Cas genomic engineering. We utilized adeno-associated virus (AAV) vectors to deliver the human *HPS1* or eGFP open reading frame both intranasally (local delivery) and intravenously (systemic administration). After various timepoints, we collected lung samples and performed histological immunofluorescent assays. Results: HPS1-KO and HPS1-KI mice exhibited deficient Hps1 expression, along with characteristic features of HPS1, including oculocutaneous albinism, hypertrophy and accumulation of lamellar bodies in ATII cells, and a bleeding diathesis. These mice also exhibited enhanced pulmonary fibrosis in response to bleomycin challenge. Using different AAV vectors, we confirmed tropism of AAV containing eGFP to ATII cells via both systemic and local administration and saw evidence of eGFP transduction of macrophages and ATI cells, suggesting that Hps1 deficient ATII cells may still act as progenitor stem cells for the alveolus. Additionally, AAV-HPS1 delivered to Hps1 mice showed normalized HPS1 expression in the lungs after two weeks and evidence of reduced collagen deposition in comparison to untreated mice when challenged with bleomycin. Treated mice also showed a reduction in ATII cell hypertrophy after six months. Conclusion: This study provides evidence for the involvement of ATII cells in the manifestation of HPS1-PF and highlights the potential for gene complementation to rescue their dysfunction. Our findings contribute to a better understanding of HPS1-PF pathogenesis and offer insights for developing novel therapeutic strategies. Ongoing experiments will explore Hps1-deficient ATII cell proliferation, differentiation, and fibroblast recruitment capabilities following gene complementation.

Session Title: Mendelian Phenotypes Poster Session I

PB4788 Heteroplasmic pathogenic m.12315G>A variant in *MT-TL2* presenting with MELAS syndrome and depletion of nitric oxide donors

Authors:

F. Scaglia[1], M. Snyder[1], Y. Manor[2], S. F. Kralik[3], E. Mizerik[3], C. Gijavanekar[3], S. Elsea[3], K. Machol[4], L. Emrick[4]; [1]Baylor Col Med, Houston, TX, [2]Chaim Sheba Med. Ctr., Tel Hashomer, Israel, [3]Baylor Coll. of Med., Houston, TX, [4]Baylor Coll. Med., Houston, TX

The *MT-TL2* m.12315 G>A heteroplasmic pathogenic variant has previously been reported in two individuals with milder clinical phenotypes predominantly associated with myopathic features. Herein we report the case of a child with this heteroplasmic variant who developed neurological regression and stroke-like episodes similar to those observed in mitochondrial encephalopathy, lactic acidosis and stroke-like episodes (MELAS). Biochemical evaluation revealed depletion of arginine on plasma amino acid analysis and low z-scores for citrulline on untargeted plasma metabolomics analysis. These findings suggest that her stroke-like episodes may be due to decreased availability of nitric oxide. The use of intravenous arginine during stroke-like episodes and enteral citrulline supplementation for home use normalized her biochemical values of arginine and citrulline. Untargeted plasma metabolomics showed the absence of nicotinamide and 1-methylnicotinamide, and plasma total glutathione levels were low, thus nicotinamide riboside and N-acetylcysteine therapies were initiated which led to normalization of metabolomics results. This report expands the clinical and metabolic phenotype associated with the rare mitochondrial variant m.12315 G>A to include neurological regression, stroke-like episodes, and decreased availability of nitric oxide donors. Individuals with this heteroplasmic variant should undergo in-depth biochemical analysis to include untargeted plasma metabolomics, plasma amino acids, and glutathione levels to help guide a timely and targeted approach to treatment.

Session Title: Mendelian Phenotypes Poster Session II

PB4789 Heterozygous variants in *ZFR* cause neurodevelopmental alterations, intellectual disability, and microcephaly.

Authors:

E. Richard[1], A. Peyrel[1], A. Acharya[2], M. Nizon[3], B. Cogne[4], J. Levy[5], T. Bharadwaj[2], Y. Rajendran[2], H. C. Shinn[2], L. Ruaud[6], M. J. Lyons[7], R. J. Louie[7], B. Peterlin[8], M. Rogac[8], S. M. Leal[2,9], B. Golob[10], I. Järvelä[11], I. Schrauwen[2]; [1]MMDN, Univ Montpellier, EPHE, INSERM, Montpellier, France, [2]Ctr. for Statistical Genetics, Gertrude H. Sergievsky Ctr. and the Dept. of Neurology, Columbia Univ. Med. Ctr., New York, NY, [3]CHU Nantes, Service de Génétique Médicale, L'institut du thorax, INSERM, CNRS, UNIV Nantes, Nantes, France, [4]Nantes Université, CHU de Nantes, CNRS, INSERM, L'Inst. du thorax, Nantes, France, [5]Département de Génétique, Hôpital Robert Debré, AP-HP, Paris, France, [6]APHP Nord, Robert Debré Univ. Hosp., Dept. of Genetics, Université Paris Cité, Inserm UMR 1141, NeuroDiderot, Paris, France, [7]Greenwood Genetic Ctr., Greenwood, SC, [8]Clinical Inst. of Genomic Med., Univ. Med. Ctr. Ljubljana, Ljubljana, Slovenia, [9]Taub Inst. for Alzheimer's Disease and the Aging Brain, Columbia Univ. Med. Ctr., New York, NY, [10]Ctr. for Mendelian Genomics, Clinical Inst. of Genomic Med., UMC Ljubljana, Ljubljana, Slovenia, [11]Dept. of Med. Genetics, Univ. of Helsinki, Helsinki, Finland

ZFR codes for an RNA-binding protein characterized by its DZF (domain associated with zinc fingers) domain and of unknown function. ZFR is ubiquitously expressed in human tissues, with a high expression in the brain from early development stages to adulthood.Using next-generation sequencing and/or molecular karyotyping, we report 6 variants in *ZFR* associated with a novel neurodevelopmental disorder in 9 affected individuals from 6 unrelated families. The variants are *de novo* or autosomal dominantly inherited and include three frameshift, one nonsense, one missense variant and one large deletion encompassing *ZFR*. All *ZFR* variants were predicted to be damaging using *in silico* tools and likely lead to a loss-of-function (LOF). ZFR exhibits a high intolerance to loss-of-function (LOF) variants (pLI=1; observed/expected score=0.05). Considering these findings alongside the observed variants, it strongly indicates that haploinsufficiency is the probable underlying mechanism of the disease. Affected individuals present with neurodevelopmental abnormalities including speech delay/abnormalities, intellectual disability or learning difficulties, and global developmental delay. In addition, microcephaly, distinctive facial features and/or hypotonia were reported for some subjects.We investigated the impact of *ZFR* loss of function variants utilizing *in vivo* functional characterization of zebrafish model. Specific splice-site targeting morpholinos were used to knock-down *zfr* expression. Consistent with the human presentation, 5dpf morphant larvae exhibited microcephaly associated with swimming behavior defects.Collectively, our data indicate that heterozygous variants in *ZFR* lead to a novel neurodevelopmental disorder that can be recapitulated by knocking-down *zfr*, the orthologous gene, in zebrafish, highlighting the role of ZFR in brain development.

Session Title: Mendelian Phenotypes Poster Session III

PB4790 HiFi read sequencing reveals causative complex structural variants in a Japanese patient with severe mental retardation and microcephaly.

Authors:

K. Yanagi[1], M. Yamamori[1], A. Igarashi[1], K. Satou[1], Y. Matsubara[2], T. Kaname[1]; [1]Genome Med., Natl. Ctr. for Child Hlth.and Dev., Setagaya, Tokyo, Japan, [2]Natl. Ctr. for Child Hlth.and Dev., Setagaya, Tokyo, Japan

[Background] Long-read sequencing has made it possible to detect complex structural variants that may be difficult to detect with short-read sequencing analysis. These include repeat expansion, large insertion, inversions, deletions or translocations. We describe a successful identification of complex structural variants that are responsible for symptoms by HiFi read sequencing. [Patients] The proband was a 1-year and 3-month-old Japanese girl born at 38 weeks and 3 days gestation, normal vaginal delivery, weight 3038 g, length 48.5 cm, and OFC 30.7 cm. Her parents were healthy and nonconsanguineous. She presented severe mental retardation, microcephaly, and hypoplasia of the corpus callosum. Her facial appearance was round with arched eyebrow, low nasal root, short nose, anteverted nares and large ears. [Whole-genome analysis] We first performed short-read sequencing on the trio and segregated a CNV gain and a loss on chromosome 14 in the proband. Since detailed analysis led to the suspicion of chromosomal abnormalities, we next conducted HiFi read sequencing on the proband. [Results] Mapping to T2T-CHM13v2.0 clearly showed a CNV loss of approximately 10.7 Mb at 14q12-q13.2 including *FOXG1* and a CNV gain of approximately 2 Mb at 14q11.2 including *SUPT16H* and *CHD8*. In addition to the CNV variants, a translocation 14q13.2-14qter to 22qter was detected.[Conclusion] HiFi read sequencing clearly presented complex unbalanced chromosomal translocation. Haploinsufficiency of *FOXG1* is known to cause Rett syndrome which phenotype includes postnatal growth retardation and microcephaly, hypotonia, epilepsy, stereotypic movements and feeding problems. On the other hand, microduplication of *SUPT16H-CHD8* was shown to cause developmental delay, intellectual disability, autism spectrum disorders and macrocephaly. The proband presented severe mental retardation and microcephaly, but not stereotypic movements. We conclude both of the deletion of 14q12-q13.2 and microduplication of 14q11.2 may be responsible for the symptoms in this case.

Session Title: Mendelian Phenotypes Poster Session I

PB4791 High-content microscopy screening for drug discovery in Cohen Syndrome.

Authors:

M. Choung[1], F. Vacca[1], R. Da Costa[1], H. Barakullah[1], H. Riezman[2], M. Ansar[1,3]; [1]Dept. of Ophthalmology, Univ. of Lausanne, Jules Gonin Eye Hosp., Fondation Asile Des Aveugles, Lausanne, Switzerland, [2]Dept. of Biochemistry, Univ. of Geneva, Geneva, Switzerland, [3]Advanced Molecular Genetics and Genomics Disease Res. and Treatment Ctr., Dow Univ. of Hlth.Sci., Karachi, Pakistan

Cohen syndrome (CS) is a rare autosomal recessive disorder caused by biallelic mutations in the *VPS13B* gene. Over 100 pathogenic variants have been identified so far, mostly truncating loss-of-function mutations. It affects an estimated 50,000 individuals worldwide and is characterised by multiple clinical features, including acquired microcephaly, developmental delay, intellectual disability, neutropenia, and retinal degeneration. VPS13B is part of the chorein domain protein family, which in mammals also includes the VPS13A, -C, and -D proteins as well as the autophagy factor ATG2. These are peripheral membrane proteins with different sub-cellular localisations, but they all share similar structural features and have been proposed to act as lipid transport proteins at organellar membrane contact sites. Precise VPS13B function and the selectivity for transported lipids are not yet established, but the protein is localised at the Golgi complex, and fragmentation of this organelle is the more evident hallmark of *VPS13B*-deficient cells. We make use of this cell-autonomous morphological defect in the Golgi complex to develop a fluorescence microscopy-based assay for high-throughput screening of small molecules. We screened the Prestwick chemical library, composed of 1280 FDA-approved compounds. We identified two different classes of compounds that efficiently recover Golgi morphology in *VPS13B* knockout Hela cells: glucocorticoid receptor agonists and lysosomotropic drugs, which act as functional inhibitors of acidic lipases. For a selected number of hits, the results were further confirmed in patient-derived fibroblasts. We performed lipidomic analysis and found a general reduction in sphingolipids in *VPS13B* KO cells. Among sphingolipids, a more striking decrease was observed in the C18 n-acyl species of sphingomyelin (SM). These SM species, while of low abundance, are strongly enriched in COPI vesicles and are proposed to play a crucial role in Golgi to ER transport, linking this observation with the Golgi phenotype. Interestingly, all tested lysosomotropic drugs were able to recover the cellular level of C18 SM species in *VPS13B* KO cells, indicating a potential mechanism of action for this class of drugs. In conclusion, we developed a robust cell-based assay for the screening of small molecules and identified potential drug candidates for CS. We propose phenotypic screening as a valuable option for drug pre-selection before testing in animal models.

Session Title: Mendelian Phenotypes Poster Session II

PB4792 Highly penetrant normal-tension glaucoma risk allele maps to chromosome 12q13 and is associated with extracellular matrix dysregulation in a large family.

Authors:

W. Presley[1], S. Garnai[1], B. Guan[2], S. E. Moroi[3], B. Blevins[1], N. Michaels[1], B. Bohnsack[4], J. A. Rosenfeld[5], R. A. Lewis[5], L. Burrage[5], S. Ketkar[5], S. H. Jeong[1], R. B. Hufnagel[2], J. E. Richards[1], J. Stein, D[1], L. Prasov[1]; [1]Univ. of Michigan, Ann Arbor, MI, [2]Natl. Eye Inst., NIH, Bethesda, MD, [3]The Ohio State Univ., Columbus, OH, [4]Ann and Robert H. Lurie Children's Hosp. of Chicago, Chicago, IL, [5]Baylor Coll. of Med., Houston, TX

Glaucoma describes a group of ocular conditions characterized by degeneration of the optic nerve and is a leading cause of vision loss worldwide. While intraocular pressure (IOP) remains the largest modifiable risk factor for this disease, an estimated 30% of cases are classified as normal-tension (pressure) glaucoma (NTG) - progressive optic neuropathy despite a normal IOP measurement during clinic hours. A subset of NTG patients fail to respond to traditional IOP-lowering therapy. Understanding the molecular pathogenesis of NTG is critical to developing new therapies, and disease gene discovery in rare, highly penetrant familial cases remains one substantive approach to this end. Our group identified a 54-member pedigree with early onset, severe NTG that continued to progress despite multiple effective IOP-lowering surgeries. Genomic linkage and segregation analyses suggest an autosomal dominant inheritance pattern, but that there are multiple NTG risk alleles of high effect present in this family. Most cases, however, map to a linkage region on chromosome 12q13 with a multipoint parametric Logarithm of Odds score of 2.6 (calculated using a dominant, fully penetrant model). This region overlaps a known optic neuropathy GWAS peak and contains approximately 50 genes. RNAseq analysis from affected and unaffected familial blood samples revealed dysregulation of extracellular matrix (ECM) genes tracking with the chromosome 12 haplotype. Specifically, a Panther Gene Ontology analysis showed a downregulation of ECM structural constituents (FDR $7.3x10^{-8}$), as well as genes that function in fibrinogen, fibronectin, and integrin binding (FDRs $1.6x10^{-2}$, $2.6x10^{-3}$, and $7.1x10^{-3}$ respectively) and collagen metabolism (FDR $2.4x10^{-2}$). We have also identified a second family with NTG that is poorly responsive to traditional therapy and appears to be attributable to a variant in the same region of chromosome 12. Functional analyses for candidate gene variants within the linkage region are underway. Overall, these data support a role for ECM dysregulation in NTG pathogenesis. ECM proteins are an intriguing target as they are the primary components of the lamina cribrosa, a structure that both provides key support for the optic nerve and represents the major site of axonal damage across the different forms of glaucoma. Uncovering the genetic basis of glaucoma in these families, particularly for ECM regulation, will inform future screening and may lead to new non-IOP based treatment options for NTG and other forms of glaucoma.

Session Title: Mendelian Phenotypes Poster Session III

PB4793 HUWE1-related neurodevelopmental disorder: expanding the phenotype.

Authors:

K. Baranano[1], K. Hogan[2], M. Li[2]; [1]Johns Hopkins Univ. Sch. of Med., Baltimore, MD, [2]Rush Univ., Chicago, IL

HUWE1 (HECT, UBA, and WWE Domain Containing E3 Ubiquitin Protein Ligase1) encodes an E3 ubiquitin ligase involved in targeting specific proteins for ubiquitination and degradation. HUWE1 is located on the X chromosome. The largest cohort (21) of individuals was described in 2018, but since that time, based on family involvement in social media groups, there are likely hundreds more who have been diagnosed, as well as undiagnosed adults. We describe a large series of previously unpublished individuals affected by HUWE1-related neurodevelopmental disorder. Common findings include a characteristic overall facial gestalt, especially in boys, global developmental delay, epilepsy, autism spectrum disorder, structural brain differences and short stature. Craniosynostosis is only seen with variants at one particular residue, R110. Cognitive function tended to be higher in affected girls than boys. We propose a set of management guidelines based on review of the literature and these findings.

Session Title: Mendelian Phenotypes Poster Session I

PB4794 Hyperammonemia, Hypocarnitinemia, Rhabdomyolysis and Pancreatitis in a Patient with Undiagnosed 3MethylcrotonylCoA Carboxylase Deficiency

Authors:

S. Yano, J. Borelli, R. McGowan; Univ. of Southern California, Los Angeles, CA

Background: 3methylcrotonylCoA carboxylase (3MCC) deficiency, which is an autosomal recessive disorder of leucine metabolism. The incidence is approximately 1in 40,000. This disorder is one of the most common organic acidemias diagnosed by newborn screening (NBS). Screening for many organic acidurias including 3MCC deficiency started in California in 2005. Adults born before the introduction of this expanded NBS were not tested for 3MCC deficiency. 3MCC deficiency is thought to be a benign condition since most cases are asymptomatic although metabolic decompensation, including hypoglycemia and hyperammonemia associated with hypocarnitinemia, is reported in rare cases. Some of the NBS programs have decided to exclude 3MCC deficiency and some of the individuals are no longer followed. We report on a 24-year-old male who developed a life-threatening episode triggered by strenuous exercise and was diagnosed with 3MCC deficiency. **Case Study:** We report a case of a 24-year-old previously healthy male who was hospitalized following a 13-mile hiking trip in harsh conditions due to dehydration, acute onset of emesis, generalized muscle weakness, altered consciousness due to hyperammonemia (217uM), respiratory failure requiring assisted ventilation, rhabdomyolysis with creatine kinase (CK) level of 2,977 (later increased to 8,887) units/L and renal failure. High plasma amylase and lipase levels suggesting pancreatitis were noted on admission. Throughout his clinical course, hypocarnitinemia was diagnosed. The patient was treated with carnitine (50mg/kg/d). Plasma acylcarnitine profile and urine organic acids analysis resulted after discharge showing increased plasma C5OHcarnitine, urine 3methylcrotonylglycine, and 3hydroxyisovalerate to establish the diagnosis of 3MCC deficiency. A low plasma C2carnitine was also noted. The whole exome study showed a homozygous pathogenic variant in the MCCC2 gene (c.994C>T). **Discussion**: While 3MCC deficiency is considered a benign condition, those not treated with carnitine may have more significant health risks than previously thought. Our case developed hyperammonemia causing altered consciousness as a symptom of 3MCC deficiency, a previously under-reported complication of this disorder. We also found the patient to have rhabdomyolysis and pancreatitis, which have not been reported to the best of our knowledge. Unexpected life-threatening episodes observed in our case may indicate the need for carnitine supplementation even for asymptomatic individuals with 3MCC deficiency and also raise a question as to whether NBS should include this condition.

Session Title: Mendelian Phenotypes Poster Session II

PB4795 Hyperproliferative lymphatic sprouting arising from *EPHB4* mutations can be rescued through mTOR inhibition in a cell-based model system of central conducting lymphatic anomaly.

Authors:

M. Battig[1], M. E. March[1], S. E. Sheppard[2], C. Kao[1], C. Seiler[1], D. Li[1], H. Hakonarson[1]; [1]Children's Hosp. of Philadelphia, Philadelphia, PA, [2]NICHD/NIH, Rockville, MD

Central conducting lymphatic anomaly (CCLA) is a severe disorder characterized by dilated lymphatic channels and lymphatic channel dysmotility, resulting in chylothorax, pleural effusions, and lymphedema. The genetic etiology of this disorder is not well characterized, though pathogenic variants in RAS/MAPK signaling pathway genes such as *KRAS* and *ARAF* have recently been described. We have previously identified a kinase-inactivating mutation in *EPHB4* (c.2334+1G>C:p.L778_G779insLMLG) found in a patient with CCLA. *EPHB4* encodes ephrin type-B receptor 4, which binds ephrin B2 and regulates vascular processes such as angiogenesis and vascular remodeling. Here, five *EPHB4* missense variants and the insertional mutant are compared through functional studies using transduced human dermal lymphatic endothelial cells to better understand the molecular mechanism by which these mutants cause CCLA. Immunoblotting revealed that these mutations led to increased Akt/mTOR signaling. In an *in vitro* model of lymphangiogenesis, spheroids harboring the *EPHB4* mutations exhibited greater sprouting of capillary-like structures than their wild-type counterpart. Additionally, when these spheroids were treated with VEGF-C, a potent inducer of lymphangiogenesis, the increase in sprouting was greater for the mutants than for the wild-type. Furthermore, while ephrin B2 was able to suppress lymphangiogenic sprouting induced by VEGF-C of spheroids expressing wild-type *EPHB4*, it failed to curtail sprouting when these mutations were present. Finally, suppression of mTOR activity was evaluated using rapamycin (mTORC1 inhibitor) and OSI-027 (dual mTORC1/2 inhibitor). Both inhibitors were able to effectively rescue the phenotype, as observed by the reduction of sprouting to basal levels. These findings offer insight into how mutations in *EPHB4* disrupt its negative regulatory function, resulting in uncontrolled lymphangiogenesis, and how mTOR inhibitors may have therapeutic benefits in patients with CCLA and other disorders stemming from upregulation of Akt/mTOR signaling.

Session Title: Mendelian Phenotypes Poster Session III

PB4796 Identification and functional analysis of *RRAS2* variants in patients with Noonan-like phenotype

Authors:

T. Kaname[1], T. Iida[1], T. Kimura[2], H. Go[3], A. Igarashi[1], K. Fukunaga[1], T. Aoki[1], K. Yanagi[1], T. Hidai[1], M. Yamamori[1], K. Satou[1], T. Kosho[4], I. Inoue[2], Y. Matsubara[5]; [1]Dept Genome Med, Natl Ctr Child Hlth Dev (NCCHD), Tokyo, Japan, [2]Natl. Inst. of Genetics, Mishima, Japan, [3]Dept Pediatr, Fukushima Med Univ Schl Med, Fukishima, Japan, [4]Shinshu Univ. Sch. of Med., Matsumoto, Japan, [5]Natl. Ctr. for Child Hlth.and Dev., Tokyo, Japan

RASopathies are a clinically defined group of genetic syndromes caused by pathogenic variants in genes that encode components or regulators of the Ras/MAPK pathway. The RRAS2 gene abnormlity has recently been reported as one of the RASopathies caused by aberrant RAS/MAPK signalling. RRAS2, a member of the R-Ras subfamily, is a signaling molecule located in the plasma membrane. Like other RAS proteins, it is thought to regulate cell proliferation and differentiation via the RAS/MAPK pathway. Here, we report two patients with Noonan-like phenotype and functional analysis of *RRAS2* variants identified in them. One patient was an 8-year-old boy. He presented with mild developmental delay, characteristic facial features (hypertelorism, down slanted palpebral fissures, broad nasal root, low-set ears, Macrotia, mid-face hypoplasia), VSD, anal atresia. Another patient was a 2-year-old, 10-month-old boy. He presented with macrocephaly (+5.1 SD), characteristic facial features (prominent forehead, thick lower lip vermilion, hypertelorism, low-set, posteriorly rotated ears), bilateral cryptorchidism, bilateral hydronephrosis grade 2 and hepatomegaly.
After written informed consent from their parents, whole-exome sequencing and filtering analyses in trios were performed and recurrent and novel de novo variants in RRAS2, p.Gly23Val and p.Gly24Glu, were identified in the patients, respectively. We functionally analysed these variants by expressing the RRAS2 proteins in HEK293 cells and zebrafish embryos. Injection of RRAS2 mRNA (wild-type (WT), p.Gly23Val) into zebrafish embryos showed significant embryonic deformation and abnormal 4-day embryo jaw development in the mutant RRAS2 expression. Next, to measure RRAS2 activity, an activity assay system was then constructed in HEK293 cells using a luciferase reporter gene built into a serum response element (SRE), the binding site for the transcription factor ELK1, which is activated by ERK downstream of the RAS/MAPK pathway. Each RRAS2 was expressed in the HEK293 cells using an expression vector, showing the activity of RRAS2 p.Gly23Val and p.Gly24Glu was 2.45-fold and 3.06-fold higher than wild-type, respectively.
We concluded that the *RRAS2* p.Gly23Val and p.Gly24Glu found in the affected patients are gain-of-function variants, and that the patients were Noonan syndrome caused by enhanced activity of the RAS signalling pathway.

Session Title: Mendelian Phenotypes Poster Session I

PB4797 Identification and Functional Evaluation of Autosomal Recessive Non-Syndromic Hearing Impairment Genes in Rwanda

Authors:

E. Uwibambe[1], A. Wonkam[2], L. Mutesa[1]; [1]Univ. of Rwanda, Kigali, Rwanda, [2]Dept. of Genetic Med., JHU, Baltimore, MD

The incidence of hereditary hearing impairment (HI) is higher in developing countries compared to developed countries. More than 120 independent genes have been identified as responsible for almost 50% of profound HI. Non-syndromic HI is the most common form accounting for 70% of HI cases of which 80% are autosomal recessive. Reported mutations in GJB2, GJB6, and GJA genes are the most common cause of HI but studies among Cameroonian and South African participants did not identify a significant association, hence the need for further genetic exploration of responsible genes in African population. In Rwanda, more than 50% of HI among children is attributed to hereditary causes but no genetic evidence has been established. This study aims at identifying and evaluating genes responsible for autosomal recessive non-syndromic hearing impairment (ARNSHI) among Rwandan population. We have recruited 80 families with HI and 100 control individuals without HI or family history of HI. One proband per affected family has been recruited together with one unaffected sibling, both parents, and at least one affected sibling or other affected relative for blood sampling. Standard audiometry including air and bone conduction has been performed for every individual recruited, and clinical history has been taken to exclude exposure to ototoxic drugs or infections including prenatal exposure and other environmental causes of HI. Twenty-seven affected families comprising 105 individuals have been sequenced thus far. The preliminary results have shown that the gene variants that are associated with HI in Rwanda include *TCM1, MYO7A, MYO15A, USH2A, CDH23, PDZD7, CACNA1H, MYO3B, COL4A3, ERCC4, DIABLO, DMXL2, ANKRD22, KCNE1,* and *NCOA3.* For each family recruited, we will exome-sequence samples of the affected members and use Sanger sequencing to follow up variants segregating in their parents and non-affected sibling, and controls. At the end of the study, we expect to establish a dataset of genes that cause ARNSHI in Rwanda with possible novel ARNSHI genes that have not yet been reported in the African population. This will help to advance the science of HI and establish appropriate medical care including proper genetic counseling necessary for affected individuals and families as well as plan for prevention and control measures against genetic HI.

Session Title: Mendelian Phenotypes Poster Session II

PB4798 Identification of a novel pathogenic *SPTLC1* variant associated with juvenile amyotrophic lateral sclerosis born to an asymptomatic father with a mosaicism.

Authors:

S. Okubo[1], H. Naruse[1], A. Sudo[1], H. Ishiura[2], J. Mitsui[3], T. Matsukawa[1], W. Satake[1], O. Yasushi[4], S. Tsuji[5], T. Toda[1]; [1]Dept. of Neurology, Graduate Sch. of Med., The Univ. of Tokyo, Tokyo, Japan, [2]Dept. of Neurology, Okayama Univ. Graduate Sch. of Med., Dentistry and Pharmaceutical Sci., Okayama, Japan, [3]Dept. of Precision Med. Neurology, Graduate Sch. of Med., The Univ. of Tokyo, Tokyo, Japan, [4]Dept. of Neurology, Natl. Ctr. of Neurology and Psychiatry, Tokyo, Japan, [5]Inst. of Med. Genomics, Intl. Univ. of Hlth.and Welfare, Chiba, Japan

Amyotrophic lateral sclerosis (ALS) is a progressive neurodegenerative disease of the motor system, affecting upper and lower motor neurons. We experienced a 21-year-old female patient presenting with progressive muscle atrophy and weakness beginning at the age of six. Both parents were healthy without any neurological abnormalities. Neurological examination revealed cognitive decline, symmetrical weakness predominant in the distal upper extremity, muscular atrophy including the tongue, absent tendon reflexes in the upper extremities, exaggerated Achilles tendon reflexes, and positive Babinski signs. Neurophysiological testing at age 21 revealed both acute and chronic denervation without evidence of sensory neuropathy. Blood and cerebrospinal fluid examinations were unremarkable. These findings met the Awaji criteria for probable ALS and the Gold Coast criteria. Employing whole genome sequence analysis, we searched for rare variants (minor allele frequencies < 0.01) in 52 reported ALS-causative genes. The patient did not possess rare or known pathogenic variants, except for a heterozygous variant, p.Ala20Thr (c.58G>A) (NM_006415), in exon 2 of *SPTLC1* gene. This variant was absent in population databases including gnomAD, 38KJPN, and 1,163 in-house healthy control individuals. Multiple in silico predictors predicted the variant as deleterious. Interestingly, p.Ala20Ser variant involving the same amino acid has been reported to cause exon 2 skipping. A reverse transcription polymerase chain reaction (RT-PCR) analysis of the total RNA prepared from cultured lymphoblastoid cell lines revealed that the size of the RT-PCR products was consistent with that of the wild type allele, indicating that this variant does not result in exon 2 skipping. While the mother had the wild-type sequence of *SPTLC1*, the father was suspected to have a mosaic variant by inspection of the electropherogram of the direct nucleotide sequence analysis. Droplet digital PCR confirmed that relative abundance of the variant was 17% in the father's peripheral blood leukocytes. Although interpretation of the pathogenicity of p.Ala20Ser as likely pathogenic is not sufficient based on the ACMG guideline, we consider that the mosaicism in the unaffected father along with the specific phenotype of the patient as juvenile ALS supports the pathogenicity. *SPTLC1* is a gene encoding a subunit of serine palmitoyltransferase (SPT).
Recently, *SPTLC1* variants clustering in exon 2 have been shown to cause juvenile ALS. Further investigation on the functions of the mutant SPTLC1 will be needed to further clarify the pathogenetic mechanisms underlying juvenile ALS.

Session Title: Mendelian Phenotypes Poster Session III

PB4799 Identification of *CSNK2A1* variants in four unrelated Okur-Chung neurodevelopmental syndrome patients with microcephaly expands the clinical phenotypic spectrum of the disorder.

Authors:

S. Ramadesikan[1], I. Showpnil[1], A. Daley[1], E. Varga[1], M. Pastore[2], E. Sites[2], D. Bartholomew[2,3], P. White[1,3], R. K. Wilson[1,3], R. Stottmann[1,3], D. C. Koboldt[1,3]; [1]Steve and Cindy Rasmussen Inst. for Genomic Med., The Res. Inst. at Nationwide Children's Hosp., Columbus, OH, [2]Div. of Genetic and Genomic Med., Nationwide Children's Hosp., Columbus, OH, [3]Dept. of Pediatrics, The Ohio State Univ. Coll. of Med., Columbus, OH

Nearly 40% of neurodevelopmental disorders are caused by *de novo* coding variants, many of which impact genes that are critical in embryonic brain development. In this study, we describe four patients from three unrelated families who shared several clinical features characteristic of an underlying syndromic neurodevelopmental condition, including microcephaly, dysmorphic features, hypotonia, intellectual disability, and developmental delays. Clinical whole exome sequencing and research genome sequencing under an IRB-approved study protocol revealed that all affected individuals harbored heterozygous pathogenic missense variants in the gene *CSNK2A1*. *CSNK2A1* encodes a Serine/Threonine kinase called casein kinase 2 alpha subunit 1 (also known as CK2A1), which phosphorylates nearly four hundred downstream target proteins in cells, thereby participating in many essential cellular processes that aid in neuronal development and homeostasis. *De novo* heterozygous alterations in *CSNK2A1* cause autosomal dominant Okur- Chung neurodevelopmental syndrome (OCNDS; OMIM #617062). Of the three variants (c.468T>A p.D156E, c.137G>T p.G46V, c.149A>G p.Y50C) identified in this study, two are novel ((c.468T>A p.D156E, c.137G>T p.G46V) and previously unreported in literature. Notably, while most reported variants are *de novo*, one of the variants in this study (c.137G>T p.G46V) was inherited and segregated across multiple affected individuals in the family, an observation reported only once previously.
This recently identified syndrome has an evolving clinical phenotype predominantly characterized by dysmorphic features, intellectual disability, global delays, gastrointestinal and immunological manifestations, with microcephaly currently considered an infrequent occurrence. Interestingly, we report microcephaly in three of our four patients, two of whom are related and share the inherited variant. Detailed phenotypic analysis revealed that microcephaly has been reported in ~35% of *all* patients reported in literature (including this study) with no apparent genotype-phenotype correlation. Our findings highlight the recurrence of microcephaly in OCNDS and the need for progressive neurological evaluation of OCNDS patients as part of their clinical management.

Session Title: Mendelian Phenotypes Poster Session I

PB4800 Identification of de Novo Variants in Genes Involved in Ubiquitylation in Patients with Chronic Recurrent Multifocal Osteomyelitis (CRMO)

Authors:

A. Cox[1], B. W. Darbro[2], X. Bing[2], E. Phillippi[2], G. Velez[3], R. M. Laxer[4], Y. Zhao[5], V. B. Mahajan[3], H. El-shanti[2], A. G. Bassuk[2], P. J. Ferguson[2]; [1]Univ. of Rochester Med. Ctr., Rochester, NY, [2]Univ. of Iowa, Iowa City, IA, [3]Stanford Univ., Palo Alto, CA, [4]Hosp. for Sick Kids, Toronto, ON, Canada, [5]Seattle Children's Hosp., Seattle, WA

Chronic recurrent multifocal osteomyelitis (CRMO) is a rare autoinflammatory disease characterized by bone pain and sterile inflammatory osteolytic lesions. Often children are also diagnosed with psoriasis, Crohn disease, or other inflammatory conditions, and these more common diseases are frequently present in close relatives of probands with CRMO. Because patients present with a family history of autoimmune/autoinflammatory disease including some families with CRMO-affected siblings, CRMO is likely a genetic disease.

Genetic analyses of both patients with CRMO and mouse models of inflammatory bone disease have implicated the involvement of *Pstpip2*, *FBLIM1*, *FGR*, and *MLKL* variants in disease pathogenesis. Recently, in collaboration, *FGR* variants were identified in both a mouse strain (*Ali18*) with paw inflammation and systemic reduced bone density, and in a proband who harbored a *de novo* missense variant. This proband is part of a cohort of 105 families with a child or children with CRMO. We performed whole exome sequencing of the probands and their parents to identify candidate genes involved in the pathogenesis of CRMO. The families were investigated for coding *de novo* variants.

Over one hundred coding *de novo* variants were identified, including multiple variants in genes known to be involved in ubiquitylation. Dysregulated ubiquitylation has been implicated in other autoinflammatory disorders including those caused by germline or somatic pathogenic variants in *OTULIN*, *TNFAIP*3, and *UBA1*. The *de novo* variants identified in the patients with CRMO are in genes directly involved in ubiquitylation and also known to regulate NF-kB and/or inflammasome activation. Functional analysis of top candidate variants is currently underway.

This study is the largest genetic analysis of patients with CRMO to date. CRMO is a rare autoinflammatory disease with a mostly elusive genetic etiology and our results highlight the genetic complexity of its pathogenesis. These findings, in combination with prior knowledge gained by analysis of other families and mouse models of inflammatory arthritis, support a linked network of genes and biological processes which are aberrant in diseased patients.

Session Title: Mendelian Phenotypes Poster Session II

PB4801 Identification of different genetic variants causing recessive neurodevelopmental syndromes with intellectual disability in Pakistan.

Authors:

A. Rasheed[1], S. Fatima[1], T. Sultan[2], F. Alzahrani[3], A. Iftikhar Butt[1], G. Seo[4], F. Alkuraya[5], S. Naz[1]; [1]Univ. of Punjab, Lahore, Pakistan, [2]Children's Hosp. & the Inst. of Child Hlth., Lahore, Pakistan, [3]King Faisal Specialist Hosp. & Res. Ctr., Riyadh, Saudi Arabia, [4]3billion inc., Seoul, Korea, Republic of, [5]King Faisal Specialist Hosp. & Res. Ctr., Riyadh, Saudi Arabia

Neurodevelopmental disorders (NDDs) affect about 3 to 5% of the population worldwide. These conditions disrupt the development, physiology, and functioning of the brain. Motor function, communication, cognition, learning, and language may also be affected. We studied NDD syndromes in nine Pakistani families. Patients were identified and recruited from a hospital in Lahore and Rahim Yar Khan. Clinical examinations were completed by doctors at the hospitals. The patients displayed intellectual disability and global developmental delay with additional phenotypes which included facial dysmorphism, microcephaly, abnormal movements of the head and eyes, hypotonia, or epilepsy in some affected individuals. We employed exome sequencing to identify the molecular basis of NDD in the patients. Exome data were filtered to obtain homozygous variants having allele frequencies of less than 1% in various population databases. Variants shared by multiple patients in a family were selected. The missense variants were further prioritized based on the conservation of affected amino acids in 100 vertebrate species. Compound heterozygous variants were also checked in case no predicted homozygous pathogenic variant was observed. Variant segregation for each family was verified by PCR amplification of DNA samples of all participants followed by Sanger sequencing. We identified known variants in *MAN2B1, SAMHD1,* and *VPS13B* as well as six novel variants in *EPG5, ERRC8, ETHE1, RNASEH2A,* and *UBE4A,* as causing various recessive neurodevelopmental syndromes. Our work extends the allelic spectrum of NDD. It facilitates genetic counseling of the participating families and will help in prenatal diagnosis to those inclined to its use.

Session Title: Mendelian Phenotypes Poster Session I

PB4803 Identification of novel mutations in *MCPH1* gene in Pakistani family with microcephaly

Authors:

S. Irshad[1], z. riaz[2]; [1]Sch. of Biochemistry and Biotechnology, Univ. of the Punjab, Lahore, New York, NY, [2]Sch. of Biochemistry and Biotechnology, Univ. of the Punjab, Lahore, Lahore, Pakistan

Background: Primary microcephaly (MCPH) is a congenital, static, and non-progressive neurodevelopmental disorder associated with reduced head circumference (< 4 standard deviations). About 28 known genes are associated with the MCPH. The study was carried out to probe molecular basis and genetic variants involved with MCPH in an affected Pakistani family to better understand the etiology and prevalence of the disorder. **Methods:** The individuals of the ascertained Pakistani family presented primary microcephaly, along with intellectual disability, speech disorder, and motor delay. By ensuring ethical compliance and patient consent, blood samples were collected from affected individuals. DNA was extracted using the salting out method followed by whole-exome sequencing and Sanger sequencing to identify causative genetic variants or mutations. In silico studies were performed to predict the effect of mutations on the structure of target proteins. **Results:** Two missense allelic variants (NM_024596.5: c.139G>C and NM_024596.5: c.211G>C) of *MCPH1* gene were detected in a Pakistani family. In silico analysis was performed to evaluate the effect of the mutant protein. The mutation in genes affects the activities of proteins NM_024596: p. Val47Leu and NM_024596: p. Val71Leu respectively by disruption in protein structure. The mutations were predicted to have higher pathogenicity scores and have significantly influenced the prevalence of MCPH. We reported two novel genetic variants for the first time from the Pakistani population causing MCPH. **Conclusions:** Mutations in *MCPH1* gene are one of the major cause of MCPH in the populations where consanguine marriages are common. The novel mutations identified in this study will help to understand the etiology of the disorder and the mechanisms of mutated proteins.

Session Title: Mendelian Phenotypes Poster Session II

PB4804 Identification of novel variants responsible for monogenic diabetes in Tunisia using Whole Exome Sequencing : impact on diagnosis and treatment.

Authors:

N. KHERIJI[1,2,3], H. Dallali[1], I. Gouiza[1,2,3,4], M. Hechmi[1], F. Mahjoub[2,5,6], M. Mrad[2,3,7], A. Krir[7], A. Bahlous[7], M. Ben Ahmed[8], H. Jamoussi[9], R. Kefi[1,2]; [1]Lab. of BioMed. Genomics and Oncogenetics, Inst. Pasteur de Tunis, BP 74, 13 Place Pasteur, 1002 Tunis, Tunisia., Tunis, Tunisia, [2]Univ. of Tunis El Manar, El Manar I, 2092 Tunis, Tunisia., Tunis, Tunisia, [3]Faculty of Med. of Tunis, Tunis, Tunisia, Tunis, Tunisia, [4]Univ. of Angers, MitoLab Team, Unité MitoVasc, UMR CNRS 6015, INSERM U1083, SFR ICAT, France, Tunis, Tunisia, [5]Res. Unit UR18ES01 on "Obesity", Faculté de Médecine de Tunis, Tunis, Tunisia, [6]Natl. Inst. of Nutrition and Food Technology, 11 rue Jebel Lakhdar, Bab Saadoun, 1007 Tunis, Tunisia., Tunis, Tunisia, [7]Lab. of Clinical Biochemistry and Hormonology, Inst. Pasteur de Tunis, BP 74, 13 Place Pasteur, 1002 Tunis, Tunisia, Tunis, Tunisia, [8]Lab. of Clinical Immunology, Inst. Pasteur de Tunis, BP 74, 13 Place Pasteur, 1002 Tunis, Tunisia., Tunis, Tunisia, [9]Inst. Natl. de Nutrition de Tunis, Tunis, Tunisia

Introduction: Monogenic Diabetes (MD) are atypical form of diabetes that consist 3 to 6% of all cases of diabetes. This prevalence is underestimated due to the overlapping clinical features with type 1 and type 2 diabetes. Hence, genetic testing is the most appropriate tool for an accurate diagnosis. Studies carried out in Europe, Asia and the United States have reported clinical, phenotypic and genetic heterogeneity in MD. However, epidemiological studies of these atypical and syndromic forms of diabetes are scarce in Low- and Middle-Income countries, such as Tunisia. The genetic structure of Tunisian population is characterized by a rich and heterogeneous landscape resulting from the admixture of ethnic groups originating from North Africa, Europe, Near East and sub-Saharan Africa. The aim of this study is to identify genetic loci and causative mutations in Tunisian suspected MD patients using Whole Exome Sequencing “WES”. Through this study, we hope to aid clinicians in making better choices for the treatment of diabetic patients. **Materials and Methods:** WES was performed among 10 diabetic patients recruited from collaborating medical center. Inclusion criteria were: young age of diabetes onset ≤ 40 years old, Family history of diabetes in at least two generations and absence or low pancreatic autoantibodies titers. The pathogenicity of genetic variation was assessed using combined filtering and bioinformatics prediction tools. The online ORVAL tool has been used to predict the potential pathogenic variants combinations. Then, Sanger sequencing was carried out to confirm likely pathogenic predicted mutations among patients and to check for familial segregation. Finally, for novel variants, we completed structural modeling to study their impact on protein function. **Results:** We identified pathogenic variants located in *ABCC8, RFX6, GCKR, PDX1, KANK 1, PPP1R3A, TTC8, ALMS1* and *WFS1*genes. We highlighted the presence of syndromic forms of diabetes including Bardet-Biedl syndrome, Alström syndrome and Wolfram syndrome as well as the presence of isolated diabetes with considerably lower penetrance for symptoms associated with Wolfram syndrome. Idiopathic type 1 diabetes was also identified in one patient. **Conclusion:** In this study, we emphasized the importance of genetic screening for MD using WES in patients with familial history of diabetes among admixed and under-represented populations. An accurate diagnosis with molecular investigation of MD may improve the therapeutic choice for better management of patients and their families. Additional researches are required to better understand phyisiopathological mechanisms of MD.

Session Title: Mendelian Phenotypes Poster Session III

PB4805 Identification of UBE3A substrates through an *in vivo* ubiquitin activated interaction trap (UBAIT) approach in Drosophila

Authors:

B. Geier, L. Reiter; Univ. of Tennessee HSC, Memphis, TN

Autism spectrum disorder (ASD) is a complex and chronic condition associated with difficulties in social interactions, communication, and obsessive interests. Although the underlying causes of ASD remains elusive, duplications of the 15q11.2-q13.1 region, also known as the Prader-Willi/Angelman critical region (PWACR), are found in approximately 1-3% of all ASD cases, making this duplication one of the most common cytogenetic abnormalities found in ASD. Within the PWACR is the paternally imprinted gene *UBE3A*, which encodes for ubiquitin ligase E3A (UBE3A/E6AP). Mutations affecting *UBE3A* maternal expression levels can result in Duplication 15q (Dup15q) or Angelman syndrome (AS). Notably, both disorders have high rates of ASD and epilepsy, emphasizing the need for substrate identification. Previous efforts to pair cell culture with immunoprecipitation and mass spectrometry have thus far only identified a select few proteins as valid UBE3A ubiquitin substrates. Our lab has implemented an in vivo approach using the ubiquitin-activated interaction trap (UBAIT) in living *Drosophila melanogaster*. The UBAIT system covalently traps proteins ubiquitinated by *Drosophila* UBE3A (Dube3a), allowing for subsequent purification and identification via mass spectrometry. Analysis of UBAIT-purified peptides revealed enrichment and overlap in our global (*actin-GAL4*), neuronal (*nSyb-GAL4*), and glial (*repo-GAL4*) cell groups. Due to the high volume of Dube3a substrate candidates and physical time constraints, we focused our downstream validation studies on proteins with strong homology between *Drosophila* and humans that were in a size range more manageable for in vitro ubiquitination assays. We are currently verifying which candidate Dube3a substrates are poly-ubiquitinated in a Dube3a-dependent manner using both *in vitro* ubiquitination and GFP pulldown assays. We found that both synapse-associated protein 47kD (Sap47) and annexin B10 (AnxB10) are true Dube3a substrates in these assays. We will continue screening candidate fly proteins *in vitro* and subsequently validate their mammalian homologs. Not only will this research contribute to our understanding of the proteins dysregulated in Dup15q syndrome, AS, and ASD, but it will provide an improved resolution regarding the cell type(s) where UBE3A plays a functional role.

Session Title: Mendelian Phenotypes Poster Session I

PB4806 Identification of variants underlying biparietal perisylvian polymicrogyria in Finnish families

Authors:

I. Järvelä[1], R. A. Paetau[2], A. Acharya[3], **Y. Rajendran**[3], I. Schrauwen[3]; [1]Dept. of Med. Genetics, Univ. of Helsinki, Helsinki, Finland, [2]Dept. of Child Neurology, Univ. of Helsinki, Helsinki, Finland, [3]Ctr. for Statistical Genetics, Gertrude H. Sergievsky Ctr., and the Dept. of Neurology, Columbia Univ. Med. Ctr., New York, NY

Bilateral perisylvian polymicrogyria (BPP) is the most common form of regional polymicrogyria (PMG) that belongs to malformations of cortical development (MCDs). BPP is characterized by overfolding of the cerebral cortex and abnormal cortical layering. The major clinical features include oromotor dysfunction, dysarthria, and hemiparesis. Cognitive impairment and epilepsy are common. By now, more than 50 genes are reported to underlie PMG covering about 20% of cases. Mosaicism has been identified in some cases. The aim of this study was to identify genetic background of BPP in the Finnish families. A total of 22 Finnish families with BPP were enrolled in the study. There were 18 trios, three duos and one singleton case. In all cases brain imaging using MRI showed BPP. All participants were studied using exome sequencing (ES) (Illumina Inc, San Diego, CA) from peripheral blood DNA. Of them, eight ES negative cases were also analyzed using optical genome mapping (OGM) (Bionano Inc,San Diego). In addition, deep sequencing (200x) of buccal DNA samples was applied for nine cases negative for conventional ES and OGM. Buccal DNA was extracted using the Oragene assisted DNA collection kit (DNA Genotek) to identify mosaic variants that may be specific to ectodermal lineage cells. The results were confirmed by Sanger sequencing. To date, pathogenic (P) or likely pathogenic (LP) variants were found in 6/22 (27 %) patients using conventional ES. Of them, four were de novo/likely de novo (*SCN3A, TUBA1A, DDX23 and TUBB2B*), one autosomal recessive (*CCDC82*), one X-linked (*TAF1*) and one VUS (*AFF2*). In addition, two novel candidate genes were found (*RUFY4 and BOC*). Preliminary analyses of deep sequencing revealed a novel variant in *NEDD4L and one in SON* in two cases respectively. Full results will be presented at the meeting. The study was approved by the ethics committees of the Hospital District of Helsinki and Uusimaa and the Institutional Review Boards of Columbia University (IRB-AAAS3433). Written informed consent was obtained from healthy adult subjects and the parents/legal guardians of minors and study subjects. The authors report no conflict of interests. Grants: Helsinki University Central Hospital subsidiaries TLK0278 and TRTR019. NIH R21NS123325.

Session Title: Mendelian Phenotypes Poster Session II

PB4807 *IFIH1* gain of function variants in three individuals: The multifaced continuum of type I Interferonopathy

Authors:

S. Shimada; NIH, North Bethesda, MD

Background: Melanoma differentiation-associated gene 5 (MDA5), encoded by *IFIH1*, plays a critical role in activating type I interferon signaling and is essential for the immune response against viral infections. Gain-of-function variants in IFIH1 have been associated with type 1 interferonopathy, leading to the development of Aicardi-Goutieres syndrome (AGS) and Singleton Merten syndrome (SMS). The presence of overlapping symptoms between the neuroinflammatory phenotype of AGS and the systemic inflammatory phenotype of SMS suggests a shared disease spectrum. In this study, we report three individuals with clinical features that overlap between AGS and SMS, and provide a comprehensive review of previously reported cases to further elucidate the spectrum of type I interferonopathy. **Methods**: Extensive clinical investigations were conducted in the individuals. Whole-exome sequencing and research variant analyses were performed in three pedigrees to investigate the clinical features associated with variants in *IFIH1*. Segregation of variants was performed using Sanger sequencing. Additionally, in vitro analyses were conducted to measure the interferon activity of the identified variants in mammalian cells. **Results**: De novo variants in *IFIH1* were identified in all three individuals. The interferon activities observed in the variants without exogenous dsRNA ligand indicated their pathogenic nature; constitutive activation of MDA5 confirmed a gain-of-function mechanism. The clinical presentations among the probands showed variability, with one individual presenting a classical phenotype of AGS type 7 and the others displaying overlapping features of AGS and SMS. Notably, there was no strict correlation between the identified genotypes and the observed clinical phenotypes, suggesting the involvement of additional factors in disease severity and comorbidities. Furthermore, glycome abnormalities observed in a proband with early mortality may signify secondary findings related to the inflammatory response in Type I interferonopathies. **Conclusion**: This study emphasizes the clinical heterogeneity of type I interferonopathy and supports the concept that AGS type 7 and SMS represent different manifestations along a disease spectrum. A comprehensive understanding of the genotype-phenotype correlations and underlying disease mechanisms is crucial for improving the diagnosis and treatment strategies for individuals with IFIH1 variants. Further research is warranted to unravel the complexities of these disorders and identify potential therapeutic targets to mitigate the impact of type I interferonopathies.

Session Title: Mendelian Phenotypes Poster Session III

PB4808 Impact of processed pseudogene insertions in genetic testing as cause of monogenic diseases: insertion in CLCN1 gene causing Myotonia Congenita

Authors:

J. Pinto Basto[1], K. Tripolszki[1], J. Martini[1], K. Kampe[1], V. Karageorgou[1], M. Almuhaizea[2], C. Pereira[1], S. Weissgraeber[1], O. Paknia[1], P. Bauer[1]; [1]CENTOGENE, Rostock, Germany, [2]• King Faisal Specialist Hosp. & Res. Ctr. (KFSH & RC), Riyadh, Saudi Arabia

In this work, we report the first case of autosomal recessive myotonia congenita associated with a homozygous insertion of a processed pseudogene in the *CLCN1* gene. Myotonia congenita is a rare skeletal muscle channelopathy characterized by muscle stiffness and an inability of the muscle to relax after voluntary contraction (myotonia). It is caused by mutations of *CLCN1* gene (7q34), which encodes human skeletal muscle chloride channel 1. We report on a 3-year-old female patient, presenting with myotonia at the age of 14 months. Her parents are consanguineous, and she has a 13-year-old brother who is similarly affected. To identify the causative variant, genome sequencing (GS) was performed. During the analysis, single nucleotide variants, copy number variations and structural variants were also considered. PCR amplification and Sanger sequencing of the inserted region and the breakpoint were also performed to further investigate the identified structural variant. GS and Sanger sequencing were also performed for the affected sibling. Using GS, only one event remained as a good candidate. We identified a deep intronic homozygous insertion of ~750bp located in intron 14 of the *CLCN1* gene. The breakpoints were detected by the structural variant caller and further investigated by the visualization tool. The group of discordant reads at the breakpoint in intron 14 of *CLCN1* have mates that map to the *UQCRH* gene on chromosome 1. The sequence of the inserted region was obtained by targeted PCR amplification and Sanger sequencing and was determined to map exclusively to the coding region and UTR of the *UQCRH* gene. Since the inserted region showed the hallmarks of a retrotransposition event (antisense poly(A) tail, target site duplication, coding sequence of *UQCRH*), the retrotransposition of a processed mRNA was suspected. Considering that the inserted sequence is located in a deep intronic position, it is possible that splicing is impacted. The same event was identified in the affected brother in homozygous form further supporting the role of this variant as a genetic cause of the phenotype in this family. To our knowledge, this is the third reported case of a processed pseudogene insertion causing a monogenic disorder. Structural variant analysis in GS is challenging, time consuming and not regularly performed in the routine diagnostic testing. Untargeted screening of processed pseudogenes insertions is effectively performed by structural variant analysis in GS. This further emphasizes the role of GS as a first-tier approach in rare disorder diagnostics.

Session Title: Mendelian Phenotypes Poster Session I

PB4809 Impaired *EPHA4* pathway and neurodevelopmental central pattern generators in idiopathic scoliosis.

Authors:

X. Yang[1,2], L. Wang[1,2], S. Zhao[3,1], J. Liu[4], K. Xu[1], Q. Li[1], J. Cai[5], Z. Zhao[5], G. Li[6], P. Zheng[4], P. Liu[3], J. Posey[3], Z. Wu[7], G. Qiu[1], S. Ikegawa[8], J. Lupski[3], J. Rios[9], C. Wise[10], J. Zhang[1], C. Zhao[4], N. Wu[1]; [1]Peking Union Med. Coll. Hosp., Beijing, China, [2]Qilu Hosp. of Shandong Univ., Jinan, China, [3]Baylor Coll. of Med., Houston, TX, [4]Ocean Univ. of China, Qingdao, China, [5]Peking Union Med. Coll., Beijing, China, [6]Peking Union Med. Coll. Hosp. Chinese Academy of Med. Sci., Beijing, China, [7]Peking Union Med. Coll. Hosp., Chinese Academy of Med. Sci., Beijing, China, [8]RIKEN Ctr. for Integrative Med. Sci., Tokyo, Japan, [9]Texas Scottish Rite Hosp. for Children, Dallas, TX, [10]Scottish Rite for Children, Dallas, TX

Idiopathic scoliosis (IS) is the most common form of spinal deformity. Although many genetic risk loci have been identified, the pathophysiological mechanisms are poorly understood. In this study, we summarized IS-associated loci based on previous genetic association studies and mapped these loci to candidate genes using location-based or function-based strategies. Enrichment of rare variants in these candidate genes was determined in 411 Chinese individuals with severe IS (Cobb > 40°) versus 3,800 Chinese controls. A genome-wide association study (cases=6,449, controls=158,068) in East Asian population and an exome-sequenced European-ancestry IS cohort were included for validation. Zebrafish *epha4a/b* mutants, *efnb3b* morphants, and *ngef* crispants were used to explore the mechanism of *EPHA4* pathway-related IS. As a result, rare variants in *EPHA4* were significantly enriched in patients with IS (P=0.045, OR=4.09), with three rare coding variants identified in three patients. Two of the three variants (c.1443+1G>C and c.2546G>A, p.Cys849Tyr) were subsequently revealed to be *de novo* and could convey deleterious effects on *EPHA4*. Via searching for noncoding variants and structural variants, we found a *de novo* splicing variant (c.1318+10344A>G) in *EPHA4* and a *de novo* 4.46 Mb deletion that includes *EPHA4* from two additional patients with IS. In addition to rare variants, common variants in *EPHA4*, after aggregation, also showed significant enrichment (P=0.023) in the East Asian IS cohort. By investigating rare variants in other genes in the *EPHA4* pathway, we identified two dominantly inherited variants in *NGEF* (c.1A>G, p.Met1? and c.857C>T, p.Ala286Val), which encodes a downstream molecule in the *EPHA4* pathway. The *epha4a/b* mutant and *ngef* crispant zebrafish recapitulated the IS phenotype. In addition, perturbations of the *epha4* pathway led to disorganized neural patterning and abnormal swimming behavior of zebrafish, suggesting a dysfunction of the central pattern generators (CPGs). Therefore the impairment of neural patterning and CPGs may underlie the pathogenesis of IS.

Session Title: Mendelian Phenotypes Poster Session II

PB4810 Indian Siblings with Waardenburg Syndrome Living Beyond the Age of 40 Years: Diagnosis and Management

Authors:

M. Pradeep Kumar[1], M. Gomathi[2]; [1]Kovai Med. Ctr. and Hosp., Coimbatore, India, [2]Karpagam Academy of Higher Ed., Coimbatore, India

Waardenburg syndrome (WS) is a rare genetic disorder of neural crest cells inherited in autosomal dominant fashion with the prevalence of 1:50,000 live births. The hallmark features of WS are pigmentary defects, dysmorphic craniofacial features, upper limb abnormalities, neurological defects, intestinal anomalies and while some individuals may exhibit sensorineural hearing loss. Based on the phenotypic variations WS is divided into four subtypes WSI to IV and confirmed by genetic testing. There are limited cases reported in India. This article aims to report the rare WS siblings' diagnosis, treatment, and genetic counselling to their family members. We present our first case of 47 years old woman born as first child to a nonconsanguineous couple, without any pregnancy complications. She presented with severe developmental delay, short stature, microcephaly, intellectual disability, dsytopia canthorum, syndactyly, skeletal disabilities, language deficits, patch of white hair at the frontal area, light colored eyes, and Hirschsprung disease. The second case of 44 years old male sibling born as second child showed intellectual disability, aphasia, hypotonia, sensorineural hearing loss, Hyperkeratotic papules over limbs, brachydactyly, dislocated shoulders, hiatus hernia, and gastroesophageal reflux disease. MRI imaging studies showed cortical malformations and ventriculomegaly in both patients. There was no similar history in their parents or other siblings. According to Read and Newton, the first case was diagnosed as WS type 4 (shah WS) and the second case diagnosed as WS type 3 (Klein WS). The genetic testing confirmed the missense variant C>T at exon 7 of paired box 3 gene (PAX3) on chromosome 2 in both patients caused WS. (NM_181458:exon7:c.C1003T:p.P335S). This finding extends the prevailing knowledge of mutations of the PAX3 gene in WS3 and WS4 patients. This is considered as the first case report of type 4 of WS associated with PAX3 mutation in India. As it is a genetic disease, there is no definitive treatment for Waardenburg syndrome, but symptomatic treatments like surgery for skeletal abnormalities and intestinal disabilities. Since it is autosomal dominant disorder, other unaffected two siblings were undergoing carrier screening test and provided with genetic counselling. <!--EndFragment-->

Session Title: Mendelian Phenotypes Poster Session III

PB4811 Insights from a systematic cohort of 145 families with epilepsy in the Indian population.

Authors:

P. Majethia[1], N. Kaur[1], S. Mascarehnas[1], L. Rao[1], D. Narayanan[1], V. Bhat[1], K. Nair[1], B. Hunakunti[1], S. Nayak[1], S. Aroor[2], R. Bhat Y[2], B. Praveen[3], B. Yatheesha[4], S. Patil[5], S. Siddiqui[6], R. Kadavigere[7], S. Bielas[8], K. Girisha[1], S. Sharma[9], A. Shukla[1]; [1]Dept. of Med. Genetics, Kasturba Med. Coll., Manipal, Manipal Academy of Higher Ed., Manipal, India, [2]Dept. of Pediatrics, Kasturba Med. Coll., Manipal, Manipal Academy of Higher Ed., Manipal, India, [3]Dept. of Pediatrics, Father Muller Med. Coll. Hosp., Mangalore, India, [4]Paediatric neurology, Dheemahi Child Neurology and Dev. center, Shimoga, India, [5]Dept. of Paediatrics, Narayana Hrudayalaya Hosp./ Mazumdar-Shaw Med. Ctr., Bangalore, India, [6]Interventional Radiology, Neurology, Yashoda Hosp., Hyderabad, India, [7]Dept. of Radiodiagnosis, Kasturba Med. Coll., Manipal, Manipal Academy of Higher Ed., Manipal, India, [8]Dept. of Human Genetics, Univ. of Michigan Med. Sch., Ann Arbor, MI, [9]Neurology Div., Dept. of Pediatrics, Lady Hardinge Med. Coll. and Associated Kalawati Saran Children's Hosp., New Delhi, India

Introduction: Disorders with epilepsy are genetically and phenotypically heterogeneous. Advancement in genetic testing has led to rapid identification of underlying etiology in epilepsy with its emerging role in the field of precision medicine. **Methods:** A total of 145 families (153 individuals) with epilepsy with/without NDDs were recruited from October 2019 to January 2023 after obtaining informed consents from the family. Epilepsy syndrome classification was done as per the International League Against Epilepsy (ILAE) 2022 definition and classification of epilepsy syndromes. Targeted and/or genomic testing was carried out after detailed clinical evaluation. Genetic counseling and therapy were evaluated for individuals with a definitive molecular diagnosis. **Results:** Our cohort comprises of 86 (56%) males and 67 (44%) females from 145 unrelated families. Consanguinity was noted in 38% (55/145) of the families. According to the recent ILAE nosology and classification, epilepsy syndrome classification could be made in 63 of 145 families (43%). Fifty-three families had a diagnosis of neonatal and infantile-onset syndromes. Of these, six were defined as genetic epilepsy with febrile seizure plus, 20 as infantile epileptic spasm syndrome, 18 as early-infantile DEE, four as epilepsy of infancy with migrating focal seizures, four as Dravet syndrome and one with etiology-specific DEE. Among the five families with childhood-onset syndromes, one was diagnosed with epilepsy with myoclonic-atonic seizures, two with Lennox-Gastaut syndrome, and two with epileptic encephalopathy with spike-and-wave activation in sleep. The other five families fell in the group of progressive myoclonic epilepsies. Definitive molecular diagnosis was achieved in 79 out of 145 families (55%): five (6%) by targeted testing, three by chromosomal microarray (4%), twelve (16%) by Mendeliome, and 59 (74%) by exome sequencing (ES). Monogenic disorders were identified in 71 families, and imprinting and microdeletion disorders in four families each. Forty-three (57%) of 75 disease-causing variants were novel. Two novel disease-gene associations with pathogenic variants in *GCSH* and *KCNH5* were identified. Reproductive counseling was offered to 91% (72/79) of families. Therapeutic implications were noted for 27% of individuals (21/79) with a definitive diagnosis.**Conclusion:** We herein present a systematic cohort elucidating the epilepsy syndrome classification and its genetic landscape observed in the Indian population. We emphasize the utility of ES as an effective first-tier test with substantial potential for precision medicine in epilepsies.

Session Title: Mendelian Phenotypes Poster Session I

PB4812 Integrating genetic, demographic, and clinical variables to predict neurodevelopmental outcomes in children with congenital heart defects.

Authors:

E. Hernandez[1], T. A. Miller[1], M. W. Russell[2], J. W. Newburger[3], W. K. Chung[4], W. Gaynor[5], E. Goldmuntz[6], J. F. Cnota[7], M. Yandell[1], M. Tristani-Firouzi[1]; [1]Univ. of Utah, Salt Lake City, UT, [2]Univ. of Michigan, Ann Arbor, MI, [3]Harvard Med. Sch., Boston, MA, [4]Columbia Univ., New York, NY, [5]Children's Hosp. of Philadelphia, Philadephia, PA, [6]Children's Hosp. of Philadelphia, Philadelphia, PA, [7]Cincinnati Children's Hosp., Cincinnati, OH

Congenital heart defects (CHD) have been a subject of intense investigation, with recent large-scale sequencing efforts shedding light on the genetic factors underlying their etiology. However, the precise impact of genetics on clinical outcomes remains less understood. In this study, we employed Bayesian Networks, an explainable Artificial Intelligence approach, to unravel the complex relationships between clinical variables, demography, and genetics in a cohort of children with single ventricle CHD (N=304). Our findings demonstrate the interconnected nature of these variables and their combined impact on clinical outcomes. As isolated variables, a damaging genetic variant in a gene related to abnormal heart morphology and prolonged ventilator support following cardiac surgery increased the probability of having a low Mental Developmental Index (MDI) score at 14 months of age by 1.9- and 5.8-fold, respectively. However, when present in combination, these variables synergistically elevated the probability of a low MDI score by 10-fold. Genetic information also had predictive value for favorable neurodevelopmental outcomes. For instance, the absence of a damaging variant in a known syndromic CHD gene and a shorter post-operative ventilator support were associated with a 1.7- and 2.4-fold increase in the probability of a normal MDI score, respectively. When combined, these factors substantially increased the likelihood of a favorable outcome by 59-fold. Our analysis underscores the modest genetic contribution to neurodevelopmental and growth outcomes as isolated variables, with an even greater impacts when considered in the context of specific clinical variables. These results emphasize the importance of comprehensively capturing and quantifying the effects of damaging genomic variants within the context of multiple conditionally dependent variables, including pre- and post-operative factors, as well as demography. This study deepens our understanding of the intricate interplay between genetics and clinical outcomes in CHD and lays the foundation for personalized patient management.

Session Title: Mendelian Phenotypes Poster Session II

PB4813 Integration of genomic testing for evaluation of a cohort of 600 families with neurodevelopmental disorders from India.

Authors:

A. Shukla[1], P. Majethia[2], L. P. Rao[1], S. Pande[3], N. Kaur[1], S. Mascarenhas[1], M. C. Rosario[1], K. V. Nair[1], B. Hunakunti[1], A. Prasannakumar[1], D. L. Narayanan[1], S. Aroor[1], R. Y. Bhat[1], L. Lewis[1], S. Sharma[4], P. BK[5], Y. BL[6], S. J. Patil[7], N. Kamath[8], S. Nampoothiri[9], S. Bajaj[10], S. V H[11], K. G. Bhat[12], H. Prasad[12], S. Bijarnia-Mahay[13], H. Shah[1], R. Kedambadi[14], A. Kumble[15], S. Soans[16], V. Kumar[1], N. Sankhyan[17], R. K R[18], A. Pai[1], D. B. Jasti[1], S. Siddiqui[19], R. Kadavigere[1], M. L. Kulkarni[20], R. D. Puri[21], K. Girisha[2], S. Bielas[22]; [1]Kasturba Med. Coll., Manipal, MAHE, Manipal, Manipal, Udupi, India, [2]Kasturba Med. Coll., Manipal, Udupi, India, [3]Kasturba Med. Coll., Manipal, Manipal Academy of Higher Ed., Manipal, India, Manipal, India, [4]Lady Hardinge Med. Coll. and Associated Kalawati Saran Children's Hosp., New Delhi, India, New Delhi, India, [5]Father Muller Med. Coll. Hosp., Mangalore, India, Mangalore, India, [6]Dheemahi Child Neurology and Dev. center, Shimoga, Karnataka, India, Shivamoga, India, [7]Narayana Hrudayalaya Hosp./ Mazumdar-Shaw Med. Ctr., Bangalore, India, Bengaluru, India, [8]Kasturba Med. Coll., Mangalore, Manipal Academy of Higher Ed., Manipal, India, Mangalore, India, [9]Amrita Inst. of Med. Sci. & Res. Ctr., Cochin, India, Cochin, India, [10]The Purple Gene Clinic, Mumbai, India, Mumbai, India, [11]SAT Hosp., Govt. Med. Coll., Trivandrum, India, Trivandrum, India, [12]Kasturba Med. Coll., Mangalore, MAHE, Manipal, Mangalore, India, [13]Inst. of Med. Genetics and Genomics, Sir Ganga Ram Hosp., New Delhi, India, New Delhi, India, [14]First Neuro Brain & Spine Super Speciality Hosp., Mangalore, India, Mangalore, India, [15]Indiana Children Inst., Indiana Hosp., Mangalore, India, Mangalore, India, [16]A.J. Hosp. and Res. Ctr., Mangalore, India, Mangalore, India, [17]Advanced Pediatrics Ctr., Postgraduate Inst. of Med. Ed. & Res., Chandigarh, IndiaKasturba Med. Coll., Manipal, MAHE, Manipal, Chandigarh, India, [18]Narayana Multispecialty Clinic, Shimoga, Karnataka, India, Shimoga, India, [19]Dept. of Neuroimaging and Interventional Radiology, STAR Inst. of NeuroSci.s, STAR Hosp., Hyderabad, India, Hyderabad, India, [20]Jagadguru Jayadeva Muragharajendra (JJM) Med. Coll., Davangere, Karnataka, India., Davangere, India, [21]Inst. of Med. Genetics & Genomics, Sir Ganga Ram Hosp., New Delhi, India, New Delhi, India, [22]Univ. of Michigan, Ann Arbor, MI

Introduction: The considerable size and complex structure of the Indian population produces a profound burden of inherited neurodevelopmental disorders (NDDs). This poses a huge challenge in health care, as India continues to be classified as a lower-middle-income country (LMIC). **Methods:** We recruited a total of 600 families with NDDs, consented under an IRB with appropriate ethical approvals starting in fall of 2019. Deep phenotyping was collected and genomic evaluation has been completed for 523 families (588 individuals). An exome first or a sequential genetic testing strategy was used, in which targeted tests or chromosomal microarray (CMA) was followed by exome sequencing (ES)/genome sequencing.**Results:** Of the 523 families which have undergone a full evaluation, consanguinity was noted in 188 (36%) families. Notably, 546 (93%) affected individuals are of pediatric age group. Disease-causing variants were detected in 321 families (61%): 29 families by targeted testing (9%), 16 by CMA (5%), 52 by Mendeliome (16%), 190 by solo ES (59%), 3 by duo ES (1%), and 31 by trio ES (10%). Overall, 230 monogenic disorders were identified in 291 families (91%) and 25 chromosomal disorders in 30 families (9%). The spectrum of monogenic disorders comprised of 31 families with epilepsy as a core symptom (10%), 60 with leukodystrophies (19%) including 28 families with hypomyelinating disorders, 32 with metabolic disorders (10%), 63 with syndromic intellectual disability (ID, 20%) including 41 families with *de novo* syndromic ID, 28 with developmental epileptic encephalopathy (9%), 23 with mitochondrial disorders (7%), 15 with neuromuscular disorders (5%), 7 with hereditary spastic paraplegias (2%) and 32 with other heterogeneous NDDs (10%). Significant phenotypic and genotypic insights of the cohort include (i) Identification of either novel or extremely rare disease-causing variants in 20 genes: *GCSH, PIDD1, TMEM163, KCNH5, FILIP1, SNRPN, DCHS2, CLIP1, TRIM23, SLC25A10, KCNK18, GPAT, NDUFB6, MRPL49, CSMD3, UNC50, FAF1, EPB41L3, VPS36, FAM160B1* (ii) Fifty-three percent novel disease-causing variants (160/305), thus contributing to population specific variants (iii) Identification of probable founder variants (*ISCA1, RNASEH2C, CYB5R3*).**Conclusion:** This study delineates a systematic cohort of deep phenotyping and pertinent use of targeted and broad-spectrum genomic tests to decipher the etiology of inherited NDDs. A definite genetic diagnosis resulted in informed genetic counselling, management, and prevention of known disorders as well as deciphering novel phenotypes and genotypes in a lesser explored population.

Session Title: Mendelian Phenotypes Poster Session III

PB4814 † Integrative transcriptomic network analysis of human iPSC-derived osteogenic differentiation identifies *KLF16* as a key regulator of bone formation.

Authors:

M. Wu[1], Y. Ru[1], M. Ma[2], X. Zhou[1,3,4], D. Kriti[1,5], N. Cohen[1,6], S. D'Souza[7,8], C. Schaniel[7,9], S. M. Motch Perrine[10], S. Kuo[10,11], D. Pinto[1,4,12], G. Housman[13,14], G. Holmes[1], E. Schadt[1,4], H. van Bakel[1,4,15], B. Zhang[1,3,4], E. W. Jabs[1,7]; [1]Dept. of Genetics and Genomic Sci., Icahn Sch. of Med. at Mount Sinai, New York, NY, [2]GeneDx, Stamford, CT, [3]Mount Sinai Ctr. for Transformative Disease Modeling, Icahn Sch. of Med. at Mount Sinai, New York, NY, [4]Icahn Genomics Inst., Icahn Sch. of Med. at Mount Sinai, New York, NY, [5]Present address: Dept. of Biochemistry and Molecular Biology, Univ. of British Columbia, Vancouver, BC, Canada, [6]Present address: Div. of Cytogenetics and Molecular Pathology, Zucker Sch. of Med. at Hofstra/Northwell, Northwell Hlth.Lab., Lake Success, NY, [7]Dept. of Cell, Dev. and Regenerative Biology, Icahn Sch. of Med. at Mount Sinai, New York, NY, [8]Present address: St Jude Children's Res. Hosp., Memphis, TN, [9]Dept. of Med., Div. of Hematology and Med. Oncology, Tisch Cancer Inst., Inst. of Regenerative Med., Black Family Stem Cell Inst., Icahn Sch. of Med. at Mount Sinai, New York, NY, [10]Dept. of Anthropology, Pennsylvania State Univ., University Park, PA, [11]Present address: Dept. of BioMed. Sci., Univ. of Minnesota, Duluth, MN, [12]Dept. of Psychiatry, Icahn Sch. of Med. at Mount Sinai, New York, NY, [13]Section of Genetic Med., Dept. of Med., Univ. of Chicago, Chicago, IL, [14]Dept. of Primate Behavior and Evolution, Max Planck Inst. for Evolutionary Anthropology, Leipzig, Germany, [15]Dept. of Microbiol., Icahn Sch. of Med. at Mount Sinai, New York, NY

Understanding the molecular mechanisms involved in physiological and pathological bone formation requires knowledge of global transcriptional changes during osteogenesis. Previous studies of osteogenesis have identified the involvement of various types of genes, such as those that encode for transcription factors (TFs) and long noncoding RNAs (lncRNAs), as well as alternative spliceoforms and multiple signaling pathways that include TGF-β/BMP, Notch, Hedgehog, NELL-1, PTHrP, FGF, WNT, HIPPO, JNK, MAPK, and estrogen receptor signaling. Here, we differentiated mesenchymal stem cells (MSCs), derived from normal human induced pluripotent stem cells, into preosteoblasts and osteoblasts and performed bulk RNA-seq from 60 samples and single-cell RNA-seq data analysis. We identified distinct and temporally dynamic gene expression profiles that are enriched for various biological functions and gene expression patterns of TF and lncRNA genes that are highly correlated during osteogenic differentiation. Expression levels of lncRNAs were found to be significantly correlated with those of their neighboring TFs, splicing factors, and protein-coding genes, which in turn are known to regulate skeletal development, ossification, and phosphorus and phosphate metabolism. Importantly, we identified *KLF16* as a key TF using TF-TF regulatory network analysis and as a key driver gene using gene co-expression network analysis. Overexpression of *Klf16* inhibited matrix maturation and mineralization during MC3T3-E1 osteogenic differentiation. Moreover, *Klf16* haploinsufficient mice showed increased whole-body bone mineral density, femoral bone volume to total tissue volume, trabecular bone thickness, and cortical bone area, which supported our predicted TF and co-expression network modules and confirmed the inhibitory role of *KLF16* in bone formation. *KLF16* previously has been identified as one TF of a set of TFs associated with osteoporosis by comparing microarray data of bone marrow MSCs from osteoporosis patients and controls (Liu *et al*., 2019. PMID: 31007729). Our approach serves as a model system for understanding and identifying candidate genes involved in normal and pathological osteogenic differentiation and developing more effective therapeutic strategies for bone diseases.

Session Title: Mendelian Phenotypes Poster Session I

PB4815 Investigating the role of *RAB5A* variants in bipolar disorders: Insights from Drosophila models

Authors:

B. Adhikari[1,2], J. Andrews[1,2], M. Wangler[1,2], J. Holder[1,2]; [1]Baylor Coll. of Med., Houston, TX, [2]Jan and Dan Duncan Neurological Res. Inst., Houston, TX

Background: Studies of neuropsychiatric diseases are challenging due to the lack of available models that accurately replicate the complexity of these disorders. Recently, genetic variants in *RAB5A* have been associated with bipolar disorder (BPD). RAB5A, a protein in the small G-protein family, plays a crucial role in intracellular vesicle trafficking. In this study, we developed *Drosophila* models expressing *RAB5A* BPD-associated variants to investigate their functional roles. We hypothesized that *RAB5A* variants would cause morphological, behavioral, and survival abnormalities in flies expressing them.Methods:*Drosophila* overexpression models were developed with multiple *RAB5A* BPD-associated variants using the UAS/Gal4 system to drive the expression of the human cDNA. Nubbin-Gal4, and Elav-Gal4 drivers were used to express *RAB5A* in wing tissues and the central nervous system (CNS), respectively. To achieve ubiquitous expression, Actin-Gal4, and Tubulin-Gal4 drivers were utilized. Lethality was assessed by comparing the number of observed eclosing adults to the number of expected progeny (O/E). Motor dysfunction was evaluated through climbing assay, while survival analysis was conducted and analyzed using the log-rank test in Kaplan-Meir plot. Results: We observed that ubiquitous expression of $RAB5A^{D136N}$ led to complete lethality (O/E=0), while expression in wings resulted in partial lethality (O/E=0.24). However, when expressed in the CNS, $RAB5A^{D136N}$ expression did not induce lethality (O/E=1.39). Surviving Nubbin-Gal4; UAS- $RAB5A^{D136N}$ progeny showed significantly altered wing morphology, including modified wing shape, missing anterior and posterior cross-veins, thickening of the longitudinal veins with the appearance of extraneous vein formation. Also, old Nubbin-Gal4 flies (>10 days) showed a significant increase in wing damage than age-matched controls. Elav-Gal4 flies showed significant motor dysfunctions upon agitation than controls, which increased with age. We assessed the lifespan of $RAB5A^{D136N}$ flies and observed a decrease in survival compared to wild-type flies (Log-rank test, P=<0.0001).Conclusion: Our study indicates that $RAB5A^{D136N}$ when expressed ubiquitously, leads to lethality and targeted overexpression within wing tissues results in significant alterations in wing morphology. Furthermore, its expression in the CNS does not induce lethality but contributes to motor dysfunction. These findings provide functional validation that *RAB5A* variants associated with BPD may be deleterious and highlight the potential of Drosophila models for further investigation of BPD and its underlying mechanisms.

Session Title: Mendelian Phenotypes Poster Session II

PB4816 Investigating the role of Rbfox2 in Splicing and the Transcriptional Network of the SWI/SNF Complex in myoblast fusion

Authors:

T. Govindarajan, S. Kawamoto; Natl. Heart, Lung, and Blood Inst. (NHLBI), Bethesda, MD

The coordinated expression of myogenic transcription factors and splicing regulators is crucial in cardiac and skeletal muscle growth, repair, and development. Rbfox2, a member of the Rbfox family proteins, regulates alternative splicing, mRNA stability, and microRNA biogenesis in cardiac and skeletal muscle. Using the BioID proximity approach, we identified novel protein interactors that are specifically bound proximal to Rbfox2 isoforms, enabling us to understand better the shared and distinctive roles of each isoform in skeletal muscle development. This study is particularly important in investigating the non-splicing functions of Rbfox2 isoforms and associated signaling pathways in skeletal muscle development.

Session Title: Mendelian Phenotypes Poster Session III

PB4817 Is*JKAMP*gene associated with autosomal recessive “intellectual disability, autism spectrum disorder and seizures?” A Compound heterozygosity for two VUSs of *JKAMP* gene in a 22 yo female with global developmental delay, profound ID, ASD, and seizures.

Authors:

M. Hajianpour[1], J. Lundberg[1], L. Cortez[1], R. J. Simmons[2]; [1]Albany Med. Coll./ Albany Med Hlth.System, Dept. of Pediatrics, Div. of Med. Genetics and Genomics, Albany, NY, [2]Community Care Physicians, Clifton Park, NY

Is *JKAMP* gene associated with autosomal recessive “intellectual disability, autism spectrum disorder and seizures?” A Compound heterozygosity for two VUSs of *JKAMP* gene in a 22 yo female with global developmental delay, profound ID, ASD, and seizures. Proband is a 22-year-old female with a history of prematurity (34 weeks gestation, weighing 6lbs), regressed speech and language at ~3 years of age. She can currently say "baba, I want" only. She has severe autism with occasional aggression, profound intellectual disability (functioning at a 2 yo child), hypotonia (noticed at 7months of age) and seizures (started at 8 months of age) with a normal EEG and brain CT scan (at 8m), and normal repeat EEG with brain MRI (at 19m). On physical examination at 22 years of age, her weight and height were in normal range with a BMI of 20.44 kg/m^2. Her head circumference was in low normal range (~10%ile for age). She had some dysmorphic features, such as flat occiput, receding forehead, low anterior and posterior hairlines, synophrys, long eye lashes; deep-set eyes; overfolded upper and lateral helix; uplifted and fleshy earlobes; low nasal bridge; micrognathia, moderate retrognathia; tapering fingers; pes cavus; talipes calcaneovarus; and mild 2-3 toes syndactyly. She is toe walking with an unsteady gait. **GENETIC WORK-UP:** Routine chromosome analysis, Chromosome microarray analysis, Molecular studies for Rett syndrome and Fragile X syndrome, Sub-telomeric FISH, Plasma amino acids, Urine organic acids, Acylcarnitine profile, Mitochondrial genome sequencing with del/dup analysis were all Normal. **Whole exome sequencing (WES)** TRIO showed two variants of uncertain significance (VUSs) in the JKAMP gene: **c.443 A>G/ p.(E148G) Pat & c.878del/ p.(F293Sfs*34) Mat. DISCUSSION:** The JKAMP gene is a candidate gene with a potential relationship to the phenotype in our proband. This gene plays a role in regulating Jun N-terminal kinase (JNK) activity. A homozygous frameshift variant has been reported in a family with two affected siblings in association with intellectual disability, epilepsy, obsessive-compulsive disorder, autism, and dysmorphic features. - These Variant have not been observed at significant frequency in large population cohorts (gnomAD). - They have not been previously published as pathogenic or benign to our knowledge. - In silico analysis supports that these variants have a deleterious effect on protein structure/ function. **CONCLUSION:** Further research is needed to explore a possible link between the identified variants in the JKAMP gene and human disease, and/or association of Kemp gene with autosomal recessive intellectual disability, ASD, and seizures.

Session Title: Mendelian Phenotypes Poster Session I

PB4818 Joint analysis of phenotypically diverse rare disease cohort reveals new diagnostic genes

Authors:

M. Moldovan[1], S. Kobren[1], R. Reimers[2], X. Li[1], D. Traviglia[1], J. Willett[3], J. Krier[4], I. Kohane[1], Undiagnosed Diseases Network Tool Building Coalition Working Group, S. Sunyaev[1]; [1]Harvard Med. Sch., Boston, MA, [2]Rady Children's Inst. for Genomic Med., San Diego, CA, [3]McGill Univ., Montreal, QC, Canada, [4]Atrius Hlth., Boston, MA

Here, we harmonized and jointly called whole genome variants from a cohort of ~4,300 patients with diverse sets of symptoms enrolled in the Undiagnosed Diseases Network (UDN). Unlike previous cohort-based analyses which focus on specific disease categories, the UDN accepts any undiagnosed patients experiencing a broad range of pediatric- and adult-onset symptoms. In addition to genomic sequencing, patients undergo thorough, on-site clinical evaluations, resulting in a wealth of clinical data available for analysis. We inferred the functional impacts of coding variants using CADD scores and intronic variants using SpliceAI scores coupled with experimental data from a massively parallel splicing assay. We identified four genes across twelve patients with a significant excess of deleterious de novo variants, recapitulating seven known diagnoses and highlighting two new diagnoses. In addition, we developed a statistical approach to identify compound heterozygous variants in individual patients and uncovered four known and two novel gene-disease associations. In our analyses, we employ a highly accurate mutational model with base-pair resolution, which increases our statistical power. Finally, by considering genes harboring deleterious variants and applying gene set enrichment analysis on phenotypically-similar patient subgroups, we identified recurrently perturbed pathways, four of which covered seventeen distinct genes across patients, recapitulating six known diagnoses.Our pipeline for identification of putative disease-causing variants includes expert clinical evaluations of the concordance between candidate diagnostic variants and affected patient presentations. To streamline and standardize this process, we developed an objective, semi-automated, quantitative approach to reduce manual case review time. When applied to candidate diagnostic variants discovered in our analyses, clinical evaluations revealed high correlation between identified gene candidates and diagnostic feasibility. Taken together, these results demonstrate that comprehensive, cohort-based analyses of diverse, suspected Mendelian disorders can provide valuable insights into new disease genes and pathways, ultimately paving the way for improved diagnostics and targeted therapeutics.

Session Title: Mendelian Phenotypes Poster Session II

PB4819 Language and neural networks in Smith Magenis Syndrome.

Authors:

X. Tran[1], A. Gropman[2]; [1]Children's Natl. Hosp., Washington, DC, [2]Children s Natl. Hosp., Washington, DC

Background: Smith Magenis Syndrome (SMS) is caused by a heterozygous deletion in chromosome 17p11.2, or mutation of the RAI1 gene. SMS affects 1:15,000 to 1:25,000 in the US and can have a highly variable presentation, often associated with mild-moderate intellectual disability, and speech/language delay (Smith *et al* 2001, Gropman *et al* 2006, Le Gouard *et al* 2021). It can present with concomitant autism spectrum disorder (ASD), behavioral and sleep disturbances, seizure, hypotonia, and distinct facial and body dysmorphologies. SMS is associated with generalized epileptiform activity (Goldman *et al* 2006), cerebellar vermis hypoplasia, thin corpus callosum and brain stem, and decreased gray matter (Boddaert *et al* 2004, Maya *et al* 2014). **Objectives:** Aim 1. Characterize language profiles in children with SMS. Aim 2. Characterize baseline neural network activity using clinical electroencephalogram (EEG). **Methods:** Subjects with SMS who had existing EEG at Children's National Hospital were selected. Behavioral profiles were obtained based on chart review. Raw EEG files were located through NeuroWorkBench, visually reviewed and exported using Persyst. EEG data files were imported into MATLAB and processed using EEGLAB (Delorme & Makeig, 2004). Artifact subspace reconstruction and Adaptive Mixture Independent Component Analysis were used to remove artifacts. Fast Fourier Transfer was applied to the cleaned EEG data. Spectral power was calculated via Welch's method and phase coherence was calculated using EEGLAB *newcrossf* function. **Results:** 5 subjects with SMS were identified, all had behavioral profiles in line with the previously described clinical spectrum of SMS (seizure, language delay, cognitive delay, ADHD, ASD, self-injurious behaviors, sleep disorder, and scoliosis). Among the 3 subjects with EEG, 1 had EEG with bioccipital epileptiform activity, while 2 had EEG without epileptiform activity. Group level analysis was not performed because raw EEG data was only available for 1 subject. Brain MRIs were notable for centrum semiovale dysplasia, abnormal myelin sheath formation, and abnormal brainstem segmentation. **Discussion:** Due to different underlying structural abnormalities seen on each subject's MRI, baseline EEGs are expected to differ in spectral power and connectivity between subjects. Source-based localization, involving co-registering each subject's EEG to their own brain MRI, may be a more optimal analysis technique compared to generalized group-level analysis. To expand our study, we are in the process of acquiring EEG and MRI from the National Institute of Health's pre-existing SMS natural history study.

Session Title: Mendelian Phenotypes Poster Session III

PB4820 Lateral meningocele syndrome without lateral meningoceles: A case report to expand the phenotype.

Authors:

D. Rubadeux[1], J. W. Owens[2], A. Shillington[2]; [1]Univ. of Cincinnati Coll. of Med., Cincinnati, OH, [2]Cincinnati Children's Hosp., Cincinnati, OH

Lateral meningocele syndrome (LMS), also known as Lehman syndrome, is caused by pathogenic variants in exon 33 of *NOTCH3*. Variants in this final exon of *NOTCH3* interrupt the regulatory PEST domain, leading to enhanced NOTCH3 signaling due to prolonged cellular half-life. Patients with LMS are expected to have multiple lateral meningoceles, developmental delay, neonatal hypotonia, dysmorphic facial features, and feeding difficulties. Common findings include ligamentous laxity, short stature, hernias, cryptorchidism, congenital cardiac abnormalities, and hearing loss. Less common findings include proptosis, Wormian bones, paresthesia, urinary incontinence, and skin hyperextensibility. To date, 20 cases of LMS have been reported, with approximately half having completed confirmatory genetic testing.

We report an 8-year-old male patient with a history of prenatal polysubstance exposure to benzodiazepines and opiates, autism, cerebral palsy, feeding difficulties, severe developmental delay, intellectual disability, and self-injurious behavior. He received an Autism/ID Xpanded gene panel at 7 years of age, which revealed a c.6663C>G (p.Y2221*) pathogenic variant in exon 33 of *NOTCH3*. Following diagnosis of LMS, a spine MRI at age 7 showed a ventral sacral extradural arachnoid cyst but no lateral meningoceles. The patient's most recent exam noted multiple dysmorphic features including prominent metopic ridging, broad forehead, down-slanting palpebral fissures, high arched palate, long narrow philtrum, mild pectus excavatum, and wide-based gait. His stature is short for his age (117.5 cm, 1%ile), with a weight of 26.4 kg (43%ile).

Here, we document the 21st reported individual with LMS and the 12th case of molecularly confirmed LMS**.** The patient shares the dysmorphic facial features, ongoing G-tube dependence, failure to thrive, developmental delay, and congenital cardiac abnormalities seen in other individuals with LMS. His overall functioning is lower than other reported LMS cases, including a 9-year-old patient without intellectual disability who shares this p.Y2221* variant, which may be related to our patient's prenatal polysubstance exposure. However, the additional features that cannot be explained by environmental factors suggest significant variable expressivity associated with truncating variants in exon 33 of *NOTCH3*. His lack of lateral meningoceles expands the phenotype for this condition, as all previously reported patients with molecularly confirmed LMS had multiple lateral meningoceles before the age of 8, with the average age of identification being approximately 4 years.

Session Title: Mendelian Phenotypes Poster Session I

PB4821 † Leigh Syndrome Spectrum: Insights from the Largest Genetically Confirmed Pediatric International Cohort

Authors:

L. Schlieben[1,2], M. Shimura[3], N. Konstantinovskiy[1], S. L. Stenton[4,5], M. Brugger[1], F. Distelmaier[6], P. Freisinger[7], M. Hempel[8], G. Hoffmann[9], C. Makowski[10], J. Mayr[11], K. Murayama[3], C. Schlein[12], M. Schülke[13], A. Rötig[14], S. Wortmann[11], F. Fang[15], H. Prokisch[1,2]; [1]Sch. of Med., Inst. of Human Genetics, Klinikum rechts der Isar, Technical Univ. of Munich, Munich, Germany, [2]Inst. of Neurogenomics, Computational Hlth.Ctr., Helmholtz Zentrum München, Munich, Germany, [3]Chiba Children's Hosp., Ctr. for Med. Genetics and Dept. of Metabolism, Chiba, Japan, [4]Broad Inst. of MIT and Harvard, Boston, MA, [5]Program in Med. and Population Genetics, Broad Inst. of MIT and Harvard, Cambridge, MA, [6]Univ. Children's Hosp., Heinrich-Heine-Univ. Düsseldorf, Dept. of Gen. Paediatrics, Neonatology and Paediatric Cardiology, Düsseldorf, Germany, [7]Hosp. Reutlingen, Dept. of Paediatrics, Reutlingen, Germany, [8]Univ. Hosp. Heidelberg, Inst. of Human Genetics, Heidelberg, Germany, [9]Div. of Neuropaediatrics and Paediatric Metabolic Med., Ctr. for Paediatric and Adolescent Med., Univ. Hosp. Heidelberg, Heidelberg, Germany, [10]Dept. of Epileptology, LMU Ctr. for Dev. and Complex Chronically Ill Children, Dr. von Haunersches Kinderspital, Munich, Germany, [11]Salzburger Landeskliniken, Paracelsus Med Univ, Salzburg, Austria, [12]Inst. of Human Genetics, Univ. Med. Ctr. Hamburg-Eppendorf, Hamburg, Germany, [13]Charité Campus Virchow Klinikum, Berlin, Germany, [14]Univ. Paris Cité, Imagine Inst., Lab. of Genetics of Mitochondrial Diseases, INSERM UMR 1163, Paris, France, [15]Dept. of Neurology, Natl. Ctr. for Children's Hlth., Beijing Children's Hosp., Capital Med. Univ., Beijing, China

Introduction: Leigh syndrome spectrum (LSS) is the most frequent clinical manifestation of pediatric mitochondrial disorders (MD) characterized by neurodegeneration, bilateral symmetrical lesions in the basal ganglia and/or brain stem and impaired oxidative phosphorylation. Defects in more than 100 genes spanning the nuclear and mitochondrial genome have been associated with LSS, suggesting a high level of genetic heterogeneity. This study aimed to provide new insights into natural history of LSS by investigating the largest genetically diagnosed LSS collection to date.Methods: Clinical, survival, onset, and neuroimaging data of 527 genetically confirmed LSS patients from China, Japan, Germany, and France were reviewed. A gradient boosting tree-based machine learning classifier was developed to predict 3-year survival.Results: Pathogenic variants were identified in 124 distinct genes, with 57% of cases showing autosomal recessive, 37% maternal, 4% X-linked, and 2% autosomal dominant inheritance. Most frequent MD genes causing LSS were *MT-ATP6*, *MT-ND3*, and *MT-ND6*, while the most common nuclear genes were *ECHS1, SURF1*, and *PDHA1*. Overall, 20% of the 527 LSS patients were deceased at the timepoint of last assessment. Median age of death was 2 years (range= 0.1-14 years), with 60% of patients dying within the first three years of life. Median age of onset was 0.75 years (range= 0-11 years) and patients were in median followed up for 40 months (range= 0-380 months). Time from onset to death was in median 1.2 years (range= 0-13 years). Phenotypic analysis revealed 280 distinct Human Phenotype Ontology (HPO) terms, with a median of 9 HPO terms (range: 3-35) per patient. Global developmental delay, hypotonia and lactic acidosis were the most frequent phenotypes. Poorest 5-year survival was associated with defects in *MT-ND5, HSD17B10, MT-ATP6* (m.8993T&gtC/G, m.9176T&gtC), and *ALDH5A1* (≤50%). The developed model predicting survival showed an average prediction accuracy of 86% and outperformed genotype, clinical and MRI information alone. Most important features of the model were age of onset, increased MRS lactate levels as well as lesions in medulla oblongata and caudate nucleus.Discussion: This study presents findings from the largest genetically defined, systematically phenotyped LSS cohort, providing natural history of LSS. The remarkable accuracy and performance of our model provide a favourable and assuring indication of its potential in guiding clinical trial design and decision-making. Finally, the predicted survival by our model offers a promising outcome measure that serves as important guidance for future clinical trials.

Session Title: Mendelian Phenotypes Poster Session II

PB4822 Leveraging an animal model of CHD8-related syndrome to interrogate brain circuits: CHD8 heterozygous mice display decreased impulsivity in the differential reinforcement of low rate responding assay, but no difference in attention or motivation, compared to wildtype mice.

Authors:

S. Dhanani[1], T. Sohrabi[1], M. Begum[1], D. F. Rivas[1], P. Balsam[1], R. A. Muhle[2]; [1]Columbia Univ./New York State Psychiatric Inst., New York, NY, [2]Columbia Univ./ New York State Psychiatric Inst., New York, NY

Loss of function mutations in *CHD8* are often associated with autism, intellectual disability, and bodily overgrowth. Mouse models of CHD8 loss have commonalities, including avoidance of open areas, increased sniff time, and larger brain volume. The use of operant conditioning to determine differences in attention, motivation, and impulsivity between *Chd8*$^{+/-}$ mice and their wildtype (WT) littermates has not been fully assessed.
We tested an all-male cohort of 8 *Chd8*$^{+/-}$ and 8 WT food-restricted mice, starting at 8 weeks old. Med Associates operant boxes were used to assess for sustained attention (SA), motivation using progressive ratio (PR), and impulsivity via the differential reinforcement of low rate responding (DRL) assays, sequentially. *SA task*: a visual cue predicting which of 2 levers could earn a reward was reduced from 10 to 0.25sec. *PR task:* the number of lever presses to receive a reward doubled after each reward received. *DRL assay:* the mice received a reward only if they abstained from pressing the lever within a specified time interval, with a training period increasing the interval by 2sec every 2-3 days, until reaching a 36sec interval that was tested for 10 days. 2-way ANOVA or an unpaired parametric t-test was used to test statistical significance. In the DRL assay, nonlinear regression curves (a one phase decay curve to latencies < 4sec, and a Gaussian curve to latencies > 4sec) were fit to histograms of the summed lever press latencies, binned in 1sec intervals.
We did not find differences in the ability of *Chd8*$^{+/-}$ mice to sustain attention to a visual cue in the SA task, or in their motivation to press increasing numbers of times in the PR task, suggesting CHD8 loss does not impair attention or motivation. In the DRL assay, *Chd8*$^{+/-}$ mice adapted more quickly to the task, which rewarded ability to wait between presses. A Gaussian curve fit to the lever press latencies of the *Chd8*$^{+/-}$ mice in the 36sec interval tasks was significantly shifted to the right of the WT mice ($p < 0.0001$). Greater numbers of burst presses (latencies < 1sec) by the WT mice was not significant during the 36sec interval testing sessions ($p = 0.3963$), but when considering the entire training period was significant for the interaction of session x genotype ($p = 0.0023$ for session X genotype), as well as for subject and session ($p < 0.0001$ for subject and for session, $p = 0.1159$ for genotype).
These aspects of DRL performance suggest that *Chd8*$^{+/-}$ are more accurate in the timing of behavior and better able to inhibit repetitive initiation of behavior. These behaviors implicate basal ganglia circuits involving reward, and can inform our understanding of behavioral differences in people with CHD8-related syndrome.

Session Title: Mendelian Phenotypes Poster Session III

PB4823 Leveraging identity-by-descent to identify novel loci associated with esophageal ulceration

Authors:

J. Baker[1], G. Evans[1], R. Bohlender[2], H-H. Chen[3], A. S. Petty[4], C. Huff[2], D. Samuels[5], J. Below[6]; [1]Vanderbilt Univ., Nashville, TN, [2]U.T. MD Anderson Cancer Ctr, Houston, TX, [3]Academia Sinica, Taipei, Taiwan, [4]Vanderbilt Univ. Med. Ctr., Nashville, TN, [5]Vanderbilt Univ. Sch. of Med., Nashville, TN, [6]Vanderbilt Univ Med Ctr., Nashville, TN

Background: Esophageal ulceration is characterized by open sores that break the lining in the esophagus. These ulcers commonly form on the lower end of the esophagus near the connection to the stomach. These ulcers can cause discomfort or pain while swallowing, acid reflux, and chest pain. Esophageal ulcers are typically the result of gastroesophageal reflux disease (GERD) and when left untreated can develop into Barrett's esophagus. Very little concrete information is available about the etiology of esophageal ulceration but heritability for GERD and Barrett's esophagus have been estimated as 30% and 40%, respectively, indicating that esophageal ulcers may also have similar genetic causes. The lack of clarity of the genetic etiology of Esophageal ulcers limits our ability to determine the definitive cause of the condition, ultimately hindering the determination of an appropriate treatment. Large DNA biobanks can provide an opportunity to elucidate some of the genetic causes of esophageal ulceration by enabling researchers to leverage cryptic relatedness through shared identical-by-descent (IBD) genomic segments resulting from recent common ancestry. **Methods and Results:** Phasing was performed using SHAPEIT4 and then pairwise IBD segments ≥3 cM were identified using hap-IBD for 69,819 individuals of European ancestry in BioVU, Vanderbilt University Medical Center's biobank. We defined cases of esophageal ulceration as individuals who have a record of the PheCode 530.12, which is composed of ICD codes 530.2, 530.20, 530.21 (ICD-9), and K22.1 (ICD-10) on two or more unique dates. Controls were defined as individuals without any mention of a PheCode in the exclusion ranges (530-530.99, 532-532.99). We performed IBD mapping genome-wide on a set of 142 cases and 48,156 controls to identify enrichment of IBD sharing in case-case pairs compared to case-control pairs. We identified one unique, genome-wide significant enrichment signal on chromosome 10 that exceeded the family-wise error rate correction (6e-5). Using UCSC genome browser we identified 11 genes within this signal. None of these genes had any previous association with related esophageal phenotypes in OMIM. There were also no variants within this signal region from the GWAS Catalog that had been associated with phenotypes relevant to esophageal ulceration pathology. **Conclusions:** We were able to identify a novel genome-wide significant signal for esophageal ulceration using IBD mapping. These results highlight how our IBD-based approach can leverage cryptic relatedness within biobanks to identify novel associations with heritable diseases.

Session Title: Mendelian Phenotypes Poster Session I

PB4824 Long telomeres promote clonal hematopoiesis with aging by delaying replicative senescence

Authors:

M. Tassia[1], E. DeBoy[2], K. Schratz[2], S. Yan[1], R. McCoy[1], M. Armanios[2]; [1]Johns Hopkins Univ., Baltimore, MD, [2]Johns Hopkins Univ. Sch. of Med., Baltimore, MD

Telomere length predicts the replicative potential of cells, and short telomere length signals cellular senescence. We identified individuals with germline heterozygous loss-of-function mutations in *POT1*, a protein which binds single-stranded DNA at the 3'-end of telomeres. These mutations affected both POT1 stability and telomere DNA-binding. *POT1* variant carriers possessed long telomere lengths relative to age-matched controls, and their leukocyte telomere length was maintained over a 2-year period when compared with unaffected relatives where expected telomere shortening was observed. Variant carriers with long telomeres developed multiple neoplasias, ranging from benign to malignant, and the tumor spectrum affected multiple organs - manifesting as melanoma, thyroid cancer, and sarcoma often associated with hematologic malignancies. To investigate the mechanisms driving these observations, we performed whole genome sequencing on whole-blood-derived hematopoietic progenitors collected from two *POT1* mutation carriers and compared these genomes to those collected from a non-carrier relative. Genomic analyses revealed elevated somatic mutation burden dominated by clock-like C>T mutation signatures, supporting the hypothesis that *POT1* haploinsufficiency delays replicative senescence. *POT1* mutation carriers also possessed somatic variants canonically associated with driving clonal hematopoiesis of indeterminant potential (CHIP), where mutations in *DNMT3A* and *JAK2* were most common. Phylogenetic analyses estimated CHIP-driver mutations emerged early in life and were sustained with age, propagating to high frequency by adulthood. Using population genetic simulations, we constructed a non-Wright-Fisher evolutionary model that extended our empirical findings - presenting an age-dependent evolutionary model where *POT1* loss-of-function mutations and long telomeres promote CHIP through replicative longevity, and telomere shortening limits the propagation of somatic mutations over time. Together, our results support a role for long telomere length in promoting clonal evolution with age; and the clinical phenotype is associated simultaneously with both cancer and CHIP predisposition. Our findings underscore the importance of telomere shortening as tumor suppressive mechanism with aging.

Session Title: Mendelian Phenotypes Poster Session II

PB4825 Long Term Renal Transplant Success is Possible in Hypoparathyroidism, Sensorineural Deafness, and Renal Dysplasia Syndrome: A Familial Case Series.

Authors:

L. Alvey[1], V. Viswanath[1], J. Owens[2], A. Shillington[2]; [1]Univ. of Cincinnati Coll. of Med., Cincinnati, OH, [2]Cincinnati Children s Hosp., Cincinnati, OH

Hypoparathyroidism, Sensorineural Deafness, and Renal Dysplasia (HDR) Syndrome, formerly known as Barakat syndrome, is caused by autosomal dominant loss of function *GATA3* variants. Renal disease is common and presents in 72.2% of patients with HDR syndrome. 40% of renal disease is due to congenital anomalies of the kidney and urinary tract, while the remaining cases demonstrate variably progressive disease. Prognosis of HDR Syndrome is primarily dependent upon the severity of renal disease, and current treatment includes routine monitoring and standard treatments for chronic kidney disease. To date, four patients with end stage renal disease from molecularly confirmed HDR syndrome have had renal transplants; however, there is minimal long-term data regarding the efficacy of renal transplants in these patients.Here we present a family of 4 genetically confirmed HDR syndrome patients, which is most notable for a patient having undergone a successful living donor renal transplant 27 years prior. In this family, a 41-year-old female with a history of hearing loss, UTIs, and hematuria was diagnosed with HDR syndrome via genetic testing which found a pathogenic c.608_609del (p.Gly203Glufs*100) variant in *GATA3*. This variant was also identified in her 5-year-old female daughter, 4-year-old male son, and 74-year-old father. Her father was diagnosed with ESRD at age 46 and had a living donor kidney transplant at age 47. His transplant was well tolerated until age 73. Workup at that time revealed acute renal failure with acute tubular necrosis, advanced global glomerulosclerosis, moderate arteriosclerosis, and severe interstitial fibrosis / chronic tubular interstitial changes. Renal failure likely occurred due to the expected lifespan of a grafted kidney rather than as a result of HDR syndrome.The reported family presented to clinic with a clinical diagnosis of Alport syndrome due to their familial renal disease and congenital hearing loss. These congenital symptoms are frequently seen in HDR syndrome; however, the kidney disease / hearing loss in Alport syndrome is explicitly progressive over years to decades and not found at birth. In addition to highlighting discrepancies between Alport syndrome and HDR syndrome, this family also establishes that long term successful renal outcomes are possible in HDR syndrome. Our findings suggest that kidney transplantation should be readily offered to individuals with end stage renal disease due to HDR syndrome.

Session Title: Mendelian Phenotypes Poster Session III

PB4826 Long-read sequencing reveals novel transcripts induced by misexpression of DUX4 in FSHD muscle

Authors:

D. Zheng[1], A. van den Heuvel[2], A. Mahfouz[1,3], S. Bennett[4], S. Tapscott[4], S. Van Der Maarel[1]; [1]Leiden Univ. Med. Ctr., Leiden, Netherlands, [2]Univ. Utrecht, Utrecht, Netherlands, [3]Delft Univ. of Technology, Delft, Netherlands, [4]Fred Hutchinson Cancer Ctr., Seattle, WA

DUX4 is a germline transcription factor that is expressed during early embryogenesis and subsequently silenced in most somatic tissues. Misexpression of DUX4 in skeletal muscle is considered the major cause of Facioscapulohumeral muscular dystrophy (FSHD). DUX4 activates a cascade of downstream events leading to muscle wasting. Besides protein-coding genes, DUX4 also regulates expression of several classes of repetitive elements, and is known to affect RNA processing. Because of these complexities using data from short-read sequencers limits our understanding of the complex transcriptional events provoked by DUX4. We combined SMRT long-read isoform sequencing with short-read RNA-seq of DUX4-inducible myoblasts to investigate the full-length transcriptome landscape inflicted by DUX4. In the presence of DUX4, we observed a more intricate transcriptional landscape, which was distinguished by the emergence of novel isoforms for the known genes, as well as a considerable number of extensive alternative splicing events. In addition, DUX4-dependent transcriptional activation of 1275 novel (i.e. unannotated) intergenic isoforms (652 intergenic loci) was identified and verified by corresponding short-read RNA-seq data, bulk RNA-seq data of primary myotubes, and embryonic scRNA-seq data. Our analysis of public DUX4 ChIP-seq data and ATAC-seq data obtained from DUX4-expressing myoblasts and human embryonic stem cells, respectively, revealed 399 intergenic isoforms (361 intergenic loci) to be directly regulated by DUX4 expression. Intergenic loci with predicted coding transcripts could be confirmed in Ribo-seq data of DUX4-expressing myoblasts, indicating that these novel intergenic transcripts are translated into novel polypeptides. Taken together, our study elaborates on the transcriptional features induced by DUX4 and reveals unannotated transcripts at transcriptome and translatome levels. These can be considered potential biomarkers for disease diagnosis, progression, and therapeutic intervention in FSHD, as well as being explored for their role in cleavage-stage embryos.

Session Title: Mendelian Phenotypes Poster Session I

PB4827 Loss of function in *RBBP5* results in a syndromic neurodevelopmental disorder associated with microcephaly.

Authors:

S. Jangam[1,2], K. Jay[1,2], Y. Huang[3], A. Huang[3], J. Wan[3], O. Chorin[4], A. Rothschild[4], O. Barel[5], M. Mariani[6], M. Iascone[7], H. Xue[8], Undiagnosed Disease Network, J. Hunag[8], C. Mignot[9], K. Boris[9], V. Saillour[10], S. Yamamoto[1,2], O. Kanca[1,2], H. Bellen[1,11], C. Palmer[3,12], S. Nelson[3], M. Wangler[1,2], J. Martinez-Agosto, A[3,12]; [1]Baylor Coll. of Med., Houston, TX, [2]Jan and Dan Neurological Res. Inst., Texas Children's Hosp., Houston, TX, [3]UCLA, Los Angeles, CA, [4]Sheba Med. Ctr., Ramat Gan, Israel, [5]Sheba Med Ctr, Ramat-gan, Israel, [6]Santa Anna Gen. Hosp., Italy, Italy, [7]Laboratorio di Genetica Medica, Bergamo, Italy, [8]Shanghai Jiao Tong Univ. Sch. of Med., Shanghai, China, [9]AP-HP Sorbonne Univ., France, France, [10]Laboratoire de biologie médicale multisites Seqoia, Paris, France, [11]Jan and Dan Duncan Neurological Res. Inst., Houston, TX, [12]Dept. of Psychiatry and Biobehavioral Sci., Los Angeles, CA

We are investigating the *RB binding protein 5 (RBBP5)* gene as a novel neurodevelopmental disorder candidate. *RBBP5* is a member of COMPASS (COMplex of Proteins Associated with SET1) epigenetic modifier protein complex that methylates Histone 3 lysine-4 (H3K4). Chromatin dysregulation has been associated with various syndromes, but *RBBP5* has not yet been implicated in human disease. We have identified four unrelated individuals through the Undiagnosed Diseases Network (UDN) with *de novo* heterozygous pathogenic variants in *RBBP5*. Two missense and two nonsense variants were identified in probands with neurodevelopmental symptoms, including global developmental delay, intellectual disability, microcephaly, and short stature. We examined the variants using bioinformatic analysis, protein structure modeling, and *in vivo* functional characterization. The expression of the frameshift variants in cell lines does not result in the expression of the RBBP5 protein, confirming that these are loss-of-function alleles. Both missense variants p.T232I and p.E296D affect evolutionarily conserved amino acids and are expected to interfere with the interface between RBBP5 and the histones. To analyze the variants *in vivo*, we used the *Drosophila melanogaster* as a model organism. Loss of the fly *Rbbp5* leads to developmental phenotypes, including microcephaly and loss of H3K4 trimethylation in the larval stage. In human cDNA rescue experiments, $RBBP5^{Ref}$, and $RBBP5^{T232I}$ fail to rescue the loss of *Rbbp5*. However, expression of $RBBP5^{Ref}$ induces a more severe reduction in brain size than *Rbbp5* null larvae. Expression of $RBBP5^{T232I}$ results in a similar microcephaly phenotype as null animals, suggesting $RBBP5^{T232I}$ is a loss-of-function allele. In overexpression paradigms, ubiquitous expression of $RBBP5^{Ref}$ leads to late larval lethality and, when expressed in the eye (tissue specifically), causes a small eye phenotype. Expression of either $RBBP5^{Ref}$ or $RBBP5^{E296D}$ results in a more severe small eye phenotype than $RBBP5^{T232I}$, again indicating $RBBP5^{T232I}$ is a loss-of-function allele. Here, by demonstrating $RBBP5^{T232I}$ is a hypomorphic allele, we are supporting that de novo heterozygous variants in RBBP5 are associated with a novel syndromic neurodevelopmental disorder. We conclude that RBBP5 nonsense and missense variants result in similar neurodevelopmental phenotypes. Our future work will focus on determining the mechanism of the missense $RBBP5^{E296}$ allele in the rescue paradigm and identifying the downstream dysregulated genes resulting from RBBP5 loss-of-function.

Session Title: Mendelian Phenotypes Poster Session II

PB4828 Loss of GFAP causes optico-retinal abnormalities and vision impairment.

Authors:

K. Grønskov[1], M. V. K. Sarusie[1], C. Rönnbäck[2], C. Jespersgaard[1], S. Baungaard[3], Y. Ali[4], L. Kessel[2], S. T. Christensen[3], K. Brøndum-Nielsen[1], K. Møllgård[3], T. Rosenberg[2], L. A. Larsen[5]; [1]RigsHosp.et - Kennedy Ctr.t, Glostrup, Denmark, [2]RigsHosp.et - Glostrup, Glostrup, Denmark, [3]Univ. of Copenhagen, Copenhagen, Denmark, [4]Univ. of, Copenhagen, Denmark, [5]Univ Copenhagen, Copenhagen, Denmark

Glial Fibrillary Acidic Protein (GFAP), is major constituent of the intermediate filament of astrocytes within the central nervous system and the retina. During mammalian brain development GFAP is expressed in radial glial cells (RGCs), but previous studies of mouse or human tissues have not identified GFAP expression in retinal neural progenitor (RNP) cells. Monoallelic missense variants in *GFAP* cause Alexander disease (AxD) which is characterized by Rosenthal fibers (protein aggregations including GFAP). Currently, no phenotypic consequence of GFAP haploinsufficiency is known, and recent rodent experiments suggest that antisense oligonucleotide-mediated suppression of GFAP expression can reverse the disease progression in AxD.We investigated a six-generation family with ten individuals presenting with visual impairment, retinal dysplasia and pseudopapilledema. Whole genome sequencing of three affected individuals revealed a rare novel loss-of-function variant in *GFAP*, which segregated with disease in the family. The variant, c.928dup, results in frameshift and translation into a 421 aa protein p.(Met310Asnfs*113) (wild type GFAP is 431-438 aa). We analysed a series of human embryonic tissues and identified strong GFAP expression in RNPs within the developing eye at 35-51 days post conception. Experiments using zebrafish models verified that the c.928dup variant does not result in extensive GFAP protein aggregation. Analysis of zebrafish loss-of-function *gfap* mutants, showed that depletion of Gfap causes vision impairment and retinal dysplasia, characterized by a significant loss of Müller glia cells and photoreceptor cells. Our findings provide novel insight into the function of GFAP in retinal development and the consequence of GFAP deficiency.

Session Title: Mendelian Phenotypes Poster Session III

PB4829 Loss of TBCK alters bone cell homeostasis resulting in altered bone morphology.

Authors:

E. Durham[1], A. Black[1], K. Katsura[1], E. M. Gonzalez[1], D. Layo-Carris[1], E. Lubin[2], A. Diaz[1], E. Bhoj[1]; [1]Children's Hosp. of Philadelphia, Philadelphia, PA, [2]Univ. of Pennsylvania/PSOM, Philadelphia, PA

In 2016, we and others discovered mutations in TBC1 Domain Containing Kinase (TBCK) that cause a complete loss of TBCK protein, leading to a multisystem disorder characterized by pediatric neurodegeneration. Recently, through clinical analysis of dozens of children with this rare disorder, we have also determined that this syndrome is associated with an abnormal craniofacial form, as well as a bone fragility phenotype resembling severe osteoporosis that is far more pronounced than would be predicted by just hypotonia/deconditioning. To study this devastating and fatal disease, we have established a biorepository of patient fibroblast samples. We found that, when given osteogenic media for 21 days, patient-derived fibroblasts both failed to differentiate and presented with altered signaling in genes related to osteoclast function, indicating that these cells have altered differentiation potential compared to control fibroblasts. Our lab also acquired a systemic knockout mouse model that recapitulates the neurologic phenotype seen in patients. Though tbck-/- mice experience perinatal lethality, we have begun to investigate the impact of TBCK loss in heterozygous adult mice (tbck+/-) relative to their TBCK (tbck+/+) littermates, as several compound heterozygous mutations have been identified to be pathogenic in patients with TBCK-encephloneuropathy syndrome. We hypothesize that, though TBCK Syndrome is associated with compete loss of TBCK protein, direct and/or indirect effects of this mutation may impact bone development and health even in heterozygotes. In an adult cohort of heterozygous (tbck+/-) and unaffected (tbck+/+) mice (n=7/genotype), we found no difference in skull endocast volume between genotypes, using uCT and 3D Slicer software. However, cephalometric analysis indicated differences in tbck+/- skulls. Heterozygous calvaria were altered with a wider and longer shape as compared to unaffected littermates. Additionally, histomorphometric investigations showed altered suture morphology. Our results confirm system-wide effects of TBCK loss, specifically in bone. This current work is a first step in establishing a novel pathway between these pathogenic mutations and systemic bone health. With adverse bone phenotypes noted in over half of the documented TBCK Syndrome patients, our continued objective is to identify therapeutically targetable mechanisms of TBCK-related craniofacial and bone abnormalities.

Session Title: Mendelian Phenotypes Poster Session I

PB4830 Maternal and *De Novo* Variants in *MYBPC3* Result in Presentation of Severe Neonatal Mixed Hypertrophic and Non-compaction Cardiomyopathy.

Authors:

A. Repp, T. Lee; Columbia Univ. Irving Med. Ctr., New York, NY

Introduction Variants in the cardiac myosin binding protein-C (*MYBPC3*) gene are the most common cause of familial hypertrophic cardiomyopathy (HCM), a leading cause of sudden cardiac death. A heterozygous autosomal dominant gene mutation of the *MYBPC3* gene typically presents milder phenotype, incomplete penetrance, and later age of onset. Our case poses a unique presentation of biallelic *MYBPC3* mutations via one *de novo* and one inherited variant.

Case Presentation A five-week-old female presented with new mixed phenotype HCM/NCCM with moderately diminished left ventricular (LV) systolic function. Non-invasive ventilatory support was used throughout her hospital course. She remained on milrinone for inotropic support, and high-dose intravenous diuretics to maintain a stable clinical status. Her echocardiogram prior to OHT demonstrated an LV ejection fraction of 32 % by 5/6 area x length [normal range 54-74%]; LV end diastolic volume 32.8 ml [Z-score 3.24]; LV mass 39.3 g [Z-score 3.95].

Genetic Testing and Results Given severity of symptoms, rapid genome sequencing (rGS) was sent versus the standard cardiomyopathy panel, and the patient was concurrently listed for cardiac transplant. rGS resulted in two variants in the *MYBPC3* gene: a *de novo* heterozygous pathogenic variant (c.1227-13 G>A) and maternal heterozygous likely pathogenic variant (c.1224-80 G>A). The maternal aunt is also a carrier of the likely pathogenic variant c.1224-80 G>A. The mother and maternal aunt both have had cardiac screening with echocardiogram and have no clinical findings.

Discussion The cardiogenetic team determined, with consideration of the patient's early-onset disease, that the *de novo* variant (c.1227-13 G>A) is the main cause of the severe, mixed HCM/NCCM phenotype, and that the maternal variant likely played little role in the patient's clinical outcome. This has informed the genetic counseling that has been provided to the family. Cardiac screening every three to five years for gene-positive phenotype-negative family members has been recommended. If the family were to do preimplantation genetic testing, they were counseled to test for the *de novo* variant.

Conclusion *MYBPC3*-cardiomyopathy is classically an autosomal dominant disorder. However, this case emphasizes the need for clinicians to consider varied inheritance patterns and additive pathogenic and likely pathogenic variants. Particularly in pediatrics, it is vital to consider how these genes play a role, especially in infants with severe disease. Clinicians should therefore maintain a lower threshold for rapid whole exome or genome sequencing in pediatric populations with severe disease presentation.

Session Title: Mendelian Phenotypes Poster Session II

PB4831 † Metabolic Liver Transplant: A Single Cooperative from 2005-2022

Authors:

C. Heuston[1], N. Yazigi[1], U. Ekong[1], K. Chapman[2]; [1]Georgetown, Washington, DC, [2]Children's Natl., Washington, DC

Liver transplant (LT) can increase quality and length of life for patients with inherited metabolic diseases (IMDs) who suffer severe and unpredictable complications despite strict dietary and medical management. With advancements in surgical techniques and immunosuppression, LT has been increasingly used as a therapeutic option for IMDs to compensate for enzyme insufficiencies. To gain a better understanding of the metabolic outcomes and potential benefits of LT, we retrospectively analyzed biochemical markers such as ammonium and serum amino acid levels, dietary treatment, and growth in 61 individuals with specific types of IMDs from our cohort of 99 who received LTs at the Medstar Georgetown University Hospital between 2005 and 2022. Of the 61 patients with intoxication-type IMDs, there was a relaxation in diet, discontinuation of nitrogen scavengers (if used prior to transplant), and overall increase in height and weight post-liver transplant. In patients with maple syrup urine disease (MSUD), the average isoleucine levels decreased from 270.6±142.9 and to 148.6±76.4 post-LT (p-value 0.006). In patients with urea cycle disorders (UCD), average serum ammonium decreased from 64.6±37.3 to 35.8±17.2 post-LT (p-value 0.01). Patients with methylmalonic aciduria (MMA) had statistically significant decreases in both isoleucine (p-value 0.04) and ammonium levels (p-value 0.03) post-LT. No statically significant changes in amino acids levels were found in patients with propionic acidemia (PA). For individuals with moderate to severe presentations of intoxication-type IMDs, LT is a viable medical option that achieves good metabolic outcomes and improves quality of life.

Session Title: Mendelian Phenotypes Poster Session III

PB4832 Microcephaly with or without Chorioretinopathy, Lymphedema and Mental retardation (MCLMR) syndrome in a patient from India

Authors:

P. Kaur, A. Kaur, A. Kumari, R. Daniel, A. Arora, D. Katoch, I. Panigrahi; Postgraduate Inst. of Med. Ed. and Res., Chandigarh, India

Microcephaly with or without chorioretinopathy, lymphoedema, or mental retardation syndrome (OMIIM #152950) is a rare autosomal dominant syndrome caused due to pathogenic variations in *KIF11* gene on chromosome 10q23. Microcephaly, eye abnormalities, intellectual disability and lymphodema have been observed in 91%, 72%, 67%and 47% respectively of *KIF11* mutation. We hereby report first case from India with deletion in KIF11 gene. A 15 months old male child presented with pedal odema which was noticed since birth. He had smooth perinatal transition and the developmental milestones were achieved at normal age. On examination his weight for age was 8.9 kg (-0.68 Z), height for age was 75 cms (-1.04 Z) and head circumference was - 39.5 cms (-4.53Z). He appeared to have dysmorphic features with microcephaly, sloping forehead, prominent metopic suture, upward slanting palpebral fissures, long philtrum and thin upper lip. Additional abnormalities detected were bilateral simian crease and bilateral pedal edema. His fundus examination revealed marked tessellation with well defined areas of white appearing chorioretinal atrophic patches. Focussed exome sequencing revealed contiguous heterozygous deletion of 643 bp, spanning genomic location chr10:g.(92613040_92613130)_(92613377_92613619) encompassing exonic regions 7 to 8 of the *KIF11* gene. To further confirm the results, c-DNA study was done. The results showed amplified PCR product of 729 bp in control, mother and father of index case while PCR product of 729 and 395 bp in index case depicting and confirming heterozygous deletion of 334 bp. Majority of the reported mutations in *KIF11* are point mutations . In 2018 first patient with MCMLR syndrome with a microdeletion (209Kbp) on chromosome 10q23.3 encompassing *KIF11* gene was reported. Chorioretinopathy is the specific eye abnormality, other findings reported are retinal folds, microphthalmia, and myopic and hypermetropic astigmatism. The lymphedema is confined to the dorsum of the feet and resolves spontaneously. The overall penetrance of the *KIF11* mutation can be considered to be high (95%) and variable expressivity is reported. In our case parents were clinically unaffected, fundus examination was normal and c-DNA PCR did not reveal deletion in KIF11 gene in parents, pointing towards the denovo origin of the deletion in proband. This report will also be of interest to the opthalmologists as the fundus appearance in this syndrome mimics Familial Exudative Viteroretinopathy, an inherited retinal vascular disease.

Session Title: Mendelian Phenotypes Poster Session I

PB4833 Molecular Characterization of short stature & its syndromes in consanguineous families from Pakistan.

Authors:

T. Ishaq[1,2], F. Alzehrani[3], G. Seo[4], H. A. Ishaq[5], F. S. Alkuraya[3], O. Makitie[6,7], S. Naz[2]; [1]Yale Sch. of Med., New Haven, CT, [2]Sch. of Biological Sci. Univ. of the Punjab., Lahore, Pakistan, [3]King Faisal Specialist Hosp. and Res. Ctr., Riyadh, Saudi Arabia, [4]3billion inc, Seoul, Korea, Republic of, [5]Dept. of Radiology, Government Teaching Hosp. Shahdara, Lahore, Pakistan, [6]Dept. of Molecular Med. and Surgery and Ctr. for Molecular Med., Karolinska Inst.,, Stockholm, Sweden, [7]Children's Hosp., Univ. of Helsinki and Helsinki Univ. Hosp., Helsinki, Finland, Helsinki, Finland

Short stature may be an isolated finding caused due to growth hormone deficiency or can be accompanied by various skeletal defects. Skeletal dysplasias are a heterogeneous group of heritable disorders characterized by abnormalities of cartilage and bones which affect the shape and size of the skeleton along with disproportion of long bones and spine abnormalities. We analyzed and characterized recessively inherited short stature disorders with or without skeletal dysplasia in twelve families from the Punjab region of Pakistan. Radiographs were obtained for the probands. DNA samples from each family were first screened for a common *GHRHR* pathogenic variant and the non-coding *RMRP* gene by Sanger sequencing. Samples from families negative for variants in these genes were selected for exome sequencing to elucidate the genetic causes. The data were analyzed using Franklin software or 3billion automated pipeline. Homozygous variants having an allele frequency less than 0.01in various public databases were selected. Several *in silico* tools were used to check the pathogenicity of each variant. Sanger sequencing was completed to ascertain co-segregation of variant with the phenotype in each family. Patients in four families had short stature only while the remaining exhibited skeletal dysplasia. Initial screening revealed that affected members of one family were homozygous for a known pathogenic variant of *RMRP* while patients in three families were homozygous for a *GHRHR* variant. Exome sequencing and subsequent analysis identified novel homozygous splice-site variants in *NPR2* and *TG* in two families. Patients of the other families had previously reported variants in *EVC2*, *GH1*, *PAPSS2* and *PCNT* with the same variant in the latter gene accounting for the disorder in two families. All variants were rare in public databases as well as absent in the DNA of 200 ethnically matched controls. Our study revealed genetic heterogeneity of short stature and its syndromes and identified few founder mutations in the Pakistani population.

Session Title: Mendelian Phenotypes Poster Session II

PB4834 Molecular detection of Sickle Cell Anemia & Classic Haplotypes in the β-globin gene *(HBB)* cluster by means of SNPs, in a group of samples from Bolívar department, for a better understanding of this rare disease in Colombia.

Authors:

M. DIAZ-MATALLANA[1], E. Benavides-Benítez[2], I. Briceño[1], J. Martínez-Lozano[1], J. Bernal[2]; [1]Univ. de La Sabana, Chía, Colombia, [2]Univ. del Sinú Elías Bechara Zainúm - Seccional Cartagena, Cartagena, Bolívar, Colombia

INTRODUCTION. The best known hemoglobinopathy is sickle cell disease (SCD), also called sickle cell anemia. According to the National Organization for Rare Disorders in the United States (NORD), the frequency of sickle cell anemia varies from country to country. Mutations in the *HBB* gene (β globin cluster) are common in people from Africa, the Mediterranean, the Middle East, India, & in people from the Caribbean and regions of Central and South America but can be found in individuals of any ethnic origin.GENERAL OBJECTIVE. Characterize hematological, biochemical, & molecularly SCD, likewise, detect classic haplotypes by DNA sequencing in a group of blood samples from the department of Bolívar, Colombia. METHODOLOGY. This work corresponds to an analytical cross-sectional study, with convenience sampling. Colombian blood samples were collected after informed consent from volunteers from several municipalities in the Bolívar department (n=287), but also samples from the Pacific region (n=13), Providencia Island (n=5), & Bogotá [abnormal hemoglobins: blood on filter paper (n=4)] were included. Data were obtained: (1) hematological (hemograms, reticulocytes); (2) Biochemical tests: dHPLC technique was used to determine Hemoglobin (Hb); (3) DNA sequencing data & classical haplotype detection through five SNP markers (designed oligonucleotides based on SNPs: *rs334, rs3834466, rs28440105, rs10128556, rs968857*).RESULTS. 103 samples have been identified by *rs334* through Sanger´s Sequencing technique. When contrasting the results of identification between *rs334* SNP Vs. Hb Chromatograms, a coincidence was found in 39 samples out of 43 samples analyzed, therefore, when comparing these techniques, a concordance of 90.7% was found. 25 samples out of 38 samples previously analyzed by *rs334* were classified into classical haplotypes CAR (12), BEN (8), CAM (2), SEN (1), & ATP-I (2). In this study, the greatest frequency of the CAR haplotype (48%), is similar to that observed in a previous study where CAR haplotype was the highest (58%). Other studies suggest that the largest proportion of individuals arriving in the Americas from Africa originated from BEN (70%) & small proportions from the CAR (17%) & SEN (13%). These SNPs-based classification have not been performed before in Colombia. The integration of different components used in this research will not only provide new knowledge into the approach & follow-up of SCD, but also a better understanding of the clinical manifestations of the disease. The results of this work will make it possible to expand the data or records of carriers & those affected, which will benefit patients & their families.

Session Title: Mendelian Phenotypes Poster Session III

PB4835 Molecular diagnosis of retinal dystrophies in a northeast population of Mexico

Authors:

R. Villafuerte-de la Cruz, J. Valdez-Garcia, A. Rojas-Martinez; Tecnologico de Monterrey, Monterrey, Mexico

Inherited retinal dystrophies (IRDs) are characterized by progressive degeneration of photoreceptors, resulting in vision loss, that may develop from birth through to late middle age. IRDs exhibit both genotypic and phenotypic heterogeneity. The purpose of this work was to know which alleles are most frequently related to IRDs in this population. All the patients had clinical diagnosis of IRDs. Using next generation sequencing (NGS), a genetic targeted panel which includes 330 genes was done to a cohort of 126 non-related patients. In total, 94 patients were solved (74.6%), 10 patients were partially solved (7.9%) and 22 patients were unsolved (17.5%). The most common mode of inheritance in this cohort was autosomal recessive. The three most common mutated genes in this cohort were *USH2A* (28.7%), *ABCA4* (15%), *RPGR* and *PROM1* (3.2%). Phenotypically, in isolated cases retinitis pigmentosa was found in 42.1%, followed by macular dystrophies 16.7% and cone rode dystrophy in 11.9%. For syndromic cases, Usher type 2 was the most common, it was found in 19.1%. It is important to emphasize that within *USH2A* patients, 25 had a pathogenic variant involving exon 13, but just 9 were homozygous. This work shows different allelic distribution compared to other works in Mexicans. Even if we didn't find any *RPE65* patient, it is possible that some other genetic therapies arrived soon for some of these patients.

Session Title: Mendelian Phenotypes Poster Session I

PB4836 Molecular outlook in RASopathies - a five years' experience

Authors:

I. Focsa[1,2], A. Pavel[2], D. Prepelita[3,2], L. Bohiltea[3,4], V. Plaiasu[4], A. Trifa[5], A. Tutulan-Cunita[2], D. Stambouli[2]; [1]Univ. of Med. and Pharmacy Carol Davila, Bucharest, Romania, [2]Cytogenomic Med. Lab., Bucharest, Romania, [3]Univ. of Med. and Pharmacy “Carol Davila”, Bucharest, Romania, [4]Natl. Inst. "Alessandrescu-Rusescu", Bucharest, Romania, [5]"Victor Babes" Univ. of Med. and Pharmacy, Timisoara, Romania

RASopathies are a group of genetic disorders caused by pathogenic variant in genes that encode components of RAS/mitogen activated protein kinase (MAPK) pathway. Inappropriate activation of the pathway leads to a series of overlapping clinical manifestation including: short stature, various degree of cognitive difficulties, cardiovascular malformations, skeletal anomalies and a predisposition to develop tumors. These group of diseases encompasses neurofibromatosis type 1 (NF1; OMIM 162200), Noonan syndrome (NS; OMIM 163950), Noonan syndrome with multiple lentigines (NSML; OMIM 151100), Noonan syndrome-like disorder with loose anagen hair (NSLH OMIM 607721), cardiofaciocutaneous syndrome (CFCS, OMIM 115150), Costello syndrome (CS, OMIM 218040), and Capillary malformation - arteriovenous malformation (CM-AVM; OMIM 608354).Here we present a cohort of 75 individuals, with a clinical suspicion of RASopathy, who were referred to our laboratory for molecular diagnosis. An NGS gene panel including RASopathies, schwannomatosis and differential diagnosis genes was run. Forty-three cases (57%) had a positive result, among which a majority, twenty-nine cases (31%), was molecularly confirmed with NF1. The second most frequent RASopathy identified in our cohort was NS (eight cases; 18,6%), five (11,6%) individuals carrying pathogenic variant in *PTPN11* while the other three (7%) cases harboring single nucleotide changes in *SOS1, KRAS and LZTR1,* respectively. The rest of the molecularly diagnosed cases was represented by one case (2,3%) each of NSLH (genomic changing in *SHOCH2*), CFCS (*BRAF*-impacted), CS (*HRAS*-affected), CM-AVM (*RASA1*-modified). Surprisingly, on two individuals clinically diagnosed with NF1 and NS, pathogenic variants in *NF2* and *CHD7* genes were detected, assigning the conclusive diagnosis as NF2 and CHARGE syndrome. Fifteen (47%) among molecularly undiagnosed cases follow the deletion/duplication analysis. Two exon deletions (exon 2 and exon 23) in *NF1* were identified. As deep intronic variants, large genomic rearrangements or variants in additional genes could be the molecular substratum of the remaining negative cases, comprehensive genomic testing may be considered.The advance of NGS technologies helps to identify more causal genes, to expand the phenotypic spectrum and to classify the disorders, being a useful tool for establishing the final diagnosis, which is crucial for an appropriate follow up and an accurate genetic counselling.

Session Title: Mendelian Phenotypes Poster Session II

PB4837 Monoallelic GATA2 mutation in a patient with multiple recalcitrant warts and critical COVID19

Authors:

S. Biglari[1], L. Youssefian[2], N. Mozafari[3], S. Zaresharifi[3], A. Saeidian[4], H. Hakonarson[5], V. Béziat[6], J-L. Casanova[7], H. Vahidnezhad[8]; [1]Dept. of Genetics and Molecular Biology, Sch. of Med., Isfahan Univ. of Med. Sci., Isfahan, Iran, Islamic Republic of, [2]UCLA, Los Angeles, CA, [3]Skin Res. Ctr., Shahid Beheshti Univ. of Med. Sci., Tehran, Iran, Islamic Republic of, [4]Children s Hosp. of Philadelphia, Philadelphia, PA, [5]Children's Hosp. of Philadelphia, Philadelphia, PA, [6]St Giles Lab. of Human Genetics of Infectious Diseases, Rockefeller Branch, Rockefeller Univ., New York, NY, [7]The Rockefeller Univ., New York, NY, [8]Univ. of Pennsylvania, Philadelphia, PA

Introduction:

The life-threatening coronavirus disease 2019 (COVID-19) affects less than 1 in 1,000 healthy people under 50 without underlying conditions. In some patients, life-threatening COVID-19 may be caused by monogenic inborn errors of immunity (IEI) to SARS-CoV-2 with incomplete or complete penetrance. Three patients with *GATA2* deficiency were hospitalized for COVID-19 pneumonia, albeit of moderate severity.

METHODS:

In this study, the patient's peripheral blood DNA was sequenced by whole-exome sequencing (WES). Additionally, with a novel bioinformatic pipeline, we utilized whole-transcriptome sequencing (RNA-Seq) to detect pathogenic viral (including 926 different viruses) concomitantly, the causal human pathogenic variant, and the consequences of detected sequence variants in the host. The proband, a 26-year-old female presented with generalized recalcitrant warts and bilateral sensorineural hearing loss, was born to healthy, non-consanguineous parents. She had no severe infections since birth, but three years later, early in Iran's COVID-19 pandemic, she appeared with fever, cough, and shortness of breath for one week at the emergency department. After four days in the ICU, she died of severe respiratory distress syndrome.

RESULTS

Analysis of both WES and RNA-Seq data, including multistep filtering of the variants, led to the identification of a heterozygous variant in *GATA2*: NM_001145661, c.1075-1102del28, p.W360Sfs*18, which was confirmed by Sanger sequencing and was an out-of-frame 28-bp deletion. Additionally, RNA-Seq unaligned reads to the human genome were applied to a catalog of 926 different viruses, and α-HPV-2 was detected in a biopsied wart, while it was absent in the normal skin of the patient.

DISCUSSION

A monogenic susceptibility to COVID-19 is considered, and some IEI patients, with autosomal TLR3 and X-chromosome-linked TLR7 deficiencies, have been reported to develop critical pneumonia. Three hospitalized *GATA2* patients with moderate COVID-19 have been reported so far. Here, we report the first *GATA2* patient with multiple warts and lethal COVID-19.

Session Title: Mendelian Phenotypes Poster Session III

PB4838 Monogenic genes in nephrolithiasis: comparative approaches in gene discovery.

Authors:

Y-R. Huang[1], J. Chang[1], E. Su[1], M. Baum[2], D. Bodner[3], F. Hildebrandt[4], F. R. Schumacher[5], C-H. Wu[6]; [1]Case Western Reserve Univ., Cleveland, OH, [2]Boston Children's Hosp., Boston, MA, [3]Univ. Hosp., Cleveland, OH, [4]Boston Children's Hosp, Boston, MA, [5]Case Western Reserve Univ, Cleveland, OH, [6]Univ. Hosp./Case Western Reserve Univ, Cleveland Heights, OH

Introduction: Nephrolithiasis, also known as urolithiasis, kidney stones or urinary stones, are solid masses made of minerals and salts that develop in the urinary tract. The etiologies of nephrolithiasis are complex and the overall contributing risk factors to this disease are driven by a combination of genetic, metabolic, and environmental factors. Evidence shows significant contribution of genetic factors and family history to kidney stones.
Methods: We analyze the discovery methods of monogenic causes of nephrolithiasis. The monogenic causes of nephrolithiasis have been summarized into five main approaches: biochemistry, candidate gene, linkage, whole exome sequencing (WES), and others. For comparison, we picked another metabolic disease, diabetes mellitus (DM), to analyze the differences among their gene discovery and perform proportion test for analysis.
Results: There are currently 57 genes identified as monogenic causes of nephrolithiasis. Before 1985, biochemistry was the only method to discover the monogenic causes of nephrolithiasis. This approach was unique for nephrolithiasis, but not seen in DM. After 1985, the most widely used strategies for gene discovery have been the candidate gene and linkage analysis approaches. Twenty out of the 57 monogenic causes of nephrolithiasis were identified by candidate gene; and 25 out of the 57 were identified by linkage analysis. There were no significant differences observed in the analysis of monogenic causes of nephrolithiasis and diabetes with regard to candidate gene and linkage analysis methods (p value= 0.92 and 0.75 respectively). The introduction of next-generation sequencing (NGS) strategies, accelerated the pace of discovery of diseases-causing genes. However, in nephrolithiasis, only 2 out of the 57 monogenic causes were identified by WES. Six out of 36 monogenic causes of diabetes were identified by WES. In comparison, WES is underrepresented in the discovery of the monogenic causes of nephrolithiasis (p value=0.028).
Discussion: Since 2013, WES and whole genome sequencing (WGS) have discovered nearly three times as many genes as conventional approaches. With the introduction of WES and WGS, the discovery of novel monogenic causes of diseases has accelerated rapidly, however, only 2 monogenic causes of nephrolithiasis were discovered by WES. This may suggest that WES was underutilized in the discovery of the monogenic causes of nephrolithiasis, and may have great potential to lead to novel discovery.

Session Title: Mendelian Phenotypes Poster Session I

PB4839 Mosaicism for two distinct *PTEN* variants in PTEN hamartoma and blood

Authors:

L. Garlisi Torales[1], K. Woodis[1], A. D. Britt[2], L. Surrey[2], A. Srinivasan[2], A. Ganguly[3], D. Li[2], D. Adams[2], S. Sheppard[4]; [1]NICHD/NIH, Rockville, MD, [2]Children's Hosp. of Philadelphia, Philadelphia, PA, [3]Univ. of Pennsylvania, Philadelphia, PA, [4]NICHD/NIH, Rockville, MD

The *PTEN* gene encodes the phosphatase and tensin homolog tumor suppressor. Pathogenic germline variants in *PTEN* cause PTEN Hamartoma Tumor Syndrome (PHTS), which can present with a range of vascular malformations, hamartomata, tissue overgrowth, and cancer predisposition, including breast, endometrial, thyroid, colorectal, and renal, and melanoma. Less commonly, patients can carry mosaic *PTEN* variants. Patients with mosaicism may be under-diagnosed but can present with major and minor criteria for PHTS.PTEN hamartoma of the soft tissue (PHOST) is a vascular anomaly that is often intra-muscular in the extremities. Histopathologically, it presents with the proliferation of fibroadipose tissue in the muscle and different types of disorganized vessels, which may be slow flow or fast flow. One of the differential diagnoses for slow flow PHOST is fibroadipose vascular anomaly (FAVA), and can be difficult to distinguish on imaging or histopathology.A 12 year-old female patient presented with a vascular malformation in her right ankle with pain and limited mobility, initially diagnosed with FAVA with negative genomics. Re-biopsy and subsequent genetic testing reveal two distinct mosaic *PTEN* variants in multiple tissues. Penn GDL Somatic Overgrowth and VM v3 panel performed on genomic DNA isolated from tissue biopsy revealed a somatic, pathogenic *PTEN* variant (c.683delA 2.0-2.1% mosaicism level). *PTEN* sequencing on genomic DNA from leukocytes revealed a different variant with higher mosaicism level (c.202_209+18delins27 (splice donor), 20-30% mosaicism level). Neither variant was found on genomic DNA isolated from saliva sample using the Penn GDL Somatic Overgrowth and VM v3 panel. She was diagnosed with PHTS with low grade mosaicism.This case demonstrates important concepts in genetic testing of patients with vascular anomalies. Genetic testing can be used to define and differentiate PTEN hamartoma from FAVA. Testing of additional samples, as well as using updated sequencing technology with deeper coverage, led to identification of the genetic cause. Furthermore, in most cases, genetic testing can direct medical care. However, to our knowledge, there are no guidelines regarding cancer screening in individuals with *PTEN* mosaicism. Further work is needed to understand the risk of cancer in individuals with *PTEN* mosaicism (i.e., whether this should be based on the distribution of the variants).

Session Title: Mendelian Phenotypes Poster Session II

PB4840 Mouse knockout shows *Sumo1* is essential for embryonic development and contributes to adult organ function.

Authors:

Y. Gu, D. L. Nelson; Baylor Coll. of Med., Houston, TX

Protein Sumoylation is a posttranslational modification that conjugates the small ubiquitin-like modifier (Sumo) to target proteins through covalent bond. The SUMO family includes *SUMO1-5*; *SUMO1-3* are ubiquitously expressed while *SUMO 4* and *5* are tissue-specific. *SUMO1* mutation has been reported in individuals with cleft lip and reduced levels of SUMO1 in heart have been found in heart failure. In a search for genetic modifiers of Fragile X associated Tremor/Ataxia syndrome (FXTAS) and Fragile X associated Primary Ovarian Insufficiency (FXPOI), we identified *SUMO1*. To understand *SUMO1* function in normal development and potential roles in modification of FXTAS and FXPOI, we created a new *Sumo1* knockout first mouse model using the EUMMR ES cell HEPD0731_5_10. Surprisingly, the majority of mid-to-high percentage chimeras (4 of 5) showed neck skin lesions in adults, but germline transmission was eventually achieved. When *Sumo1* heterozygous males were crossed with *Sumo1* heterozygous females, no homozygous mice were produced, indicating embryonic lethality. *Sumo1* homozygous embryos die before embryonic day 11.5 (E11.5). Curiously, among 135 offspring from *Sumo1* heterozygous mating, 103 were heterozygous while only 32 were wild type, with no homozygous knockouts. Recovery of wild type pups was reduced (33% verse 24%) possibly indicating that gametes carrying the *Sumo1* mutation had an advantage over the wildtype gametes during fertilization. To confirm this result, we set up mating of heterozygous mice with C57BL/6 wildtype mice. We identified 68 wildtype (42%) and 93 heterozygous (58%) pups out of 161 offspring, an 8% deviation from the expected numbers, consistent with results from crossing heterozygous mice and supporting the notion of a gametic advantage. *Sumo1* heterozygous mice showed a significant reduction in weight, especially in mid-to-late adulthood. Heterozygous males have small testes and kidneys while enlarged hearts in heterozygous females indicate heart failure, with increased weight in their spleen, kidneys and lungs. This new *Sumo1* knockout model provides a novel tool to study protein Sumoylation in embryonic development, fertilization, growth and adult organ function. In ongoing work, we found brain weight reduction in *Sumo1* heterozygous females and uneven expression of *Sumo1* in cerebral cortex at advanced ages, suggesting roles in protein Sumoylation in mature brain function. Studies of the effects of *Sumo1* reduction on a mouse FXTAS model are underway.

Session Title: Mendelian Phenotypes Poster Session III

PB4841 Multi-omics uncover dysregulated cerebellar pathways associated with neurodegeneration in a mouse model of Chediak-Higashi Syndrome.

Authors:

M. Talbert, S. Greene, F. Frost, M. Morimoto, D. Maynard, C. Toro, W. Gahl, M. Malicdan; NIH, Bethesda, MD

Background: Chediak-Higashi Syndrome (CHS) is a rare autosomal recessive disorder caused by bi-allelic mutations in the lysosomal trafficking regulator (*LYST*) gene. It is characterized by partial oculocutaneous albinism, bleeding diathesis, immune deficiency, and progressive neurodegeneration. The underlying function of LYST remains unclear. Animal models, such asthe beige mouse (*bg*), with a spontaneous loss-of-function mutation in the C-terminal region of LYST, have provided valuable insights into the pathogenesis of CHS. While the *bg* mouse exhibits immunologic and bleeding phenotypes characteristic of CHS, it lacks a neurodegenerative phenotype.**Methods:** To further investigate the function of LYST, we utilized CRISPR/Cas9 genomic engineering to generate a *Lyst* knockout model ($Lyst^{\Delta/\Delta}$). Our approach involved targeting exons 4 to 53, effectively deleting a majority of the open reading frame of murine *Lyst*. We conducted molecular and phenotypic characterization, as well as comprehensive multi-omics profiling.**Results:** Molecular characterization of tissues from $Lyst^{\Delta/\Delta}$ mice confirmed the deletion of *Lyst*, as both transcript and protein levels were undetectable. Phenotypic characterization of $Lyst^{\Delta/\Delta}$ mice revealed that they are viable and, similar to *bg* mice, have lighter pigmentation of the coat and skin and exhibit prolonged bleeding times. Compared to *bg*, the $Lyst^{\Delta/\Delta}$ develop signs of ataxia by 9 months, mirroring the neurodegenerative aspects of CHS. To further investigate the function of Lyst, we performed transcriptomic and lipidomic analyses on the cerebellum and cortical brain tissues from $Lyst^{\Delta/\Delta}$ and wild type mice at 1 month and 18 months of age. Bulk RNA sequencing analysis revealed a higher number of differentially expressed genes in the cerebellum of older $Lyst^{\Delta/\Delta}$ mice, suggesting progressive cerebellar involvement. Untargeted lipidomics highlighted significant differences in lipid levels between wild type and $Lyst^{\Delta/\Delta}$ in mice in cerebral cortex and cerebellum.**Conclusion:** Our findings demonstrate that the $Lyst^{\Delta/\Delta}$ model recapitulates key features of the human CHS phenotype, including CHS-associated neurodegeneration. This model provides a reliable platform for studying the function of LYST and exploring future therapies for CHS. Through comprehensive molecular, phenotypic, and multi-omic characterization of $Lyst^{\Delta/\Delta}$, we have gained a better understanding of the underlying mechanisms driving CHS, paving the way for the development of targeted interventions for this devastating disorder.

Session Title: Mendelian Phenotypes Poster Session I

PB4842 *MUSK*-related disorder severity is predicted by protein domains and allele combinations

Authors:

B. Cocanougher, L. Francescatto, S. Liu, A. Behura, M. Anneling, D. G. Jackson, K. Deak, C. Hornik, M. K. ElMallah, C. E. Pizoli, E. C. Smith, K. Tan, M. T. McDonald; Duke Univ., Durham, NC

Purpose: Biallelic loss-of-function variants in the muscle specific kinase (*MUSK*) gene lead to clinically variable neuromuscular weakness. This variability in *MUSK*-related disorders ranges from fetal demise to adult-onset limb-girdle weakness. How combinations of specific variants contribute to disease severity remains incompletely explored. Individual domains are important for separate functions, such as agrin binding (Ig1 domain), Wnt binding (Frz domain), DOK7 binding (juxtamembrane domain), and phosphorylation of downstream signaling targets (tyrosine kinase domain). Are variants that impact one domain more severe than variants that impact another?
Methods: We report one new case of neonatal onset *MUSK*-related congenital myasthenic syndrome (CMS). In the literature, we identify 45 additional cases of *MUSK*-related disorders; 23 of these cases were *MUSK*-related CMS and 21 were *MUSK*-related fetal akinesia deformation sequence. We analyze 1) null variants for which no protein is expressed and 2) missense variants and intragenic deletions, which result in altered protein products. For missense variants, we ask if disease severity is correlated with the protein domain that is directly impacted by the mutation. Next we ask whether the combination of variant types, such as null + null, null + missense, or missense + missense, predicts severity. Lastly, we combine these analyses to ask whether the combination of protein domain and allele combination types synergistically contribute to disease severity.
Results: Biallelic null variants are the most uniformly severe allele combination, resulting in fetal loss. Homozygosity for the *MUSK* p.I575T variant, which disrupts DOK7 binding, also results in fetal loss and is the most severe missense variant. Variants that disrupt the Ig1 domain are the most severe type of missense variant in *MUSK*-related CMS. The kinase domain is the most commonly mutated domain; kinase domain mutations have the widest range of disease severity. A mathematical model to predict disease severity by integrating variant location and allele combinations is developed.
Conclusions: The severity of *MUSK*-related disorders is impacted by both the location of missense variants within protein domains and the combination of allele types. The approach developed may be generally applicable to autosomal recessive disorders related to genes which code for proteins with multiple domains and multiple functions.

Session Title: Mendelian Phenotypes Poster Session III

PB4844 Natural history and molecular characterization of patients with Metaphyseal Enchondromatosis with D-2-Hydroxyglutaric aciduria (MCHGA)

Authors:

C. Montano[1], E. Wohler[2], J. Kim[2], G. Gottesman[3], L. Burrage[4], G. Ascone[1], C. Ferreira[1], N. Sobreira[2]; [1]Natl. Human Genome Res. Inst./NIH, Bethesda, MD, [2]McKusick-Nathans Dept. of Genetic Med., Johns Hopkins Univ. Sch. of Med., Baltimore, MD, [3]Shriners Hosp. for Children, Tampa, FL, [4]Baylor Coll. Med., Houston, TX

Metaphyseal enchondromatosis with D-2-hydroxyglutaric aciduria (MCHGA; OMIM 614875) is characterized by D-2-hydroxyglutaric aciduria and multiple enchondromas. MCHGA has been reported in 12 probands, all presenting in infancy with widespread enchondromatosis. Five probands had global developmental delay, and four had vascular malformations. While its molecular cause is not well known, four probands had somatic mosaic isocitrate dehydrogenase 1 (*IDH1*) gain-of-function variants identified in blood. One patient developed acute myeloid leukemia, but other malignancies had not been reported in probands with MCHGA to date. To better understand the phenotypic spectrum of patients with MCHGA, we characterized the natural history and molecular findings of ten additional patients investigating their germline and/or somatic DNA with whole exome sequencing and/or digital PCR targeting the p.Arg132Cys and p.Arg132His variants in *IDH1* with a sensitivity of >0.1% VAF.
All probands presented with widespread enchondromas in early childhood and normal cognition. While four had motor developmental delay secondary to skeletal deformities, and three had vascular anomalies, none had hypotonia or other dysmorphic features as previously described in patients with MCHGA. Two adult probands presented with chondrosarcoma, and a third pediatric patient had an enchondroma of the skull base with borderline chondrosarcoma characteristics. Ours is the first report of an association between MCHGA and chondrosarcomas. Eight patients underwent molecular testing; most (7/8) had positive findings. Two patients had *IDH1*-p.Arg132Cys variants identified in their enchondromas (VAF 26-46%) as well as in blood (0.8%) or saliva (1.5%). Three patients with either negative or unavailable somatic testing had positive molecular testing in saliva or blood (VAF 0.4-5.5%). On the other hand, two patients with an *IDH1*-p.Arg132Cys variant in the tumor tissue had negative testing for that variant in blood or saliva. The proband with the most severe phenotype exhibited a third *IDH1* variant (*IDH1*-p.Arg132Gly) in bone marrow (VAF 37%), which had not been previously described.
Our novel findings of the association of chondrosarcomas and vascular anomalies with MCHGA, as well as the identification of low levels of mosaicism detected in saliva or blood, highlight the exciting possibility of early non-invasive diagnosis of MCHGA with digital PCR and the importance of thoughtful tumor surveillance in these patients.

Session Title: Mendelian Phenotypes Poster Session I

PB4845 Neuroaxonal dystrophy with osteopetrosis syndrome caused by a novel biallelic missense homozygous variant in *SCYL2*

Authors:

D. Chitayat, O. Greenberg, Y. Fisher, A. Staines, P. Shannon; Univ. of Toronto, Toronto, ON, Canada

Neuroaxonal dystrophy with osteopetrosis syndrome (OMIM # 600329) was first by Fitch et al., (1973) in a consanguineous Moroccan Jewish family. Further cases were reported by Ambler et al., (1983), Jagadha et al., (1988), and Reese et al., (1995) suggestive of a single gene disorder with autosomal recessive inheritance. However, to date, no genetic variant was linked to this disease. We report sibs, born to consanguineous Pakistani parents. Both were identified prenatally with cerebral ventriculomegaly and agenesis of the corpus callosum. Autopsy done on both showed similar abnormalities including facial dysmorphism including high arched palate, flattened nasal tip with anteverted nose, a long philtrum and micrognathia. Hands showed bilateral clinodactyly. Radiographic images were suggestive of osteopetrosis. Brain histopathology showed dilated ventricles, agenesis of the corpus callosum and hypoplastic pyramids and peduncles. Microscopic examination revealed multiple axonal swellings in the central and peripheral nerve systems. The bones demonstrated thin metaphyseal cortical bone with calcified chondroid and hypertrophic osteoclasts with intracytoplasmic inclusions. Electron microscopy from the spinal cord revealed aggregates of irregular membrane bound organelles in neuronal axons. Trio whole exome sequencing identified SCYL2 (NM_017988) c.902G>A; p. Arg301His, variant, heterozygous in parents and homozygous in affected fetuses. This variant was previously classified as a variant of unknown significance. SCYL2 is a regulatory protein pseudokinase linked to Clathrin-coated vesicle trafficking. Mutations in other proteins of this system have been associated with neuro-developmental disorders SCYL2 truncating variants have recently been implicated in arthrogryposis as well. We suggest the SCYL2, p. Arg301His variant as the first described pathogenic mutation in neouroaxonal dystrophy with osteopetrosis syndrome. Ref: Fitch et al., Arch. Path. 95: 298-301, 1973; Ambler et al., Neurology 33: 437-441, 1983. Jagadha et al., Acta Neuropath. (Berlin) 75: 233-240, 1988; Rees et al., Pediat. Neurosurg. 22: 321-327, 1995

Session Title: Mendelian Phenotypes Poster Session II

PB4846 Neurofilament light chain levels in cerebrospinal fluid is useful for evaluating the disease activities of cerebral adrenoleukodystrophy

Authors:

T. Kakumoto[1], T. Matsukawa[1], H. Ishiura[2,1], H. Mori[3], S. Tsuji[1,4], T. Toda[1]; [1]The Univ. of Tokyo, Bunkyo-ku, Tokyo, Japan, [2]Okayama Univ., Okayama, Japan, [3]Jichi Med. Univ., Tochigi-ken, Japan, [4]Intl. Univ. of Hlth.and Welfare, Chiba-ken, Japan

Objective: Adrenoleukodystrophy (ALD) has a poor prognosis when it progresses to the cerebral form (CALD). The aim of this study is to investigate whether cerebrospinal fluid (CSF) neurofilament light chain (cNfL) serves as a sensitive biomarker for evaluating the disease activity of ALD and for assessing response to hematopoietic stem cell transplantation (HSCT). **Methods:** First, we conducted a cross-sectional study of 45 male ALD patients confirmed to carry pathogenic variants in *ABCD1*. The cNfL levels in the patients with the cerebral form of ALD (CALD) or the cerebello-brainstem form of ALD (CBALD) were compared with those in patients with adrenomyeloneuropathy (AMN). The Spearman's rank correlation coefficient was used to investigate the correlation between cNfL levels and brain-MRI-based Loes severity scores. Next, a longitudinal analysis was performed on a subset of the cross-sectional study population that underwent multiple CSF examinations. A longitudinal analysis of cNfL levels was further conducted to investigate the response to HSCT. **Results:** The cNfL levels in 22 patients with CALD were significantly higher than those in 18 patients with AMN (median, 5,550 vs 1,420 pg/mL; $p<0.001$). The cNfL levels in five patients with CBALD were markedly high (median 10,100 pg/mL). Logistic regression analysis revealed that the cutoff cNfL level of 1,920 pg/mL showed good sensitivity (95.5%) and specificity (83.3%) for distinguishing CALD from AMN. cNfL levels showed a positive correlation with the Loes scores of brain lesions ($p<0.001$). A longitudinal analysis revealed that the cNfL levels in three AMN patients who later converted to CALD was increased above the cutoff cNfL level at the time of conversion or increased even during the period of AMN, whereas the cNfL levels in four AMN-staying patients were consistently below the cutoff cNfL level for 3.5-10.9 years. In ten ALD patients who underwent HSCT, their cNfL levels decreased 3-24 months after HSCT. In the two patients who showed progressive cognitive decline, whose substantial increase in the cNfL was observed. **Conclusions:** The cNfL level is useful for evaluating the disease activities of CALD and CBALD. Longitudinal measurement of cNfL level may be a good biomarker to detect the conversion of AMN to CALD. The cNfL level may reflect the response to HSCT.

Session Title: Mendelian Phenotypes Poster Session III

PB4847 Neurogenetic syndromes with cerebral palsy revealed by whole-exome sequencing

Authors:

N. Okamoto[1], E. Nishi[1], Y. Hasegawa[1], K. Yanagi[2], T. Kaname[2], F. Miya[3], M. Yamada[4], H. Suzuki[5], T. Takenouchi[6], K. Kosaki[3]; [1]Dept. of Med. Genetics, Osaka Women's and Children's Hosp., Izumi, Osaka, Japan, [2]Dept. of Genome Med., Natl. Res. Inst. for Child Hlth.and Dev., Tokyo, Japan, [3]Ctr. for Med. Genetics, Keio Univ. Sch. of Med., Tokyo, Japan, [4]Ctr. for Med. Genetics, Keio Univ., Tokyo, Japan, [5]Inst. of Med., Univ. of Tsukuba, Tsukuba, Japan, [6]Dept. of Pediatrics, Keio Univ. Sch. of Med., Tokyo, Japan

[Background] Cerebral palsy (CP) is a group of permanent disorders characterized by abnormalities in movement or posture. It is attributed to non-progressive disturbances that occurred in the developing fetal or infant brain. CP-like disorders may clinically resemble CP, but do not fulfill CP criteria and often have a progressive course and/or neurodevelopmental regression. Common environmental and genomic factors may cause CP. The number of genetic conditions associated with CP has rapidly increased with the development of molecular analyses. The aim of the present study was to delineate the genotypic and phenotypic spectra of children with CP. We performed whole-exome sequencing (WES) or whole-genome sequencing (WGS). We established a molecular diagnosis in 21 out of 31 families with CP or CP-like disorders. **[Methods]** Blood samples were collected from patients and their parents after they provided written informed consent. WES or WGS was used to identify causative variants. The present study was included in the Initiative on Rare and Undiagnosed Diseases project, which was established in Japan to provide accurate diagnoses, discover etiologies, and, ultimately, provide cures for rare and undiagnosed diseases. **[Results]** Causative variants in 11 different genes were identified in 21 (68%) out of 31 probands: *CTNNB1* (n = 4), *KIF1A* (n = 3), *SPAST* (n = 3), *NKX2-1* (n = 2), *CYP2U1* (n = 2), *ATP1A3*, *GNAO1*, *HPDL*, *PLP1*, *PYCR2*, and *TBCD* (n=1). Brothers with *TBCD1* variants showed cerebellar atrophy. **[Conclusion]** Some CP or CP-like disorders may have unique characteristics in the form of a constellation of clinical symptoms, which may facilitate their recognition and shorten the diagnostic process. For example, pathogenic variants in the *CTNNB1* gene have been identified in patients with various diseases, including neurodevelopmental disorder with spastic diplegia and visual defects. The present results will broaden the known clinical manifestations of neurogenetic syndromes with CP or CP-like disorders. A definitive genetic diagnosis is beneficial for the genetic counseling and clinical management of individuals. The present results indicate that CP without apparent environmental factors needs to be included in the current recommendation of WES or WGS in the diagnostic evaluation of individuals with neurodevelopmental disorders.

Session Title: Mendelian Phenotypes Poster Session I

PB4848 New phenotypic features in Osteoglophonic dysplasia.

Authors:

A. Othman[1], H. Babcock[1], C. Gill[2], J. Fraser[3,4,5], D. Regier[5], K. Simpson[5], C. Ferreira[1]; [1]Skeletal Genomics Unit, Natl. Human Genome Res. Inst., NIH, Bethesda, MD, [2]Dept. of Orthopaedic Surgery, Univ. of Texas Southwestern Med. Ctr. and Scottish Rite Hosp. for Children, Dallas, TX, [3]Prenatal Pediatrics Inst., Children's Natl. Hosp., Washington, DC, [4]Ctr. for Genetic Med. Res., Children's Natl. Hosp., Washington, DC, [5]Rare Disease Inst., Children's Natl. Hosp., Washington, DC

Osteoglophonic dysplasia (OGD) is a rare autosomal dominant skeletal disorder with 16 patients described. It is caused by distinct monoallelic *FGFR1* pathogenic gain-of-function variants—currently five causative variants are known. The *FGFR1* gene encodes a fibroblast growth factor receptor protein crucial for the osteogenesis of the axial and craniofacial skeleton. Classical features of OGD include pancraniosynostosis, prominent forehead, proptosis, widely-spaced eyes, low-set ears, midface retrusion, short nose, prognathism, impacted teeth, and rhizomelic shortening of the limbs. Radiographs show a characteristic copper-beaten appearance of the skull and multiple long bone lytic lesions corresponding to non-ossifying fibromas. Patients can develop hypophosphatemia related to increased levels of FGF23, a phosphaturic hormone. Here, we present two previously unreported, unrelated pediatric patients with OGD due to the pathogenic variant c.1141T>C, p.(Cys381Arg) in exon 9 of the *FGFR1* gene, which is predicted to disrupt the transmembrane domain of the gene. Both patients were initially given a diagnosis of Pfeiffer syndrome, an allelic but distinct syndromic craniosynostosis disorder caused by other gain-of-function variants in *FGFR1*. Both patients exhibited the traditional clinical and radiological features of the syndrome. Interestingly, the two patients showed increased frontal temperature and sensitivity to heat with excessive sweating. Their physical examination also revealed a common bilateral overlapping pattern of the toes with the second and fourth toe either overlapping or underriding the third. These two features have not been reported in the literature before. This report aims to highlight the importance of distinguishing OGD from allelic disorders (Pfeiffer syndrome and others) and expand upon the previously reported clinical phenotype of OGD including possible genotype-phenotype correlations.

Session Title: Mendelian Phenotypes Poster Session II

PB4849 NGS driven identification of a rare homozygous variant in the *GNE* gene, causing hereditary inclusion body myopathy type 2

Authors:

Z. Ali[1], J. Klar[2], A. Fatima[3], U. Abdullah[4], K. Khan[5], S. Baig[6], N. Dahl[7]; [1]Univ. of Swat, Pakistan, Charbagh, Swat, Pakistan, [2]IGP, Uppsala, Sverige, Sweden, [3]Aga Khan Univ., Karachi, Karachi, Pakistan, [4]UIBB, PMAS-Arid Agriculture Univ., Rawalpindi, Pakistan, [5]Lurie Children s Hosp., Chicago, IL, [6]Dept. of Biological and BioMed. Sci., Aga Khan Univ., Karachi 74000, Pakistan, Karachi, Pakistan, [7]Uppsala Univ, Uppsala, Sweden

Hereditary inclusion body myopathy type 2 (HIBM2; OMIM: 600737) is a rare autosomal recessive neuromuscular disorder caused by mutations in the *GNE* (OMIM: 603824) gene. The associated clinical manifestations are non-specific including foot drop, gradual progressive weakness in distal and proximal extremities of lower and upper limbs (sparing quadriceps muscles). Muscle biopsy show inclusion bodies and rimmed vacuoles within affected muscles. The clinical presentation in inclusion body myopathies are non-specific and genetic investigations using whole exome sequencing (WES) has now emerged as an important diagnostic tool. We identified a novel homozygous *GNE* missense variant in three Pakistani siblings with myopathy. Two additional siblings and parents were healthy. Data from WES were filtered against a recessive model of inheritance, disease relevance, presence in population databases, or predicted to be benign by various bioinformatics tools. The analysis revealed a single homozygous gene variant in the *GNE* gene (NM_001190384:c.289T>C; p.(Tyr97His)) in the three affected children. The two healthy siblings were either heterozygous or homozygous for the wild type allele and the parents were heterozygous for the *GNE* gene variant. The identified variant is highly conserved and was predicted to be ’deleterious’ by various bioinformatics prediction tools. Bidirectional Sanger sequencing was used to confirm the segregation of the *GNE* gene variant. The present study expands the mutations spectrum in the *GNE*-gene, with possible implications for improved understanding of disease pathophysiology. Furthermore, the current study highlights the importance of NGS, in particular WES, for a specific diagnosis in hereditary myopathy.

Session Title: Mendelian Phenotypes Poster Session III

PB4850 *NKX6-2*- related leukodystrophy: natural history study, clinical phenotype, biomarkers, and gene therapy

Authors:

K. Zhelcheska, V. Chelban, H. Houlden; UCL Queen Square Inst. of Neurology, London, United Kingdom

Background/ Objectives: Hypomyelinating leukodystrophies are a heterogeneous group of genetic disorders characterized by detectable myelin deficiency in the brain with a wide spectrum of phenotypes. A phenotype associated with bi-allelic mutations in the *NKX6.2* gene that leads to spastic ataxia 8 (SPAX8), autosomal recessive, with hypomyelinating leukodystrophy, has been described. Despite the discovery of various other genetic forms of hypomyelinating leukodystrophies, many remain unsolved.**Methods:** The study included affected individuals with spastic ataxia and hypomyelination from unrelated families, and homozygosity mapping and exome sequencing were implemented to identify and characterize the causal variants in *NKX6-2*.**Results:** 33 individuals from 21 families carrying compound heterozygous and homozygous pathogenic variants in *NKX6-2* were identified. The key neuroimaging feature in most cases was of hypomyelination.**Conclusion:** Phenotypic and neuroimaging expression in *NKX6-2* mutations vary from a complex, neonatal onset at the severe end to a childhood onset at the milder end of the spectrum. We plan to expand our analysis to conduct a natural history study to evaluate disease progression, and to identify serial biomarkers. There is an important need to develop gene transfer methods to replace and rescue *NKX6-2* protein loss.

Session Title: Mendelian Phenotypes Poster Session II

PB4852 Novel and recurrent variants in *PAX6* in four patients with ocular phenotypes from Southeast Asia

Authors:

E. Tan, J. Goh, A. Lai, H. Wei, S. Jamuar; KK Women's and Children's Hosp., Singapore, Singapore

Aniridia is an autosomal dominant condition characterized by the complete or partial absence of the iris, often with additional presentations such as foveal hypoplasia, nystagmus, cataract, glaucoma, and other ocular abnormalities. Most cases are caused by heterozygous mutations in the paired box 6 gene (*PAX6*) which codes for a transcription factor that regulates eye development. We report the identification of *PAX6* variants in four patients of different ethnic backgrounds with ocular phenotypes. Two of the variants are recurrent single-nucleotide substitutions - one is a substitution in the first codon which had been reported in a Chinese patient with complete aniridia, while our patient had bilateral aniridia. Another recurrent variant maps to intron 5 of the canonical transcript and also exon 5a of an alternative transcript. The substitution is predicted to facilitate its transcription over the canonical transcript, causing it to be expressed at a level that is higher than normal. This patient did not have aniridia but had left eye proptosis and rotatory nystagmus. He also had olfactory neuroblastoma, and eventually passed away from the recurrence and progression of a frontotemporal tumor. The other two pathogenic variants are novel, and are found in two patients with isolated aniridia. One is a single nucleotide duplication in exon 11 while the other involves duplication of 8 nucleotides in exon 13. Both are predicted to lead to premature termination of translation. For the recurrent variants, the comparison of phenotypes for patients with identical variants would shed light on the mechanisms of pathogenesis; and the discovery of two novel variants expands the spectrum of *PAX6* mutations.

Session Title: Mendelian Phenotypes Poster Session III

PB4853 Novel biallelic splice-site variants in *ATP8A2* causing cerebellar ataxia, impaired intellectual development and disequilibrium syndrome-4.

Authors:

D. Raghavan MS[1], V. K Gowda[2], U. kinhal[2], A. S Manjunath[2], .. Babu G[3], A. Srivastava[4]; [1]Sanjay Gandhi Post graduate Inst. of Med. Sci.sSanjay Gandhi Postgraduate Inst. of Med. Sci., Lucknow, India, [2]Indira Gandhi Inst. of child Hlth., Bangalore, India, [3]Head bioinformatics Progenesis India .Ltd, Chennai, India, [4]Sanjay Gandhi Postgraduate Inst. of Med. Sci., Lucknow, India

Recessively inherited variants in ATP8A2 are the genetic basis of cerebellar ataxia, impaired intellectual development and disequilibrium syndrome 4 (CAMRQ4: OMIM#615268). ATP8A2 is a P4 ATPase (adenosine triphosphate) that function in the transport of aminophospholipids, and plays an important role in vesicular trafficking, cellular survival of neurons and cell signaling through active transport of phospholipids across the cell membrane. We included five cases from four families with biallelic pathogenic / likely pathogenic variants in ATP8A2 gene identified by solo whole exome sequencing. The variants identified were validated by Sanger sequencing in patients and parents. The age of presentation ranged from 1 to 14 years with onset of disease in neonatal period to infancy in all cases. Male: female ratio was 4:1. The common features identified in all were severe developmental delay with none attaining neck control, speech delay with absence of expressive and receptive language, hypotonia noted since birth, feeding difficulties, visual impairment, ptosis, areflexia, involuntary movements and microcephaly. The onset of movement disorder was before 6 months of age: choreoathetosis in four and dystonia in one case respectively, Visual complaints noted were optic atrophy (3/5), ophthalmoplegia (3/5), hypopigmented iris (1/5) and infero-nasal lens dislocation (1/5). Facial dysmorphism was noted in all with a varying combination of broad forehead, up turned nose, low set ears, triangular face, and gum hypertrophy. Hearing was impaired in 3/5 cases. Neuroimaging, Electroencephalogram and nerve conduction studies were normal in all. Splice variants (c.780-2A>G, c.3272+1G>A and c.3019-1G>A) in homozygous state were identified of which two were novel. Variant c.3272+1G>A was recurrent and appeared in two unrelated families. All were pathogenic/ likely pathogenic as per ACMG classification. Minigene assay followed by Western blot confirmed pathogenicity of the identified splice site variants. AutoMap analysis showed shared regions of homozygosity suggesting a founder effect for the recurrent variant. The unusual features reported were facial dysmorphism in all, lens dislocation and iris hypopigmentation in one case. To conclude, The ATP8A2 mutation series shows multi system involvement including both upper and lower motor neurons, basal ganglia, cerebellum, optic nerve and retina, auditory system. This series is the first from India and adds to the existing literature on the disease.

Session Title: Mendelian Phenotypes Poster Session I

PB4854 Novel facial gestalt and cancer risk in patients with Beckwith-Wiedemann spectrum caused by CDKN1C mutations

Authors:

S. Klein[1], E. Hathaway[1], A. George[1], J. Kalish[2]; [1]CHOP, Philadelphia, PA, [2]Children s Hosp. of Philadelphia, Philadelphia, PA

Introduction: Beckwith-Wiedemann Spectrum (BWSp) is caused by genetic and epigenetic alterations in genes on chromosome 11 that regulate cell growth and division. Dysregulation of the maternally expressed cell cycle regulating gene *CDKN1C*, cyclin-dependent kinase inhibitor 1C, occurs in about 60% of cases of BWSp. However, given the variable phenotype in BWSp due mosaic alterations in most patients with BWSp, the specific phenotype and role *CDKN1C* is not well defined. Here, we study the role of *CDKN1C* in the broader BWSp phenotype caused by *de novo* or maternally inherited variants. In families, the mutation can be present for several generations prior to an affected child being born from a carrier mother. Notably, patients with *CDKN1C* variants may not present with classic BWS and tumor risk compared to the broader BWSp population. **Methods:** We performed a comprehensive literature search using the terms "BWS", "Beckwith-Wiedemann Syndrome/Spectrum", and "*CDKN1C*" to identify 116 cases of BWS-*CDKN1C*. We then assessed a previously unreported cohort of 15 patients with BWSp with *CDKN1C* variants to identify a facial dysmorphia pattern that can be distinguished from other patients with BWSp. The facial features were described by at least two geneticists and compared/contrasted to reach a consensus list of features for each patient. Cardinal and suggestive BWS features were also evaluated for all patients reported and tumor risk was established based on available reports. **Results:** The most common phenotypes were macroglossia (96%), macrosomia (48%), ear creases (86%), and abdominal wall differences (74%). Facial features distinct from other epigenotypes of BWSp included down-slanting palpebral fissures with almond-shaped eyes, hypertelorism, and telecanthus. Tumors within this group, of 131 total patients, included four cases of neuroblastoma (3%), one case of acute lymphoblastic leukemia (<1%), and one case of melanoma (<1%). No cases of hepatoblastoma or Wilms tumor (the most common BWSp tumors) were reported. **Conclusion:** Overall, this study identifies unique features associated with *CDKN1C* variants in BWSp, specifically a facial gestalt, enabling more accurate clinical correlation and management. These findings also clarify the cancer risk associated with BWSp and *CDKN1C* variants, guiding surveillance, and treatment strategies. Lastly, identifying the etiology enables more precise counseling and family planning.

Session Title: Mendelian Phenotypes Poster Session II

PB4855 Novel mutations in BLOC-1 and BLOC-2 HPS genes in patients with non-syndromic ocular albinism: using NGS to improve diagnosis and follow-up

Authors:

R. Artuso[1], L. Tiberi[1], C. Rocca[2], S. Bargiacchi[1], G. Bacci[3], E. Marziali[3], P. Fortunato[3], S. Landini[1], E. Dirupo[1], L. Zaroili[1], D. Vergani[1], E. Sieni[4], M. Coniglio[4], L. Pieri[5], R. Caputo[3], A. Peron[1,2]; [1]Med. Genetics Unit, Meyer children's Hosp. IRCCS, Florence, Italy, [2]Dept. of Clinical and Experimental BioMed. Sci. "Mario Serio", Univ. of Florence, Florence, Italy, [3]Pediatric Ophtalmology Unit, Meyer Children's Hosp. IRCCS, Florence, Italy, [4]Dept. of Paediatric Oncohematology, Meyer children's Hosp. IRCCS, Florence, Italy, [5]Ctr. for Bleeding Disorders and Coagulation, Careggi Univ. Hosp., Florence, Italy

Background: Hermansky-Pudlak syndrome (HPS) is an autosomal recessive disorder characterized by oculocutaneous albinism (OCA) or ocular albinism (OA), a bleeding diathesis, and, in specific subtypes, organ involvement (pulmonary fibrosis, granulomatous colitis, or immunodeficiency). To date, 11 HPS subtypes (HPS-1 to HPS-11) have been molecularly identified. Methods: We performed trio exome sequencing in paediatric patients diagnosed with OCA/OA at Meyer Children’s Hospital, Florence. Results: In seven7 paediatric patients with apparently isolated OCA/OA we unexpectedly identified pathogenic variants in HPS-associated genes. Six were novel variants in *BLOC1S5, HPS3, BLOC1S3, HPS6* and three were previously deported changes in *HPS6 and HPS5.* All the novel variants were classified as pathogenic and most of them are frameshift, as previously reported.All patients showed an apparently isolated ocular phenotype with no organ involvement. Only one of two siblings with variants in *BLOC1S5* developed a pilocytic astrocytoma, deemed to be unrelated to HPS. Conclusions: Our results reinforce genotype-phenotype correlations for BLOC-1 and BLOC-2 deficiency, improving also the recognition of mild phenotypes. Reporting novel variants especially in *BLOC1S5* and *BLOC1S3* is important given the small number of patients described in the literature. Also, exome sequencing allowed reaching a molecular diagnosis that was more accurate than the clinical one and starting specific hematologic follow-up.

Session Title: Mendelian Phenotypes Poster Session III

PB4856 Novel pathogenic variants identified in BBS genes through Genome Sequencing - BBS1 exon 10-11 deletion, BBS9 stop-gain SNV.

Authors:

J. Li[1], E. Tavares[1], J. L. Duncan[2], A. Vincent[1], E. Heon[1]; [1]The Hosp. for Sick Children, Toronto, ON, Canada, [2]UCSF, San Francisco, CA

Background: Bardet-Biedl syndrome (BBS) is a rare autosomal recessive disorder that causes severe visual loss in affected patients. Approximately 80% of the clinically examined BBS cases have been associated with variants in over 21 BBS genes with ~ 20% cases remaining unsolved by the clinical genetic testing. This study aims to identify the causative variants in the unsolved BBS cases using whole genome sequencing (GS). Method: Two BBS cases were recruited after negative clinical genetic testing results. Genomic DNA of the patients was extracted from peripheral blood. GS was analysed with a comprehensive variant filtering protocol. Structural variants, transposable elements, and single nucleotide variants (SNVs) were first filtered with BBS gene panel. Pathogenic variants identified in BBS genes were confirmed by visualizing read files and segregation analysis via Sanger sequencing. Results: Case 1: A female affected with BBS carried the most common pathogenic variant in BBS1, NM_024649.5(BBS1): c.1169T>G; p.(Met390Arg), identified through clinical genetic testing. Following GS, comprehensive variant filtering identified a heterozygous novel structural variant NM_024649.5(BBS1): c.830+554_1110+1052del; p.(Asp278Metfs*3). The 3kbp deletion in BBS1 was predicted to cause a loss in exons 10-11, resulting in frameshift and premature stop codon. A different BBS1 exon 10-11 deletion has been reported in ClinVar as pathogenic. Segregation analysis of the deletion confirmed maternal inheritance. Case 2: This BBS patient did not have any pathogenic variants identified through clinical genetic testing. Comprehensive variant filtering of the GS identified two rare (gnomAD AF < 0.01%), highly conserved (phylopVert100 > 9) stop-gain SNVs in exons 8 and 9 of BBS9. Segregation analysis confirmed that NM_198428.3(BBS9): c.724G>T; p.(Gly242*) was on the paternal allele, and NM_198428.3(BBS9): c.966G>A; p.(Trp322*) was inherited from the mother. While p.(Trp322*) was reported once in ClinVar, p.(Gly242*) has not been reported yet. p.(Gly242*) is located in the PTHB1 N-terminal domain, and loss-of-function BBS9 variants are known to be pathogenic by causing an absent or disrupted protein product. Conclusion: Here we report a novel large exonic deletion on BBS1 and a novel stop-gain SNV on BBS9, both suggested pathogenic by ACMG classification. Unsolved BBS patients could benefit from GS for a more conclusive diagnosis. The identification of the two novel BBS variants reinforces the importance of GS in rare disease diagnosis.

Session Title: Mendelian Phenotypes Poster Session I

PB4857 Novel pathogenic variants identified in TRAPP genes in families with Neurodevelopmental Disorders

Authors:

R. Mohammad, R. Maroofian, H. Houlden; UCL Queen Square Inst. of Neurology, London, United Kingdom

Transport Protein Particle (TRAPP) complex plays a pivotal role in ER-to-Golgi transport and contributes to the regulation of multiple membrane trafficking pathways. The precise interaction of different proteins between cellular compartments is fundamental to conducting specialised functions. In humans, TRAPP proteins fall under two related complexes: TRAPP II and TRAPP III. While TRAPP proteins are all related, it is evident that some proteins function independently of the complex. As a result, variants in the genes that encode these proteins have been associated with a spectrum of human diseases, including an emerging set of disorders known as TRAPPopathies, which comprise of Neurodevelopmental Disorders (NDDs). This study used a large-scale, genotype-first approach where 33 causal candidate variants were found in 41 patients with undiagnosed rare NDDs, some of which have been collected as part of the SYNAPS project. Exome sequencing data was screened in 14 known TRAPP genes linked to NDDs. Seven novel homozygous variants were identified in TRAPPC6B c.267+3A>G, c.445+5G>T, TRAPPC9 c.676T>G (p.Tyr226Asp), c.1351G>A (p.Gly451Ser), c.2279-3C>G and TRAPPC11 c.560+3A>T, c.373A>G (p.Arg125Gly) that show evidence of pathogenicity with some of them confirmed by Sanger sequencing. This study provides valuable genomic data for future functional studies to better understand the role of the TRAPP complex in membrane trafficking pathways and how pathogenic variants in TRAPP genes can lead to NDDs. By developing our understanding of the pathogenic mechanisms involved, we hope that this will aid in the development of potential therapeutic targets.

Session Title: Mendelian Phenotypes Poster Session II

PB4858 Novel phenotype associated with homozygous pathogenic variant in the *POP1* gene

Authors:

M. Michelson, K. Yosovich, S. Bahar, M. Ginsberg, D. Lev; Wolfson Med Ctr, Holon, Israel

The biallelic variants of the *POP1* gene are associated with the anauxetic dysplasia (AAD OMIM 607095), a rare skeletal dysplasia, characterized by prenatal rhizomelic shortening of limbs and generalized joint hypermobility. Affected individuals usually have normal neurodevelopmental milestones. Here we present three cases from the same family with a pathogenic homozygous *POP1* variant and completely novel phenotype: a girl with global developmental delay and autism, microcephaly, peculiar dysmorphic, features, and multiple congenital anomalies. Two consequent pregnancies were terminated due to multiple congenital malformations. Fetal DNA samples revealed the same homozygous variant in the *POP1* gene. RNA expression studies revealed decreased transcript of the POP1 compared to the control. To our knowledge, this is the first report of a new syndrome, associated with a pathogenic variant in *POP1*. Our findings expand the phenotypic spectrum of POP1-related disorder.

Session Title: Mendelian Phenotypes Poster Session III

PB4859 Novel *PRKACA* variant in a patient of with multiple congenital anomalies

Authors:

Y. Saito, Y. Kuroda, Y. Enomoto, T. Naruto, K. Kurosawa; Kanagawa Children's Med. Ctr., Yokohama, Japan

Introduction: *PRKACA* and *PRKACB* code for the homologous Cα- and Cβ-subunits of cyclic AMP (cAMP) - dependent protein kinase A (PKA). Heterozygous variants in *PRKACB* cause Ellis-van Creveld syndrome-like multiple congenital anomalies. Clinical features comprise intellectual disability, bilateral postaxial polydactyly, congenital heart defects, and teeth abnormalities. In 2020, Palencia-Campos et al. reported three unrelated patients had the similar phenotype with a heterozygous recurrent variant in *PRKACA* (NM_002730.4) c.409 G>A, p.(Gly137Arg). We describe a female patient with an Ellis-van Creveld syndrome-like phenotype associated with a novel variant in *PRKACA*.Case report: The proposita, first child of nonconsanguineous Japanese parents with an unremarkable family history, was born at 40 weeks gestation. The birth weight was 2725 g (-1.1 SD), length 48.7 cm (-0.6 SD), and occipitofrontal circumference 31.0 cm (-2.0 SD). She had bilateral postaxial polydactyly, syndactyly, atrioventricular septal defect, and right renal agenesis. Her development was delayed; she began speaking meaningful words at 17 months and walking at 23 months. At age three years six months, her gross IQ was 66. She had facial dysmorphism, hirsutism, hyperpigmentation, and hypodontia. Karyotype was 46,XX. Whole exome sequencing revealed a de novo heterozygous variant c.601G>T, p. (Gly201Cys) in *PRKACA*. The variant is not included in the public databases gnomAD, jMorp, ClinVar, and Human Genetic Variation Database. The identified variant in *PRKACA* shows a high CADD score of 26.3. *PRKACA* is a constrained gene intolerant of missense variation (z-score: 2.97). The variant was classified as likely pathogenic according to the ACMG guidelines for the interpretation of sequence variants (PS2, PM2, PP3, PP4). Conclusion: Notably, both the variant detected in our patient and those reported previously are in the functionally important protein kinase domain. Further, the present patient had additional features not previously reported including distinct facial appearance and intellectual disability. Together with these results, our patient will confer to delineate the phenotypic spectrum of the *PRKACA*-related Ellis-van Creveld syndrome-like phenotype.

Session Title: Mendelian Phenotypes Poster Session I

PB4860 Novel *SCN1A* and *SCN1B* variants lead to Dravet syndrome and generalized epilepsy with febrile seizures plus type 1

Authors:

S. Zokaei[1], H. Aryan[2], E. Razmara[3], M. Keykhaei[4], M. Mohammadi[5], B. Yarali[5], M. Zarif-Yeganeh[6], D. Farhud[7], A. Tavasoli[8]; [1]Neurogenetic Diseases Lab., Kennedy Krieger Inst., The Johns Hopkins Univ., Baltimore, MD, [2]Natl. Inst. of Genetic Engineering and Biotechnology, Tehran, Iran, Islamic Republic of, [3]Dept. of Med. Genetics, Faculty of Med. Sci., Tarbiat Modares Univ., Tehran, Iran, Islamic Republic of, [4]Johns Hopkins Univ. Sch. of Med., Baltimore, MD, [5]Ctr. of Excellence for Pediatrics, Dept. of Pediatric Neurology, Children's Med. Ctr., Tehran Univ. of Med. Sci., Tehran, Iran, Islamic Republic of, [6]Cellular and Molecular Endocrine Res. Ctr., Res. Inst. of Endocrine Sci., Shahid Beheshti Univ. of Med. Sci., Tehran, Iran, Islamic Republic of, [7]Sch. of Publ. Hlth., Tehran Univ. of Med. Sci., Tehran, Iran, Islamic Republic of, [8]Pediatric Neurology Div., Children's Med. Ctr., Pediatrics Ctr. of Excellence, Myelin Disorders Clinic, Tehran Univ. of Med. Sci., Tehran, Iran, Islamic Republic of

Background: Dravet syndrome and generalized epilepsy with febrile seizures plus type 1 (GEFSP1) are two important epileptic syndromes in children whose etiology is often imputed to variants in SCN1A and SCN1B, respectively. Herein, we aimed to identify the genetic causes of epilepsy in 5 Iranian families with complex phenotypes.**Methods**: Five patients from five unrelated families were recruited and their clinical features, brain magnetic resonance imaging, and electroencephalography data were analyzed. To investigate the underlying genetic factor(s), the patients were subjected to solo whole-exome sequencing (WES) and in silico analysis were performed to predict the pathogenicity of each identified variant.**Results**: Four novel homozygous variants in SCN1A were identified including NM_001202435.1:c.3139G>T;p.(Glu1047*), NM_001202435.1:c.2632delA;p.(Met878Cysfs*2), NM_001165964.1:c.2331+1G>T;p.(Leu794fs*11), NM_ NM_001165964.1:c.1105_1106insCA;p.(Ser369Glnfs*11). These candidate variants were confirmed using Sanger sequencing. In silico analyses categorized all novel variants as likely pathogenic. We also identified a few clinical manifestations associated with Dravet Syndrome including communication delay, hearing problem, diminished eye contact, and autistic spectrum disorder. Besides, dysphagia, severe digestive problems, and malnutrition were identified in association with GEFSP1.**Conclusion**: We introduced rare extra-neurological findings of Dravet syndrome and GEFSP1 in five unrelated patients harboring variants in SCN1A/1B. This study highlights the beneficial application of solo WES that in complex childhood epileptic syndromes.

Session Title: Mendelian Phenotypes Poster Session II

PB4861 Novel*TAF1*variants associated with reduced RAP74 phosphorylation in patients with X-linked intellectual disability

Authors:

T. Wang[1], M. Han[1], C. Skinner[2], R. Wang[1], S. Srikanth[2], F. Abidi[2], T. Niranjan[1], R. Stevenson[2], C. Schwartz[2]; [1]Johns Hopkins Univ., Baltimore, MD, [2]Greenwood Genetic Ctr., Greenwood, SC

TATA box binding protein associated factor 1 (TAF1) is the largest subunit of the basal transcription factor II D (TFIID), which directs assembly of the pol II preinitiation complex. TAF1 consists of multiple functional domains that possess activities involving promoter binding, histone acetylation, and protein phosphorylation. Previous studies showed that both N- and C-terminal domains have serine kinase activity that selectively phosphorylates transcription factor, RAP74 of TFIIF complex. Pathogenic *TAF*1 variants have been reported in patients with X-linked intellectual disability (XLID) and dystonia-parkinsonism syndromes, however, the underlying mechanism remains poorly understood. We characterized two novel *TAF1* missense mutations that co-segregate with intellectual disability phenotypes in large, multiplex XLID families. One, c.A61C (p.Met21Leu), is in the N-terminal domain and the other, c.4283A>C (p.Gln1428Pro), is in the C-terminal domain. Both variants involve highly conserved amino acid residues and are predicted to be deleterious by *in silico* analysis. Affected males from these families share variable intellectual disability and dysmorphic features. To assess impact of these two variants on TAF1 function, we investigated RAP74 phosphorylation in lymphoblasts from affected males and normal controls using a RAP74 phosphorylation-site specific antibody. Our studies identified a consistently reduced RAP74 phosphorylation in lymphoblasts from one male carrying Met21Leu and two males carrying Gln1428Pro from the two XLID families as compared to normal controls. RAP74 regulates basal transcription by facilitating the assembly of preinitiation complex and modulating Pol II-elongation and pausing. Transcriptome studies are being carried out to identify the transcriptional profile of dysregulated genes that are critical to development of intellectual function. These studies imply a novel mechanism of reduced RAP74 phosphorylation due to pathogenic variants in *TAF1* associated with intellectual disability.

Session Title: Mendelian Phenotypes Poster Session III

PB4862 Null and missense mutations of *ERI1* cause a recessive phenotypic dichotomy in humans

Authors:

P. Campeau[1], L. Guo[2], S. Salian[3], N. Rath[4], J. Rousseau[5], S. Ehresmann[1], S. Moosa[6], J. Clayton-Smith[7], S. Banka[8], A. Jackson[9], T. Brunet[10], I. Hüning[11], H. Ohashi[12], C. Bupp[13], N. Miyake[14], N. Matsumoto[15], R. Mendoza-Londono[16], G. Costain[17], N. Di Donato[18], T. Yamada[19], G. Nishimura[20], B. Wollnik[21], A. Megarbane[22], J. Rosenfeld[23], J. Rosenfeld[24], V. Heissmeyer[4], S. Ikegawa[25]; [1]CHU Sainte Justine, Montreal, QC, Canada, [2]Xi'an Jiaotong Univ., Xi'an, China, [3]CHU Sainte Justine Res. Ctr., Montreal, QC, Canada, [4]German Res. Ctr. for Environmental Hlth., Munich, Germany, [5]Res. Ctr. of the Sainte-Justine Univ. Hosp., Montreal, QC, Canada, [6]Stellenbosch Univ., Tygerberg, South Africa, [7]St Mary's Hosp, Manchester, United Kingdom, [8]Univ. of Manchester, Sale, United Kingdom, [9]The Univ. of Manchester, Manchester, United Kingdom, [10]Ludwig Maximilian Univ. of Munich, Munich, Germany, [11]Univ. of Lübeck, Lübeck, Germany, [12]Saitama Children's Med Ctr., Saitama, Japan, [13]Spectrum Hlth., Grand Rapids, MI, [14]Natl. Ctr. for Global Hlth.and Med., Tokyo, Japan, [15]Yokohama City Univ Grad Sch Med, Yokohama, Japan, [16]The Hosp. for Sick Children & UofT, Toronto, ON, Canada, [17]The Hosp. for Sick Children, Toronto, ON, Canada, [18]Faculty of Med., TU Dresden, Dresden, Saxony, Germany, [19]Hokkaido Univ. Hosp., Sapporo, Japan, [20]RIKEN Ctr. for Integrative Med. Sci., Tokyo, Japan, [21]Univ Cologne, Cologne, Germany, [22]Inst. Jerome Lejeune, Paris, France, [2324]Baylor Coll. of Med., Houston, TX, [25]IMS, RIKEN, Tokyo, Japan

ERI1 is a 3’-to-5’ exoribonuclease involved in RNA metabolic pathways including 5.8S rRNA processing and turnover of histone mRNAs. Its biological and medical significance remain unclear. Here, we uncover a phenotypic dichotomy associated with bi-allelic *ERI1* variants by reporting eight affected individuals from seven unrelated families. A severe spondyloepimetaphyseal dysplasia (SEMD) was identified in five affected individuals with missense variants, but not in those with bi-allelic null variants, who showed mild intellectual disability and digital anomalies. The *ERI1* missense variants cause a loss of the exoribonuclease activity, leading to defective trimming of the 5.8S rRNA 3’ end and a decreased degradation of replication-dependent histone mRNAs. Patient-derived induced pluripotent stem cells (iPSCs) showed impaired *in vitro* chondrogenesis with down-regulation of genes regulating skeletal patterning. Our study establishes a previously unreported entity and provides a model showing more severe effect of missense alleles than null alleles within recessive genotypes, suggesting a key role of ERI1-mediated RNA metabolism in human skeletal patterning and chondrogenesis.

Session Title: Mendelian Phenotypes Poster Session I

PB4863 Panel based sequencing of the undiagnosed leukodystrophies in Iran.

Authors:

N. Mahdieh[1], B. Rabbani[2], M. Soveizi[3], A. Tavasoli[4], A. Rabbani[2], M. Ashrafi[4]; [1]Rajaie cardiovascular research center, Tehran, Iran, Islamic Republic of, [2]Growth and Dev. Res. Ctr., Tehran, Iran, Islamic Republic of, [3]Azad Univ., North Tehran Branch, Tehran, Iran, Tehran, Iran, Islamic Republic of, [4]Tehran Univ. of Med. Sci.s, Tehran, Iran, Islamic Republic of

Leukodystrophies are a large heterogeneous group of genetic diseases affecting white matter of central nervous system. They are associated with motor loss and cognition problems. Their diagnosis is challenging and half of the patients are undiagnosed. Genetic testing helps to confirm the diagnosis. 152 patients referred to the tertiary hospital were evaluated for genetic testing. Among 44 patients which were not genetically diagnosed, further they underwent panel sequencing and Whole exome sequencing (3 patients). *ARSA* and *SURF1* gene was mutated in 3 patients; *GALC, FUCA1, POLG, EIF2B5, GJC2* each had causal variant in 2 patients, *SUMF1, RNASET2, PEX7 , PEX13, HSD17B4, EIF2B3, EIF2B4, POLR3A, POLR3B , FAM126A, PLA2G6 , RARS, HEXB , GLB1, PPT1,CLN6 , NAGLU , NDUFS1, NDUFS7, and SDHAF1* each gene had pathogenic variant in 1 patient. In 40 patients (90%) the gene and the causal variants were determined; 4 patients did not show any variant that should go for WES. This study was performed from 2016 to 2019; although there have been more patients genetically diagnosed throughout the country but they are being published in future and should further extend our knowledge of the genetic heterogeneity in this region. Early diagnosis of these patients will provide family planning and further future treatments, and managements at earlier stage for some of the patients.

Session Title: Mendelian Phenotypes Poster Session II

PB4864 Patient-derived iPSC model of CLN3 disease.

Authors:

E. Dwojak; Eunice Kennedy Shriver Natl. Inst. of Child Hlth.and Human Dev., Bethesda, MD

CLN3 disease or Juvenile Neuronal Ceroid Lipofuscinoses (JNCL) is a rare, autosomal recessive, pediatric disease. Classic syndromic presentation of the disease includes vision loss, cognitive decline, behavioral dysfunction, seizures, loss of motor function, and significantly reduced lifespan. A vision-only presentation involves later-onset and slower progression of vision loss, without other systemic involvements. The disease is caused by changes on the *CLN3* gene which encodes the lysosomal/endosomal membrane protein, battenin. Over 100 disease-associated variants have been identified in the *CLN3* gene to date, with the most common being a 966-base pair deletion. The exact function of battenin is still being characterized, but its absence is associated with abnormal glycerophosphodiesters accumulation and a lysosomal disorder phenotype. The pathophysiology is complex and not completely understood therefore there is a need for better disease models. We developed and characterized CLN3 patient-derived iPSCs to use as a model for elucidating the disease pathophysiology in different associated cells and tissues. Skin biopsies from six study participants with various genotype and phenotype in an NIH IRB-approved CLN3 natural history study (NCT03307304) were expanded into fibroblasts and reprogrammed into iPSCs using non-integrating Sendai virus technology. CLN3 genotype was confirmed by Sanger sequencing. G-banding karyotype analysis verified that the cells are diploid without significant chromosomal abnormality. Immunocytochemistry and flow cytometry verified pluripotency of the iPSCs. Histopathology of excised teratomas confirmed pluripotent potential of the reprogrammed iPSCs to differentiate into all three germ layers. In addition to these characterizations, we will further discuss characterization of CLN3-specific cellular phenotype, preliminary transcriptomic analyses, and future applications of these iPSC cell models.

Session Title: Mendelian Phenotypes Poster Session III

PB4865 Peripheral nervous system disruption revealed by tissue-specific transcriptome analysis in a humanized mouse model of familial dysautonomia

Authors:

R. Harripaul[1], E. Morini[2], M. Salani[3], E. Logan[2], E. Kirchner[2], J. Bolduc[4], J. Bolduc[2], A. Chekuri[2], B. Currall[5], R. Yadav[6], S. Erdin[2], M. Talkowski[2], D. Gao[2], S. Slaugenhaupt[2]; [1]Massachusetts Gen. Hosp., Cambridge, MA, [2]Massachusetts Gen. Hosp., Boston, MA, [3]MGH, SOMERVILLE, MA, [4]Canada, [5]Boston, MA, [6]MGH, HMS and Broad Inst., Boston, MA

Familial Dysautonomia (FD) is a rare recessive neurodevelopmental disease caused by a splice site mutation in the Elongator acetyltransferase complex subunit 1 (*ELP1*) gene. The mutation induces tissue-specific skipping of exon 20 and reduction of *ELP1* protein primarily in the central and peripheral nervous systems (CNS and PNS, respectively). FD patients exhibit complex neurological phenotypes due to loss of proprioceptive neurons including severe gait ataxia, respiratory failure, and spinal deformities. While the PNS involvement in FD has been well-documented at the histological level, the mechanisms that govern the tissue-specificity and preferential neuronal loss due to Elp1 reduction have not been characterized. This study shows *ELP1* reduction has a direct impact and convergent dysregulation in PNS tissues. We investigated the molecular mechanisms underlying FD using a humanized phenotypic mouse model that recapitulates the human mRNA splicing defect with varying tissue-specific *ELP1* protein levels in a collection of neuronal tissues, including the dorsal root ganglion (DRG), trigeminal ganglion (TG), medulla (MED), cortex, and spinal cord (SC). FD PNS transcriptomes (DRG and TG) had significantly more differentially expressed genes (DEGs) than those in the FD CNS transcriptomes (MED, SC, and Cortex) when compared to control neuronal mouse tissues. In addition to DEGs in each tissue, we identified genes tightly co-expressed and functionally dependent on the level of full-length *ELP1* transcript, defined as *ELP1* dose-responsive genes. Combining the DEGs and *ELP1* dose-responsive genes yielded tissue-specific dysregulated FD signature genes, further forming FD-dysregulated gene networks in PNS. These networks highlighted direct connections between *ELP1*, tRNA genes and a group of genes related to amine- metabolism and synaptic signalling, unifying the seemingly unrelated functions of *ELP1*. Notably, the transcriptomic dysregulation in PNS tissues was similar and enriched for neuronal subtype markers of peptidergic nociceptors and myelinated neurons, pinpointing the specific types of neurons known to be affected in FD. This study identified critical, tissue-specific peripheral nervous system gene networks that underly neuronal development and FD molecular pathology.

Session Title: Mendelian Phenotypes Poster Session I

PB4866 Personalized precision medicine of strategic health action in Niemann-Pick type a/b disease: Case report

Authors:

L. Moreno Giraldo[1], J. Sanchez-Vargas[2]; [1]Univ. del Valle, BOGOTA, Colombia, [2]Univ. Libre, Cali, Colombia

Niemann-Pick diseases (NPD) are lysosomal storage diseases caused by acid sphingomyelinase (ASM) deficiency, which catalyzes the hydrolysis of sphingomyelin (SM) to ceramide and phosphocholine. As a result, the SM and its precursor lipids accumulate in the lysosomes of the cells of the reticuloendothelial system, which produces an abnormal functioning, leading to the inability to degrade macromolecules, forming intracellular inclusions that are deposited in different organs such as the liver, spleen, lungs and brain. There are 5 subtypes, NPD-A and NPD-B are caused by variants in the sphingomyelin phosphodiesterase 1 (SMPD1) gene, which lead to abnormal or defective protein formation, which prevents the movement of lipids out of cells. It has been described that NPD worldwide affects 1 in 120,000-150,000 people and is characterized by autosomal recessive inheritance. We present the case of a 20 month old lactating female patient, with a diagnosis of global developmental delay associated with chronic protein-calorie malnutrition, dysmorphic fasciae, and hepatosplenic syndrome. In the initial approach, infectious and lymphoproliferative diseases were ruled out. Given the clinical complexity of the patient, a targeted clinical exome was requested, finding two variants, one of clinical pathogenic significance and the other probably pathogenic in the SMPD1 gene (compound heterozygous). The enzymatic activity of lysosomal enzymes was requested, pathologically increased biomarker lyso-SM-509 was found, and pathologically decreased ASM activity. Results with which a phenotype-genotype correlation is made with the NPD-A/B. With a defined and precise diagnosis, it is possible to guide health actions, follow-up guidelines, heritability risk assessment through an index case in order to find other possible carriers, carry out complete genetic counseling, implement and start targeted treatments that reduce the morbidity and mortality associated with this pathology, given that there are currently several studies in different phases of research on molecules that may intervene in the natural history of the disease.

Session Title: Mendelian Phenotypes Poster Session II

PB4867 Perturbation of the mTOR signaling pathway due to missense variants in POLR3A cause primary microcephaly

Authors:

E. Makhdoom[1,2], M. Asif[3], A. Moawia[3], U. Abdullah[4], T. Batool[5], S. Khan[6], S. Irshad[5], F. Latif[6], S. Alawbathani[3,7], W. Höhne[3], I. Anjum[8], A. Fatima[9], J. Baig[6], T. Georgomanolis[3], B. Budde[10], D. Zastrow[11,12], J. Kohler[13,12], M. Wheeler[14,12], A. Bashir[15], J. Bernstein[16,17], S. Cirak[18,19], P. Nuernberg[3,19], S. Baig[20,21], M. Hussain[3,19]; [1]Neurochemicalbiology and Genetics Lab. (NGL), Dept. of Physiology, Faculty of Life Sci., Government Coll. Univ., Faisalabad, Pakistan, [2]Cologne Ctr. for Genomics, Cologne, Germany, [3]Univ. of Cologne, Cologne, Germany, [4]UIBB, PMAS-Arid Agriculture Univ., Rawalpindi, Pakistan, [5]Univ. of the Punjab, Lahore, Pakistan, [6]Natl. Inst. for Biotechnology and Genetic Engineering, Faisalabad, Pakistan, [7]GenAlive Lab, Riyadh, Saudi Arabia, [8]Kinnaird Coll., Lahore, Pakistan, [9]Aga Khan Univ., Karachi, Karachi, Pakistan, [10]Univ. zu Köln, Cologne, Germany, [11]Invitae, San Bruno, CA, [12]Ctr. for Undiagnosed Diseases, Stanford Univ., Stanford, CA, [13]Stanford Med., Palo Alto, CA, [14]Stanford Univ, Palo Alto, CA, [15]Punjab dental Coll., Lahore, Pakistan, [16]Stanford Univ Sch Med, Stanford, CA, [17]Div. of Med. Genetics, Dept. of Pediatrics, Stanford Univ., Stanford, CA, [18]Div. of Neuropediatrics, Dept. of Pediatrics and Adolescent Med., Univ. Med. Ctr. Ulm, Ulm, Germany and Med. Faculty Ulm Univ., Ulm, Germany, [19]Ctr. for Molecular Med. Cologne (CMMC), Univ. of Cologne, Faculty of Med. and Univ. Hosp. Cologne, Cologne, Germany, [20]Aga Khan Univ., Karachi, Pakistan, [21]Pakistan Sci. Fndn. (PSF), 1- Constitution Avenue, G-5/2, Islamabad, Pakistan

Background/objectives: Primary Microcephaly is the observation of head circumference >2.5 SDs below the mean for a given sex, age, and ethnicity. It is caused by biallelic variants of 30 genes.
Methods: We analyzed the family through exome sequencing and linkage analysis. The consequences of the candidate variant were explored at cellular and molecular levels by immunofluorescence, immunoblotting, transient expression of tagged proteins, pulldown assay coupled with mass spectrometry, RNA-seq, and knockdown of Polr3a in chicken embryos.
Results: We studied a large six-generation Pakistani family with seven affected members manifested primary microcephaly. The genetic analysis identified a biallelic variant (NM_007055.3: c.40A>G; p.(Lys14Glu)) of *POLR3A*. Studying further patients, we were able to find four compound heterozygous variants of *POLR3A* in two unrelated patients manifesting microcephaly and variable features of global developmental delay, seizures, hypotonia, and dental abnormalities, featuring leukodystrophy. Pathogenic variants in *POLR3A* are known to cause hypomyelinating leukodystrophy but have never been implicated in the etiology of microcephaly. Patient-derived primary fibroblasts and ectopically expressed HeLa cells revealed reduced immunoreactivity of POLR3A. The variant plausibly compromises the interaction of POLR3A with crucial proteins involved in mTOR signaling and gene expression pathways and exerts effects on the differential expression of genes involved in translation and cell cycle. Knocking down of Polr3a in chicken embryos shows striking features of reduced brain size.
Conclusion: Our findings suggest that biallelic disruption of *POLR3A* should be considered in molecular diagnosis of primary microcephaly. Notably, mTOR signaling contributes to determining normal brain size.

Session Title: Mendelian Phenotypes Poster Session III

PB4868 Phenome-wide association study of Sickle Cell Trait in NIH's All of Us Research Program

Authors:

A. Awan[1], H. C. Ramirez[1], T. M. Ferrara[1], A. Williams[1], A. Buscetta[1], R. P. Naik[2], P. Joseph[3], J. M. Keaton[1], J. C. Denny[1], V. L. Bonham[1], H. Mo[1]; [1]NHGRI/NIH, Bethesda, MD, [2]Johns Hopkins Univ., Baltimore, MD, [3]NIAAA/NIH, Bethesda, MD

The HbS variant of HBB gene is highly prevalent worldwide, especially among individuals of African descent. Although individuals with Sickle Cell Trait (SCT), carrying one HbS allele, typically remain asymptomatic, rare cases link carrier status to adverse clinical outcomes. The NIH's All of Us Research Program is a disease-neutral biobank that includes a diverse group of participants with whole genome sequencing (WGS) data linked to Electronic Health Records (EHRs). The goal of this study was to investigate the association of SCT with clinical outcomes, especially vascular occlusive complications. We identified carriers of HbS with rs334. Among the 39,020 individuals of genetically predicted African descent with both WGS and EHR data, 3,363 individuals carried one HbS allele and 73 carried two. Those who had two HbS alleles were excluded from further analysis. We then performed a phenome-wide association study (PheWAS) to compare carriers of a single HbS allele with non-carriers. The phenotypes that had the strongest association with HbS carriers were anemia during pregnancy (OR=2.35 [1.90-2.91]), iron deficiency anemias (OR=1.49 [1.28-1.73]), aseptic necrosis of bone (OR=2.20 [1.58-3.06]), and pulmonary embolism and pulmonary infarction (OR=1.67 [1.34-2.08]). To evaluate the impact of other concomitant hemoglobinopathies, we examined ClinVar variants of the entire HBB gene and extracted the clinically relevant SNPs from All of Us. We excluded individuals who carried any non-HbS hemoglobinopathy variants, especially beta-thalassemia and HbC, from the cohort, resulting in 3,162 single HbS carriers and 31,798 non-carriers. We performed a second PheWAS and observed that the phenotypes most strongly associated with this cohort were anemia during pregnancy (OR=2.27 [1.81-2.84]), iron deficiency anemias (OR=1.49 [1.27-1.74]), gout (OR=1.48 [1.21-1.81]), and abnormal results of kidney function study (OR=2.81 [1.71-4.62]). The association with aseptic necrosis of bone changed to (OR=1.06 [0.68-1.67]), and pulmonary embolism and infarction to (OR=1.40 [1.10-1.78]), below Bonferroni cut-off). Overall, our findings are consistent with previous research. This study demonstrates that the previously reported associations of SCT with vascular occlusion phenotypes, except for the kidney, may be related to concomitant hemoglobinopathies. The current findings provide new insights into phenotypic variations among individuals with SCT. Considering the high prevalence of HbC and beta-thalassemia alleles, it is important for clinicians to investigate other hemoglobinopathies to determine the risks of vascular occlusive complications in SCT.

Session Title: Mendelian Phenotypes Poster Session I

PB4869 Phenotype associations from a comprehensive database of long-chain fatty acid oxidation disorder gene variants

Authors:

H. Richbourg[1], J. Vockley[2], O. K. Japalaghi[1], V. Rangel Miller[1], M. AlSayed[3], M. Kiel[4], S. Mercurio[4], P. Baker, II[5], S. Daugherty[1], A. Khan[6], L. Korngut[7], D. McKnight[8], N. Miller[1]; [1]Ultragenyx Pharmaceutical Inc., Novato, CA, [2]UPMC Children s Hosp. of Pittsburgh and Univ. of Pittsburgh Sch. of Med., Pittsburgh, PA, [3]Coll. of Med., Al-Faisal Univ., Riyadh, Saudi Arabia, [4]GENOMENON, Ann Arbor, MI, [5]Univ Colorado, Aurora, CO, [6]M.A.G.I.C. (Metabolics and Genetics in Canada) Clinic Ltd, Calgary, AB, Canada, [7]Dept. of Clinical NeuroSci.s, Univ. of Calgary, Calgary, AB, Canada, [8]Invitae Corp., San Francisco, CA

Background: Long-chain fatty acid oxidation disorders (LC-FAOD) are a group of six rare, life-threatening conditions typically detected through elevations in plasma acylcarnitine profiles, identified either through newborn screening (NBS) or by testing patients who present clinically. Combined, they are the most common inborn errors identified by NBS. Confirming a diagnosis with genetic testing is critical as limited functional testing is available. However, correlation of clinical and biochemical features with genotype is incomplete, and numerous variants of uncertain significance (VUS) have been reported. This abstract presents findings from a comprehensive database of disease-associated variants.
Methods: This comprehensive database of variants in the six genes associated with LC-FAOD (*ACADVL*, *CPT1A*, *CPT2*, *HADHA*, *HADHB*, *SLC25A20*) integrates data from a systematic review of published medical literature using the Mastermind Genomic Search Engine, a sponsored LC-FAOD gene panel clinical testing program, NBS results, clinical & biochemical phenotypes, and ACMG variant classifications. These data were used to analyze relationships between genotype and phenotype.
Results: As of APR 2023, the database reported 5324 variants from 3146 individuals with one or more LC-FAOD gene variants, representing 947 unique variants: 640 (68%) pathogenic (P) or likely pathogenic (LP) and 294 (31%) VUS. 1830/3146 (58%) patients had a genetic diagnosis (at least two P/LP): *ACADVL*, 678/37%; *HADHA*, 552/30%; *CPT2*, 410/22%; *SLC25A20*, 70/4%; *CPT1A*, 61/3%; *HADHB*, 59/3%. Of 722 unique genotypes observed, 501 (69%) were seen in only one patient. Phenotypic data were reported for 870 patients and overall, the five most frequent phenotypes include hypoglycemia (26%), rhabdomyolysis (22%), myalgia (21%), myoglobinuria (18%), elevated creatine kinase (16%). Beyond these, key phenotypes include*: -* *ACADVL* (n=251): cardiomyopathy (35%)- *CPT2* (n=253): rhabdomyolysis (50%)
- *HADHA* (n=228): retinopathy (21%)
-*HADHB* (n=49): peripheral neuropathy (35%)-*SLC25A20* (n=47): hyperammonemia, arrhythmia (32%)
- *CPT1A* (n=42): liver abnormality (50%).
Notably, for patients under 1 year, hypoglycemia was most common (128; 45%), and for patients over 12 years, myoglobinuria was most common (56; 52%).
Conclusions: Databases are critically important in rare diseases where clinical information on variants is scarce, and VUS are frequent, making it challenging to establish a molecular diagnosis. Genotype-phenotype data increase the diagnostic yield from genetic testing and are an important tool for clinicians to provide prognostic information to patients with LC-FAOD.

Session Title: Mendelian Phenotypes Poster Session II

PB4870 Phenotype semantic similarity-based approach for patient clustering and disease prioritization

Authors:

Y. Hu[1], C. Liu[2], C. Weng[2], K. Wang[3]; [1]Children s Hosp. of Philadelphia, Philadelphia, PA, [2]Columbia Univ., New York, NY, [3]Children's Hosp. of Philadelphia, Philadelphia, PA

Human Phenotype Ontology (HPO)-based analysis has recently become popular for the genomic diagnostics of rare diseases. Such analysis relies mainly on statistical approaches and machine learning to prioritize a list of disease-causing genes for a patient. However, these methods do not effectively use the information on the phenotypes of diseases. It is challenging to efficiently measure similarity between a patient's phenotypes to another patient and match a patient's phenotypes with disease phenotypes. Here, we propose EHRclust, a patient clustering and disease ranking algorithm that measures patient-patient/patient-disease similarity scores based on HPO concepts for EHR data analysis. EHRclust removes time-based batch effects and explicitly models posterior odds by quantifying the likelihood of a pre-specified disease based on the observed set of phenotypes. To evaluate the performance of EHRclust, we simulated HPO terms for a group of patients based on 30 complex diseases. In simulation data, we weaken the underlying phenotypic difference between patients by generating data with "imprecision" and "noise". The disease prediction results of EHRclust on 5 different similarity measures showed that they consistently achieved high accuracy (>95%) in the top 3 candidate diseases. We also applied EHRclust to real datasets with 149 Friedreich Ataxia, Neurofibromatosis, or Marfan Syndrome patients. By using the batch effect removal algorithm, EHRclust is able to identify cluster variations with less informative data. The separation of patients agrees well with diseases, with only 9(6%) patients misclassified in other disease clusters. Current evaluations show that EHRclust offers a proper balance of clustering accuracy and stability. As the scale of EHR studies continues to grow, we believe EHRclust will offer a valuable tool for biomedical researchers to disentangle complex clinical heterogeneity and improve the diagnosis accuracy.

Session Title: Mendelian Phenotypes Poster Session III

PB4871 Phenotype-based prioritization strategy accelerates the identification of real causal variants in Titinopathy patients.

Authors:

Y. Lei[1], W. Yang[2], S. Chan[1]; [1]Dept. of Paediatrics and Adolescent Med., Li Ka Shing Faculty of Med., The Univ. of Hong Kong, Hong Kong, China, [2]Univ. of Hong Kong, Pokfulam, China

The application of next-generation sequencing has led to the identification of over 600 genes associated with neuromuscular diseases (NMD). However, despite this progress, many patients still lack genetically diagnosis due to clinical and genetic heterogeneity of NMD. The traditional NGS data analysis strategy involves annotating and filtering variants based on population frequency, inheritance model, and predicted pathogenicity. Nonetheless, a substantial number of variants that spanning multiple genes will remain even after filtration. It is laborious to investigate each variant manually, and the presence of different genes further complicates the identification of true causal variants. In this study, using phenotype-based prioritization strategy, 6 patients from 4 families were identified casual variants in *TTN* gene. Whole-genome sequencing was conducted for all patient, and whole-exome sequencing was done for their parents. Alignment was performed using the Burrows-Wheeler Aligner, and the GATK best practice pipeline was employed for SNV/Indel calling. The Human Phenotype Ontology (HPO) website was used to convert the clinical presentations into HPO terms, and Exomiser was used for variants prioritization. All identified variants were validated via mRNA and Sanger sequencing. Applying the phenotype-based prioritization strategy, we found 7 variants in the *TTN* gene in the patients and parents. These variants included 4 nonsense variants, 2 splicing variants, and 2 frameshift variants. Further mRNA and Sanger sequencing validated the existence of these 7 variants. Clinical presentations play a crucial role in NMD diagnosis. However, the traditional analysis strategy fails to consider them comprehensively, requiring researchers to manually incorporate clinical information while investigating each variant. This strategy is inefficient for heterogeneous diseases like NMD. To overcome this limitation, we leveraged the Human Phenotype Ontology (HPO) and Exomiser to prioritize variants based on phenotype-disease similarity and genotype pathogenicity. The proposed strategy efficiently identifies true causal variants. This strategy significantly improves the NMD diagnosis rate and shortens the diagnostic odyssey for patients.

Session Title: Mendelian Phenotypes Poster Session I

PB4872 Phenotypic Characterization of Patients with Cleidocranial Dysplasia (CCD)

Authors:

E. Mitchell[1], K. Wade[1], C. Stole[1], N. Sobreira[2], I. Ickow[1], J. Hoover-Fong[1]; [1]Johns Hopkins Univ., Baltimore, MD, [2]Johns Hopkins Univ, Crownsville, MD

Cleidocranial Dysplasia (CCD) is a rare mendelian disease with autosomal dominant inheritance associated with severe abnormalities in bone development. CCD is caused by heterozygous pathogenic variants in *RUNX2* and *CBFB*, proteins that form a heterodimeric transcription factor which regulates osteogenesis. Clinical hallmarks of CCD include hypoplastic or absent clavicles, delayed closure of fontanelles, mild proportionate short stature, and dental anomalies including supernumerary and missing teeth. There is variable expressivity within and among families which has led to inconsistent phenotypic characterization of patients with CCD. In order to elucidate phenotypic variation in CCD, our group has assembled a cohort of 117 patients with CCD from 76 families across five countries for comprehensive phenotyping. The overall goal is to utilize these data with genomic analysis to establish genotype-phenotype correlations between variants in RUNX2 and CBFB as well as potential genetic modifiers of the CCD phenotype. We first conducted a literature search to identify previous phenotypic findings, and then we established a standardized protocol for both physical and oral examination. To better understand the psychosocial and socioeconomic effects of CCD participants, and their caregivers/family members, participants also completed validated quality of life surveys. In total, participants completed 1) a comprehensive physical examination by a skeletal dysplasia specialist, 2) an oral examinations from a craniofacial orthodontist or dentist, 3) a survey of medical history and quality of life, 4) provided DNA samples and detailed family history in the form of pedigrees and 5) conferred access to previous medical records. Physical data was collected manually then transcribed and stored in a Research Electronic Data Capture (REDCap) database. Survey data were collected using Qualtrics. Here we describe four novel and/or poorly defined phenotypes seen in our cohort: osteoporosis/osteopenia, anterior crossbite, knee deformities, and hearing loss/impairment. Future directions include combining these and other phenotypic findings with genomic data to identify genotype-phenotype correlations as well as potential genetic modifiers of CCD disease manifestations. Overall these results will help us better define the CCD phenotype and understand the pathophysiology.

Session Title: Mendelian Phenotypes Poster Session II

PB4873 *PHOX2B* non-polyalanine repeat mutations detected in autopsy cases of sudden unexpected infant death.

Authors:

A. Ueda[1], H. Naito[1,2], Y. Kakimoto[1], M. Osawa[1]; [1]Tokai Univ., Isehara, Kanagawa, Japan, [2]Kindai Univ., Osakasayama, Osaka, Japan

Congenital central hypoventilation syndrome (CCHS) is characterized by impaired ventilatory response to hypercapnia and hypoxemia during sleep. This autosomal dominant disorder is caused by the single gene of paired-like homeobox 2B (*PHOX2B*), which is in turn associated with various phenotypic disorders. Most patients carry expansions of the repetitive alanine tract. Furthermore, the minor part carries non-polyalanine repeat mutation (NPARM) of frameshift, nonsense, and missense, which more often result in severe complications of Hirschsprung disease and neural crest-derived tumors. The diagnosis of sudden infant death syndrome (SIDS) as a cause of death has decreased in mortality statistics for decades. One reason for the decrease is circumvention by examiners because of indistinguishable disorders such as accidental suffocation in bed. Other broad criteria such as sudden unexpected infant death (SUID) and sudden unexpected death in infancy are preferably chosen to indicate causes for such equivocal unclassifiable infant deaths. The involvement of CCHS in SIDS/SUID has been investigated for forensic autopsy cases because of its similarity of nocturnal onset. No apparent involvement of CCHS in SIDS cases has been demonstrated in earlier studies. This report describes a specific investigation of the relation of *PHOX2B* variation with SIDS/SUID. For this study, We analyzed 93 DNA samples of less than one-year-old Japanese SUID cases that were autopsied in our department. These 58 male and 35 female subjects had a mean age of 3.4 months. Unrelated adult Japanese volunteers (n = 942) were used as the control. No polyalanine tract expansion was detected in SUID cases. The allelic frequencies of repeat contractions and SNP (rs28647582) in intron 2 were not significantly different from that in those control group. Further extensive sequencing revealed a heterozygous NPARM of c.905A>C in a sudden death case of a one-month-old male infant. This missense mutation (p.Asn302Thr), registered as rs779068107, was annotated to ‘Affected status is unknown’ but it might be associated with sudden death. NPARM was more plausibly related to sudden unexpected death than expansions because of severe clinical complications. This finding indicates possible CCHS involvement in forensic autopsy cases without an antemortem diagnosis.

Session Title: Mendelian Phenotypes Poster Session III

PB4874 Pontocerebellar hypoplasia type 9: A report of 17 new patients

Authors:

M. Zaki[1], M. Y. Issa[1], H. M. Elbendary[1], S. Abdel Ghafar[2], K. Rafat[1], M. S. Abdel-Hamid[2]; [1]Clinical Genetics Dept., Human Genetics and Genome Res. Inst., Natl. Res. Ctr., Cairo, Egypt, [2]Med. Molecular Genetics Dept., Human Genetics and Genome Res. Inst.,Natl. Res. Ctr., Cairo, Egypt

Pontocerebellar hypoplasia (PCH) is a clinically and genetically heterogeneous group of neurodegenerative disorders of intrauterine onset with 17 different subtypes (PCH1-17). PCH9 is characterized by global developmental delay, postnatal microcephaly, epileptic encephalopathy, and distinctive neuroimaging features comprising cortical atrophy, hypogenesis of corpus callosum, hypomyelination, variable degrees of cerebellar and pontine hypoplasia and figure 8 shape hypoplastic midbrain. It is caused by a biallelic variant in the adenosine monophosphate deaminase 2 (*AMPD2*) gene which is also convicted in autosomal recessive spastic paraplegia 63. Herein, we describe 17 patients from 15 unrelated families from Egypt with PCH9. They were 10 males and 7 females with ages ranging from 3 months to 4y. Parental consanguinity was in 93% of patients. Microcephaly, severe global developmental delay, spasticity, myoclonic, and less frequent generalized seizures were evident in all patients. Fourteen patients were identified by target gene sequencing based on clinical-radiological features while 3 by whole exome sequencing due to the absence of the classic neuroimaging findings. Molecular results presented 14 different variants in the *AMPD2* gene including 12 novel ones. Our report expands the radiological and mutational spectrum of *AMPD2*. To our knowledge, this is the largest series of PCH9 from the same ethnicity.

Session Title: Mendelian Phenotypes Poster Session II

PB4876 Prediction of future Mendelian disease genes: implications for rare and common diseases

Authors:

J. X. Chong[1,2], A. J. Marcello[1], S. Khanam[1], B. M. Werness[3]; [1]Univ. of Washington, Seattle, WA, [2]Brotman-Baty Inst., Seattle, WA, [3]Amazon (unrelated to work at Amazon), Seattle, WA

Approximately ~4,500 genes have been implicated to date to underlie Mendelian conditions (MCs) and ~150 new genes for MCs are discovered each year. In Mendelian gene discovery research, annotations including selective constraint, predicted function, and conservation, as well as variant-level annotations like pathogenicity predictors and frequency are used to prioritize candidate variants and genes in order to identify the underlying gene. Widely-used gene-level constraint metrics include LOEUF, pLI, MOEUF, UNEECON-g, and s_{het}, a per-gene measure of the reduction in fitness for heterozygous carriers of a loss of function variant. These metrics have, by design, lower or minimal sensitivity for genes that underlie MCs that are recessive, have an age-of-onset past reproductive age, have modest or low effects on reproductive fitness, or are only caused by gain-of-function variants (i.e., not caused by loss-of-function). So, while these metrics help prioritize candidate genes for many severe childhood MCs, they are likely to miss many future gene discoveries across the full landscape of MCs.

We used machine learning to develop an ensemble model to predict protein-coding genes that are likely to underlie MCs or "future discoveries." Our model incorporates thousands of features spanning many metrics including conservation, constraint, gene expression, co-expression, gene structure, protein structure, protein-protein interactions/connectivity, regulatory structure, and pathogenicity predictors. All features were selected to minimize leakage, or accidentally training our model on a feature that may directly reveal whether a gene is already known to underlie a MC. We validated our model by excluding 889 genes that were first implicated in a MC or reported as not associated with a phenotype in mutant mice after 2016, followed by training on the remaining genes. Then, we tested our model's performance in the excluded set of 889 genes and achieved high precision (0.977) and recall (0.985). Our model predicts a total of 15,695 protein-coding (>85% of all coding) genes underlie a MC, suggesting that >11,000 genes underlying a MC remain to be discovered. These predictions have implications not just for Mendelian genomics but other areas of human genetics that utilize known genes for Mendelian conditions to identify or prioritize among a gene set of interest.

Session Title: Mendelian Phenotypes Poster Session III

PB4877 Prenatal exome sequencing and in vivo modeling in zebrafish reveals new candidate genes to explain fetal brain anomalies.

Authors:

H. Yousaf[1,2], K. L. Gilmore[3], B. C. Powell[4], N. L. Vora[3], E. E. Davis[1]; [1]Stanley Manne Children's Res. Inst., Ann & Robert H Lurie Children's Hosp. of Chicago, Chicago, IL, [2]Natl. Inst. for Biotechnology and Genetic Engineering Coll., Pakistan Inst. of Engineering and Applied Sci. (NIBGE-C, PIEAS), Faisalabad, Pakistan., Faisalabad, Pakistan, [3]Dept. of Obstetrics and Gynecology, Div. of Maternal-Fetal Med., The Univ. of North Carolina Sch. of Med., Chapel Hill, NC, [4]Dept. of Genetics, Univ. of North Carolina at Chapel Hill, Chapel Hill, NC

Fetal brain anomalies cause considerable healthcare burden with ~1/1,000 affected newborns every year in the US. The current paradigm for prenatal diagnosis, using karyotype and chromosomal microarray, explains the disease in ~40% of cases. This leaves the majority of the cases undiagnosed, thus complicating counseling, prenatal management, and parental decision-making. To identify underlying etiology of this population, we combined prenatal imaging with prenatal trio exome sequencing (ES) and have achieved a diagnostic rate of ~22%. Here, we present our experience with 39 fetus-parent trios with brain malformations that were undiagnosed as defined by a negative microarray, and no pathogenic or likely pathogenic variants in phenotypically overlapping OMIM genes. In these 39 trios, prenatal imaging revealed cerebellar hypoplasia in 16 (41%) and ventriculomegaly in 15 (38%) cases as the most prevalent findings. Corpus callosum agenesis and absent *cavum septi pellucidi* were observed in 13 (33%) fetuses. Less common manifestations include microcephaly (17%), hydrocephaly, anencephaly and holoprosencephaly in less than 12% of cases. Out of the 39 trios, we found that 14 (35%) have isolated brain malformations while 25 (64%) showed additional extracranial features; skeletal anomalies are the most prevalent finding, occurring in 80% of multisystemic cases. In our cohort, four families have at least two pregnancies with similar or related clinical presentation. Only one out of 39 fetuses are living while the rest did not survive either due to intrauterine fetal demise, elective termination, or neonatal death. ES data underwent unbiased filtering, revealing candidate genes in 22/39 (56%) of the trios leaving 17 families without a plausible molecular candidate gene for the primary phenotype. Candidate genes have been prioritized for in vivo modeling in zebrafish based on identification of other cases from gene matching platforms, presence of a zebrafish ortholog, gene size, and embryonic expression pattern. We are systematically performing gene targeting with CRISPR/Cas9 genome editing and transient suppression to establish relevance to brain anomaly and test pathogenicity of missense variants. Using this paradigm, we have identified *SNAPIN* as a novel cause for fetal brain anomalies and are pursuing additional candidates such as *TXNDC11*, *WWC3*, and *ZNF667*. Together, our genetic analysis paradigm coupled to an experimentally tractable *in vivo* model has the potential to improve diagnostic rates and to expand our understanding of genes critical for fetal brain development.

Session Title: Mendelian Phenotypes Poster Session II

PB4879 Protein family FAM241 (C4orf32 and C10orf35): a potential role in lysosomes.

Authors:

Q. Doctrove, Y. N. Park, G. M. Lenk, M. H. Meisler; Univ. of Michigan, Ann Arbor, MI

Inactivation of *FIG4*, *VAC14*, and other genes that regulate lysosome function, result in enlargement of the lysosome compartment and the appearance of vacuolated cells. In order to identify additional genes that regulate lysosome size and acidity, we carried out a genome-wide screen in HAP1 cells using the GeCKOv2 CRISPR library (Lenk et al., Sci. Rep. 2019). The knockout of FAM241B (C10orf35) resulted in enlarged lysosomes, suggesting a role in lysosomal homeostasis. The two members of this protein family, FAM241 A and B, encode small proteins of 132 and 121 residues with 25% amino acid sequence identity and 53% sequence similarity. There is a predicted transmembrane domain near the C-terminus. Both genes are present in vertebrate genomes, beginning with jawed fish. *FAM241B* is deleted from the bird genome. We confirmed widespread expression of *FAM241* A and B in mouse tissues, at an abundance of approximately 4 and 40 transcripts per million in whole brain, respectively. *FAM241B* encodes two transcripts of comparable abundance that differ by one noncoding exon of 79 bp. Knockout of Fam241 A or B, and the double knockout, did not impair fertility, reduce lifespan, or result in visible abnormalities. RNAseq of brain RNA from double knockout mice detected only minor differences from wildtype mice in gene expression. Analysis of the subcellular distribution of FAM241 A and B is in progress.
Supported by NIH GM24872.

Session Title: Mendelian Phenotypes Poster Session III

PB4880 Proteomic analysis reveals differential protein expression with respect to blood transfusion in Sickle Cell Disease (SCD) patients.

Authors:

A. Kulshrestha, M. E. Garrett, M. W. Foster, M. J. Telen, A. E. Ashley-Koch; Duke Univ. Med. Ctr., Durham, NC

Sickle cell disease (SCD) is characterized by a single mutation in the beta hemoglobin gene. Although patients have the same underlying mutation, the clinical outcomes are highly variable. Blood transfusions are commonly used to ameliorate the symptoms of SCD. To understand the relationship between use of blood transfusion as therapy for SCD and protein expression, we performed a mass spectrometry-based proteomic analysis of plasma samples from 109 adult patients with known transfusion histories from the Outcome Modifying Genes in SCD (OMG-SCD) cohort, quantifying 1190 proteins. We performed logistic regression to identify differentially expressed proteins, incorporating age, sex, and disease genotype as covariates.

Thirteen proteins,
namely *VASN*, *IBP7*, *PI16*, *CADM1*, *FBLN3*, *HEG1*, *APOL1*, *CD44*, *PCOC1*, *CO7*, *LYVE1*, *ATL4*, and *PTGDS*, were differentially expressed in the highly transfused subset of patients (n=37 with > 20 transfusions) after adjusting for False Discovery Rate (FDR), with *APOL1* being expressed at a lower level and the remaining proteins expressed more highly. To exclude false signals arising from other SCD symptoms, we also conducted proteomic analyses on stroke, acute chest syndrome, leg ulcers, and pain. None of the identified proteins were FDR significant in these other analyses. We found that two proteins, *CO7* and *PCOC1*, were also FDR significant proteins identified in proteomic analysis of a composite end-organ damage severity score, suggesting their potential relevance to disease severity. We further investigated the proteomic landscape in a subset of severe SCD cases (SS genotype, n = 91) with respect to transfusion history. In addition to the previously identified 13 proteins, three novel proteins, *NEGR1*, *COMP*, and *CAD13*, were identified.

Our study highlights the utility of plasma proteomics to uncover the complex molecular landscape underlying SCD. The identified proteins, including some previously associated with SCD pathophysiology, as well as others newly identified by this study, offer valuable targets for further investigation and potential development of precision therapeutic interventions. To further elucidate the biological implications, we are exploring the pathways associated with these identified proteins. By unraveling the intricate biological mechanisms and pathways involved in severe SCD, we aim to contribute to the advancement of personalized medicine approaches and provide valuable insights into the molecular profiles specific to severe SCD.

Session Title: Mendelian Phenotypes Poster Session I

PB4881 PTPRM is essential for dendritic arborization in neurons and its loss-of-function underlies neurodevelopmental disorders

Authors:

H-G. Kim[1], N-Y. Lee[2], S-H. Lim[2], V. Gupta[1], A. Ben-Mahmoud[1], D. Martine[3], B. Keren[4], M. Thompson[5], J. Lévy[6], D. Germanaud[7], S. Parikh[8], R. Lebel[9], I. Kushima[10], P. Ellul[11], B. Ku[12], J-R. Lee[2,13]; [1]Qatar BioMed. Res. Inst., Hamad Bin Khalifa Univ., Qatar Fndn., Doha, Qatar, [2]Rare Disease Res. Ctr., Korea Res. Inst. of BioSci. and Biotechnology, Daejeon, Korea, Republic of, [3]CHU de Nantes, service de génétique médicale, Nantes,L'institut du thorax, INSERM, CNRS, UNIV Nantes, CHU de Nantes, CHU de Reims, service de génétique, Reims, Nantes, France, [4]Hôpital Pitié-Salpêtrière, APHP. Sorbonne Université 83 Boulevard de l'hôpital, Paris, France, [5]Hudson Alpha Inst. for Biotechnology, Huntsville, AL, [6]Genetics Dept., AP-HP, Robert-Debré Univ. Hosp., Paris, France, [7]Dept. of Genetics, Ctr. of Excellence InovAND, AP-HP, Robert-Debré Hosp., Paris, France, [8]Neurological Inst., Cleveland Clinic, Cleveland, OH, [9]SUNY Upstate Med Univ, Fayetteville, NY, [10]Dept. of Psychiatry, Nagoya Univ. Graduate Sch. of Med., Med. Genomics Ctr., Nagoya Univ. Hosp., Nagoya, Japan, [11]InovAND Excellence center for autism and neurodevelopmental disorders, Robert Debré Hosp., Paris-Cité Univ., Paris, France, [12]Disease Target Structure Res. Ctr., Korea Res. Inst. of BioSci. and Biotechnology, Daejeon, Korea, Republic of, [13]KRIBB school, Daejeon, Korea, Republic of

Large-scale genomic studies have reported sporadic variants of the PTPRM gene in various neurodevelopmental disorders, including autism spectrum disorder, developmental disorder, epilepsy, and schizophrenia. To substantiate the pathological role of PTPRM, we recruited 10 copy number variants (CNVs) encompassing deletions, duplication, and triplication. These CNVs, ranging in size from 14.2 kb to 1.18 Mb, were observed in 11 individuals with neurodevelopmental disorders, who exhibited a variety of associated features, including mood disorders, ADHD, depression, autism, learning difficulties, and language/speech delays. Genomic alignment of these CNVs identified the smallest overlapping region at 18p11.23, pinpointing PTPRM as the common gene affected in all CNVs (deletions, duplication, and triplication). In addition to this discovery, our exome sequencing analysis identified three missense and one nonsense variants in PTPRM in unrelated individuals with syndromic intellectual disability. The presence of overlapping clinical features among individuals with both single nucleotide variants and CNVs involving PTPRM provides compelling evidence for the contribution of PTPRM haploinsufficiency to the observed clinical presentations in these patients.To validate this hypothesis, we knocked down PTPRM in cultured hippocampal neurons. Remarkably, this manipulation resulted in attenuated dendritic arborization and disrupted synaptic formation, underscoring the critical role of PTPRM in these essential neuronal processes. Additionally, structural modeling of missense variants located within functional domains of PTPRM predicted to induce severe steric hindrance, ultimately destabilizing the protein. Collectively, our findings strongly establish PTPRM as the causative gene underlying neurodevelopmental disorders, emphasizing its pivotal role in the pathogenesis of these conditions. This research shed significant light on the crucial involvement of PTPRM in normal brain development, highlighting its importance in understanding neurodevelopmental processes.

Session Title: Mendelian Phenotypes Poster Session II

PB4882 Puzzling out the genetic bases of hereditary cardiovascular diseases: Application of an integrative approach in a deeply clinically characterised Italian cohort.

Authors:

B. Spedicati[1,2], S. Lenarduzzi[2], M. Dal Ferro[3], A. Paldino[3], D. Mazzà[2], S. D'Errico[1], G. Sinagra[1,3], P. Gasparini[1,2], G. Girotto[1,2]; [1]Dept. of Med., Surgery and Hlth.Sci., Univ. of Trieste, Trieste, Italy, [2]Inst. for Maternal and Child Hlth.–I.R.C.C.S. "Burlo Garofolo", Trieste, Italy, [3]Cardiovascular Dept., Azienda Sanitaria Univ.ria Giuliano Isontina (ASUGI), Trieste, Italy

Background: Hereditary cardiovascular diseases (hCVDs) are a major cause of morbidity and mortality worldwide. They include several nosological entities which may also lead to sudden cardiac death, as cardiomyopathies (CMs), arrhythmic disorders (i.e. channelopathies), and vascular disorders (i.e. aortopathies and pulmonary arterial hypertension). In this light, a large cohort of deeply clinically characterised Italian hCVD patients was selected with the final goal of providing a detailed genetic picture of these disorders in Italy.
Methods: 440 Italian hCVDs patients were referred to the Medical Genetics Unit of the I.R.C.C.S. "Burlo Garofolo" (Trieste, Italy) and nine subjects were recruited through the Regional Register of Sudden Cardiac Deaths of Friuli-Venezia Giulia (Italy). All patients underwent Whole Exome Sequencing (WES) analysis firstly focused on an *in-silico* panel of 66 genes responsible of hCVDs. Multiplex Ligation-dependent Probe Amplification (MLPA) analysis of selected genes was performed in patients affected by hypertrophic (HC) and arrhythmogenic right ventricular cardiomyopathies (ARVC) and negative to WES analysis. Copy Number Variants (CNVs) were confirmed by Nanopore technology (MinION - Mk1C).
Results: 89% of patients were affected by CMs, 7% by channelopathies, and 4% by aortopathies and pulmonary arterial hypertension. WES allowed the molecular characterisation of 23% of patients, with the highest detection rate in familial cases (38% vs 15% in sporadic ones). *TTN*, *MYH7*, *MYBPC3*, *FLNC*, *PKP2* and *DSP* resulted the most commonly mutated genes explaining 75% of cases. Noteworthy findings of this study are: 1) the identification in two patients affected by HC and coming from the Veneto region of the c.913_914delTT, p.(Phe305Profs*27) heterozygous variant in the *MYBPC3* gene (NM_000256.3): a founder effect of this variants in patients from the Veneto region has been described and our finding further supports this hypothesis; 2) the discovery of a novel 3Mb deletion including *DSG2*, *DSC2*, and *TTR* genes firstly hypothesised through WES CNV analysis and confirmed by Nanopore technology in a patient affected by ARVC; 3) the detection in 3/9 sudden death patients of three novel missense variants in the *RYR2* gene, highlighting how the first clinical manifestation of *RYR2*-related arrhythmias may frequently be sudden death.
Conclusions: This study underlines the importance of an integrative approach to provide the correct molecular diagnosis in hCVDs patients, thus emphasising the complexity of the genetic background of these disorders and providing crucial information for patient management and recurrence risk estimation.

Session Title: Mendelian Phenotypes Poster Session III

PB4883 Rare coding variants in the CTSO gene, putatively involved in arterial remodelling, in familial forms of intracranial aneurysm

Authors:

R. Blanchet[1], M. FRENEAU[1], S. Benichi[1], M-A. Mrad[1], S. P. R. Batta[1], M. Rio[1], S. Bonnaud[1], P. Lindenbaum[1], F. Laporte[1], S. Cuenot[2], J. Deleuze[3], C. Dina[1], S. Chatel[1], E. Bourcereau[4], S. Jouan[4], H. Desal[4,1], A-C. Vion[1], R. Bourcier[1,4], G. Loirand[1], R. Redon[1]; [1]Nantes Université, CHU Nantes, CNRS, INSERM, l'institut du thorax, Nantes, France, [2]Nantes Université, CNRS, Inst. des Matériaux Jean Rouxel IMN, Nantes, France, [3]Université Paris-Saclay, CEA, Ctr. Natl. de Recherche en Génomique Humaine (CNRGH), EVRY, France, [4]Dept. of Neuroradiology, CHU Nantes, F-44000 Nantes, France, Nantes, France

Intracranial aneurysm (IA) is a local dilatation of cerebral arteries which preferentially occurs at arterial bifurcations and affects around 3% of the population. The weakening of the artery wall induced by IA can lead to sub-arachnoid haemorrhage, a severe type of stroke. Genetic risk factors have been linked to IA formation, without yet explaining its pathophysiology. Thus, identifying novel susceptibility genes based on the analysis of familial cases may help in elucidating IA pathophysiology.
We recruited a large pedigree with 6 IA carriers, and sequenced the whole exomes of 3 affected relatives. We identified 16 rare functional variants shared between these family members. To reduce the number of variants of interest, we performed identity-by-descent (IBD) analysis from genotyping array data on 5 affected members. Only one variant remained shared between all 5 carriers, which resides in the *CTSO* gene (c.946G>A). Searching for other *CTSO* variants, we analysed the WES data from 93 additional index cases with familial forms of IA. We found one other *CTSO* variant (c.128C>T) in a second large pedigree with 5 IA carriers. The presence of each variant was then confirmed by capillary sequencing in all affected members of each family.
Next, we investigated the function of the papain-like cathepsin encoded by *CTSO*. Its depletion in vascular smooth muscle cells (VSMCs) was associated with increased levels of fibronectin in the extra-cellular space as well as with augmented VSMCs stiffness. The same effect was observed with the expression of the two mutated forms of CTSO identified in familial cases, the protein being poorly secreted in these conditions.
In conclusion, by familial investigations, we identified *CTSO* as a putative novel susceptibility gene for IA, which encodes for a cathepsin acting as an extracellular protease targeting fibronectin. The reduced CTSO secretion observed in the presence of the familial variants may affect the composition and mechanical properties of the extracellular matrix of the cerebral artery wall. This could contribute to the increased risk of developing IA in subjects expressing the *CTSO* variants.

Session Title: Mendelian Phenotypes Poster Session I

PB4884 Rare variants in *PPFIA3* are associated with delayed development, intellectual disability, autistic features, and epilepsy

Authors:

M. Paul[1,2,3], S. Michener[1,2,3], H. Pan[4,2], J. Pfliger[1,2,5], J. Rosenfeld[4], V. C. Lerma[1,2,6], A. Tran[4], M. Longley[1,2], R. A. Lewis[1,4,7], M. Weisz-Hubshman[4], M. Bekheirnia[4,8], N. Bekheirnia[8], L. Massingham[9], M. Zech[10,11], M. Wagner[11,10], H. Engels[12], K. Cremer[12], E. Mangold[12], S. Peters[12], J. Trautmann[12], J. Mester[13], M. Guillen Sacoto[13], R. Person[13], P. McDonnell[14,15], S. Cohen[14], L. Lusk[14], A. Cohen[16], J-B. Le Pichon[17], T. Pastinen[18,16], D. Zhou[19], K. Engleman[20], C. Racine[21,22], L. Faivre[22], S. Moutton[22,23], A-S. Denommé-Pichon[21,22], R. Maroofian[24], H. Koh[25], A. Poduri[25], J. Bolton[25], S. Schuhmann[26], G. Vasileiou[26], S. Russ-Hall[27], I. Scheffer[27,28], G. Carvill[29], H. Mefford[30], Undiagnosed Diseases Network, C. Bacino[4,31], B. Lee[4,32], H-T. Chao[1,2,3,4,33]; [1]Dept. of Pediatrics, Baylor Coll. of Med., Houston, TX, [2]Jan and Dan Duncan Neurological Res. Inst., Texas Children's Hosp., Houston, TX, [3]Cain Pediatric Neurology Res. Fndn. Lab., Jan and Dan Duncan Neurological Res. Inst., Houston, TX, [4]Dept. of Molecular and Human Genetics, Baylor Coll. of Med., Houston, TX, [5]Graduate Program in Dev., Disease Models, and Therapeutics, Baylor Coll. of Med., Houston, TX, [6]Dept. of Psychology, Univ. of Houston, Houston, TX, [7]Dept. of Ophthalmology, Baylor Coll. of Med., Houston, TX, [8]Renal genetics Clinic, Baylor Coll. of Med., Houston, TX, [9]Rhode Island Hosp. and Hasbro Children's Hosp. Providence, Rhode Island, RI, [10]Inst. of Human Genetics, TU Munich, Munich, Germany, [11]Inst. of Neurogenomics, Helmholtz Zentrum München, Munich, Germany, [12]Inst. of Human Genetics, Sch. of Med., Univ. Hosp. Bonn, Univ. of Bonn, Bonn, Germany, [13]GeneDx, Gaithersburg, MD, [14]Epilepsy NeuroGenetics Initiative (ENGIN), Div. of Neurology, Children's Hosp. of Philadelphia, Philadelphia, PA, [15]Dept. of Neurology, Perelman Sch. of Med., Univ. of Pennsylvania, Pa, PA, [16]Children's Mercy Res. Inst., Kansas City, MO, [17]Dept. of Pediatrics, Children's Mercy Kansas City, Kansas City, MO, [18]Children's Mercy Kansas City, Kansas City, MO, [19]Children's Mercy Hosp., Kansas City, MO, [20]Children's Mercy Hosp, Kansas City, MO, [21]Univ. Hosp., Dijon, France INSERM UMR1231 GAD "Génétique des Anomalies du Développement," FHU-TRANSLAD, Univ. of Burgundy, Dijon, France, [22]Functional Unit for Diagnostic Innovation in Rare Diseases, FHU-TRANSLAD, Dijon, France, [23]Dept. of Genetics and Reference Ctr. for Dev. Disorders and Intellectual Disabilities, FHU-TRANSLAD and GIMI Inst., Dijon Bourgogne Univ. Hosp., Dijon, France, [24]Dept. of Neuromuscular Diseases, UCL Queen Square Inst. of Neurology, London, United Kingdom, [25]Dept. of Neurology, Harvard Med. Sch., Boston, MA, [26]Inst. of Human Genetics Univ.sklinikum Erlangen, Friedrich Alexander Univ. Erlangen Nürnberg, Erlangen, Germany, [27]Epilepsy Res. Ctr., Dept. of Med., Univ. of Melbourne, Austin Hlth., Victoria, Australia, [28]Dept. of Pediatrics, Univ. of Melbourne, Royal Children's Hosp., Florey and Murdoch Children's Res. Inst.s, Melbourne, Australia, [29]Northwestern Univ., Chicago, IL, [30]St. Jude Children`s Res. Hosp., Memphis, TN, [31]Texas Children's Hosp., Houston, TX, [32]Dept. of NeuroSci., Baylor Coll. of Med., Houston, TX, [33]McNair Med. Inst., The Robert and Janice McNair Fndn., Houston, TX

PPFIA3 encodes the Protein-Tyrosine Phosphatase, Receptor-Type, F Polypeptide-Interacting Protein Alpha-3 (PPFIA3), which is a member of the LAR protein-tyrosine phosphatase-interacting protein (liprin) family involved in synaptic vesicle transport and presynaptic active zone assembly. The protein structure and function are well conserved in both invertebrates and vertebrates, but human diseases related to PPFIA3 dysfunction are not yet known. Here, we report 16 individuals with rare heterozygous *PPFIA3* variants who have developmental delay, intellectual disability, hypotonia, dysmorphisms including microcephaly or macrocephaly, autistic features, and epilepsy. To determine the pathogenicity of *PPFIA3* variants *in vivo*, we generated transgenic fruit flies expressing either human PPFIA3 wildtype (WT) or variant protein using GAL4-UAS targeted gene expression systems. Ubiquitous expression with Actin-GAL4 showed that the *PPFIA3* variants had variable penetrance of pupal lethality, eclosion defects, and anatomical leg defects. Neuronal expression with elav-GAL4 showed that the *PPFIA3* variants had seizure-like behaviors, motor defects, and bouton loss at the 3rd instar larval neuromuscular junction. Altogether, in the fly overexpression assays, we found that the *PPFIA3* variants in the N-terminal coiled-coil domain exhibited stronger phenotypes compared to those in the C-terminal region. In the loss-of-function fly assay, we show that the homozygous loss of fly Liprin-α leads to embryonic lethality. This lethality is partially rescued by the expression of human PPFIA3 WT, suggesting human PPFIA3 function is partially conserved in the fly. However, the rare *PPFIA3* variants failed to rescue lethality. Altogether, the human and fruit fly data reveal that the rare *PPFIA3* variants are dominant negative loss-of-function alleles that perturb multiple developmental processes and synapse formation.

Session Title: Mendelian Phenotypes Poster Session II

PB4885 RASopathy-linked variants in men with spermatogenic failure

Authors:

A-G. Juchnewitsch[1], K. Lillepea[1], K. Pomm[2], A. Dutta[1], A. Valkna[1], E. Tamp[3], E. Mahyari[4], V. Kübarsepp[1,2], L. Nagirnaja[4], G. Belova[1], O. Poolamets[2], K. Aston[5], D. Conrad[4], L. Kasak[1], M. Punab[1,2], M. Laan[1]; [1]Univ. of Tartu, Tartu linn, Estonia, [2]Tartu Univ. Hosp., Tartu linn, Estonia, [3]East Tallinn Central Hosp., Tallinn, Estonia, [4]Oregon Hlth.& Sci. Univ., Beaverton, OR, [5]Univ of Utah, Salt Lake City, UT

RASopathies are congenital syndromes caused by germline disease-causing variants in 21 genes of the RAS/MAPK pathway[1]. Patients are typically diagnosed in childhood based on phenotypic features, including distinct facial features, affected development and growth, neurologic, cardiac, gastrointestinal, or dermatological conditions. Cryptorchidism (CR) has been observed in ~80% of male cases with Noonan syndrome (NS), the most prevalent RASopathy. We have recently reported a cryptorchid adult patient with spermatogenic failure (SPGF) carrying a likely pathogenic (LP) variant in *SOS1*, the 2nd most frequently affected gene in NS, and proposed the RAS/MAPK pathway as a novel genetic etiology in congenital CR among SPGF patients[2].

We aimed at a systematic analysis of RASopathy-linked variants in idiopathic SPGF patients with (n=151) and w/o CR (n=368), as well as normozoospermic reference subjects (n=323). All men were recruited to the ESTonian ANDrology (ESTAND) cohort and phenotyped at the Andrology Clinic, Tartu University Hospital[3]. The prevalence of CR in Estonian full-term newborn boys is ~1%[4]. Rare disease-causing variants were filtered, interpreted, and prioritized based on known and in silico predicted clinical /functional consequences. The final pathogenicity assessment followed the RASopathy-specific variant interpretation guidelines[5], including gathered extended clinical and family data.

Eight SPGF patients and one subfertile normozoospermic man carried five previously reported and four novel LP/P variants in six RAS/MAPK pathway genes (*PTPN11*, *SOS1*, *SOS2*, *MAP2K1*, *LZTR1*, *SPRED1*). None of these cases were diagnosed with RASopathies before the study. All men presented congenital genitourinary anomalies, including six patients with CR. The prevalence in CR cases (1 in 25) was over 40-fold higher compared to the general population (1 in 1000), whereas no difference was observed for SPGF cases w/o CR and the reference men. RASopathy genes with identified LP/P variants are expressed in adult testis, including spermatogenic and somatic cell types. These data suggest that disease-causing variants in the RAS/MAPK pathway represent a novel identified genetic cause of SPGF and are enriched in patients with CR. Translational research is needed for genetic testing in CR patients to facilitate timely personalized management.

References [1]Tartaglia, Aoki, & Gelb (2022) Am J Med Genet C Semin Med Genet 190:425-439 [2]Laan et al (2021) Clin Endocrinol 94:656-6663 [3]Punab et al (2017) Hum Reprod 32:18-31 [4]Kübarsepp et al (2022) Andrology 10:303-309 [5]Gelb et al (2018) Genet Med 20:1334-1345

Funding: Estonian Research Council, grant PRG1021.

Session Title: Mendelian Phenotypes Poster Session III

PB4886 Reduced *clic4* and *itgb1* expression in a zebrafish crispant model of *crim1* loss of function.

Authors:

T. Le[1], S. Htun[2], J. Lauzon[3], Y. Sun[4,5], E. Ullah[6], R. Hufnagel[6], B. P. Brooks[6], S. Baranzini[4,5], A. Slavotinek[1,2]; [1]Cincinnati Children's Hosp. Med. Ctr., Cincinnati, OH, [2]Div. of Med. Genetics, Dept. Pediatrics, UCSF, San Francisco, CA, [3]Alberta Children's Hosp, Calgary, AB, Canada, [4]Dept. of Neurology, UCSF, San Francisco, CA, [5]Inst. of Human Genetics, UCSF, San Francisco, CA, [6]Natl. Eye Inst., NIH, Bethesda, MD

Heterozygous deletions predicting haploinsufficiency for *CRIM1* have been identified in two families with autosomal dominant, colobomatous macrophthalmia with microcornea and coloboma (MACOM). We present an additional patient with retinal coloboma and a heterozygous deletion involving *CRIM1* as further evidence for the role of *CRIM1* in the etiology of coloboma. We also report 4 members of a family without eye defects, but with hydronephrosis/megaureter that required surgery. All 4 had del(2)(p22.3p22.2) chr2:g.35816791_37319473del that included *CRIM1* and 5 other genes. *Crim1* encodes a type I transmembrane protein that is expressed at the cell membrane of lens epithelial and fiber cells at the stage of lens pit formation. Murine lines with decreased *Crim1* expression have reduced numbers of lens epithelial cells and defective adhesion between lens epithelial cells and between lens epithelial and fiber cells, among other findings. We used CRISPR/Cas9 to make a stable model of *crim1* deficiency, generating zebrafish that were homozygous for a 2 basepair (bp) deletion, c.379_380delCT: p.Leu127Gly*fs**2 (NM_212821.1) and verified reduced *crim1* expression with qRT-PCR. A comparison of eye to head size in homozygous *crim1* crispants and controls showed smaller eyes in the crispants, but the ratio of eye to head size at 48 hpf was significantly increased. This difference was not significant at 72 hpf and we did not observe colobomas. RNA-Seq using whole eyes from *crim1* crispants and controls at 72 hpf showed downregulation of *clic4*, a gene that is expressed at the cell membrane and that is required for integrin-mediated cell adhesion. Prior published work crossing *Crim1* and *Itgb1 floxed* mice showed that 1/4 of the compound heterozygous mice displayed iris coloboma with detachment of the lens epithelial cells from the fiber cells, consistent with an interaction between *Crim1* and *Itgb1*. qRT-PCR in *crim1* crispants and controls showed significantly reduced *clic4* and *itgb1* expression in the *crim1* crispants. As *Itgb1* is expressed in the lens and anchors the anterior epithelium and lens fibers to the lens capsule, we hypothesize that the reduced *clic4* expression in c*rim1* crispants disrupts lens cell adhesion by compromising *itgb1* function and *itgb1*-mediated cell adhesion. Our findings provide data on additional patients with *CRIM1* haploinsufficiency and the underlying pathogenesis of MACOM.

Session Title: Mendelian Phenotypes Poster Session I

PB4887 Reduced genomic diversity in functionally relevant genes modifies the risk of neurodevelopmental phenotypes compared to cancer in individuals with germline *PTEN* variants

Authors:

A. Kim[1,2], L. Yehia[1], C. Eng[1,3,4,5,6]; [1]Genomic Med. Inst., Lerner Res. Inst., Cleveland Clinic, Cleveland, OH, [2]Cleveland Clinic Lerner Coll. of Med., Case Western Reserve Univ., Cleveland, OH, [3]Ctr. for Personalized Genetic Hlth.care, Med. Specialties Inst., Cleveland Clinic, Cleveland, OH, [4]Taussig Cancer Inst., Cleveland Clinic, Cleveland, OH, [5]Dept. of Genetics and Genome Sci., Case Western Reserve Univ. Sch. of Med., Cleveland, OH, [6]Germline High Risk Cancer Focus Group, Case Comprehensive Cancer Ctr., Case Western Reserve Univ., Cleveland, OH

Individuals with germline *PTEN* alterations (PHTS) have a wide spectrum of clinical features ranging from cancer to neurodevelopmental disorders (NDD), including autism spectrum disorder (ASD). However, factors that underlie such phenotypic variability remain elusive, making it impossible to predict clinical outcomes. Here, we investigate germline genomic diversity (locus-homozygosity) as a modulator for the development of cancer versus NDD in PHTS. Genotype data of 376 individuals of European ancestry with PHTS were collected. Study participants included individuals with PHTS and NDD (n=117) and no-NDD (n=259), the former including n=57 with ASD. Homozygosity burden was measured by the ratio of the number of homozygous variants to the total number of genotyped loci. Sets of functionally relevant genes pertinent to NDD were curated for targeted analyses. Conservative filtration of homozygous variants was applied to identify variants in exonic or splicing sites for pathway analysis and with deleterious effects for collapsed variants analysis. Logistic regression models were performed with 10-fold cross validation to predict NDD/ASD phenotype. We found genome-wide level of homozygous common variants (MAF≥0.01) but not rare variants (MAF<0.01) was non-significantly increased in the NDD group. However, targeted analyses revealed significant enrichment of homozygous common variants in genes involved in inflammatory processes and low-confidence genes associated with NDD. Notably, in the ASD subgroup, homozygous common variants were enriched in genes involved in differentiation, inflammatory processes and chromatin structure regulation, high-confidence genes associated with NDD, and regions with pathogenic copy number variants. Pathway enrichment analysis of genes harboring selected homozygous variants revealed pathways germane to NDD/ASD, including neuroinflammation, axonal guidance, and synaptogenesis signaling. Collapsing analysis with more stringent filtration of homozygous variants identified suggestive candidate NDD/ASD genes, including *GABRA4* (p=0.029) and *CX3CR1* (p=0.016). Finally, prediction models for NDD/ASD using homozygosity burdens as predictors, after adjusting for sex, were validated. These models achieved 72% accuracy and 72% area under the curve (AUC) for NDD classification and 89% accuracy and 80% AUC for ASD classification. Thus, a higher burden of homozygous common variants, implying reduced genomic diversity, in genes with biological functions pertinent to NDD/ASD exists, may modify the risk of developing NDD/ASD phenotype, and could serve as a predictive marker of NDD/ASD in patients with PHTS.

Session Title: Mendelian Phenotypes Poster Session III

PB4889 Residual Neuropathy target esterase activity predicts phenotypic onset among *PNPLA6* disorders

Authors:

J. Liu[1], Y. He[2], C. Madu[1], M. Han[1], C. Lwin[1], H. Qian[3], Z. Yong[3], L. Dong[4], G. Arno[5], A. Moore[6], A. Webster[7], M. Michaelides[8], L. Huryn[1], B. Guan[9], R. Hufnagel[10]; [1]Ophthalmic Genetics and Visual Function Branch, Natl. Eye Inst., Bethesda, MD, [2]Fermentation Facility, Biochemistry and Biophysics Ctr., Natl. Heart, Lung, and Blood Inst., Bethesda, MD, [3]Visual Function Core, Natl. Eye Inst., Bethesda, MD, [4]Genetic Engineering Core Facility, Natl. Eye Inst., Bethesda, MD, [5]UCL Inst. of Ophthalmology, London, United Kingdom, [6]Ophthalmology Dept., UCSF Sch. of Medicin, San Francisco, CA, [7]Moorfields Eye Hosp, London, United Kingdom, [8]UCL Inst. of Ophthalmology, Univ. Coll. London, London, United Kingdom, [9]NIH, Bethesda, MD, [10]Natl. Eye Inst., NIH, Bethesda, MD

Pathogenic variants in the *patatin like phospholipase domain containing 6* (*PNPLA6*) gene cause a spectrum of neurological disorders that include spastic paraplegia type 39, Gordon-Holmes, Boucher-Neuhäuser, Laurence-Moon, and Oliver-McFarlane syndromes. Patients display a pleiotropy of clinical characteristics that include vision loss, gait disturbance, anterior hypopituitarism, and hair anomalies. *PNPLA6* encodes Neuropathy target esterase (NTE), an enzyme involved in phospholipid homeostasis in the nervous system. Yet it remains unclear how pathogenic variants in the *PNPLA6* gene affect NTE that ultimately contributes to the observed spectrum of patient phenotypes. Here, we investigated the relationship between NTE activity and clinical manifestations in human and mouse.

Clinical meta-analysis was performed on reported (PubMed, 2008-2023) and newly recruited patients. DNA constructs encoding full-length NTE underwent site directed mutagenesis to produce the patient specific variants. NTE activity was determined as previously published (PMID:12791540). Mouse visual function and structure assays were performed on the Espion E3 system, Optodrum, and Spectralis. Statistical analyses were performed in Prism.

Biallelic pathogenic *PNPLA6* variants were detected in a cohort of 23 patients. Clinical meta-analysis of 118 individuals indicated that missense variants located within the enzymatic domain associate with the severe forms of the disease, revealing a key role NTE activity has on disease pathogenesis. Measuring the esterase activity of 46 disease-associated and 20 benign variants observed across *PNPLA6*-associated clinical diagnoses and general population databases (gnomAD v2.1.1, 2023), respectively, resolved uncertainty in classification of all variants tested. Calculating the overall NTE activity of affected individuals revealed a striking relationship between NTE activity and clinical manifestations, where activity in patients with retinopathy and endocrinopathy was significantly lower to those without visual or endocrinological symptoms. This phenomenon was recaptured in a *Pnpla6* murine allelic series, where mice with similar activities to patients with retinopathy experienced reduced visual function and retinal thickness compared to littermate controls and mice who had similar activities to patients without retinopathy. Overall, our research has discovered not only a novel genotype:activity:phenotype relationship of the *PNPLA6* disorder spectrum predicated on the residual activity of NTE, but also a robust functional assay that can unambiguously classify *PNPLA6* variants.

Session Title: Mendelian Phenotypes Poster Session I

PB4890 Resolving missing heritability among individuals with oculocutaneous albinism: A rare haplotype, comprised of two common *TYR* variants (p.S192Y and p.R402Q), and evidence of OCA individuals with digenic inheritance.

Authors:

S. Loftus[1], M. F. Gillis[1], L. Lundh[1], L. L. Baxter[2], D. E. Watkins-Chow[3], F. X. Donovan[4], NISC Comparative Sequencing Program, W. S. Oetting[5], W. A. Gahl[1], W. J. Pavan[3], D. R. Adams[6]; [1]Med. Genetics Branch, NHGRI, NIH, Bethesda, MD, [2]Ctr. for Precision Hlth.Res., NHGRI, NIH, Bethesda, MD, [3]Computational and Statistical Genomics Branch, NHGRI, NIH, Bethesda, MD, [4]Cancer Genomics Unit, NHGRI, NIH, Bethesda, MD, [5]Dept. of Experimental and Clinical Pharmacology, Univ. of Minnesota, Minneapolis, MN, [6]Office of the Clinical Director, NHGRI, NIH, Bethesda, MD

Oculocutaneous albinism (OCA) is a rare, heterogeneous disorder characterized by reduced pigmentation of the skin, hair and eyes. Critically, individuals also present with visual-developmental changes including foveal hypoplasia and optic nerve misrouting, which contribute to low vision. The majority of OCA individuals possess biallelic variants in *TYR* (OCA types 1A and 1B) or *OCA2* (OCA type 2). Additional genes found to contribute to OCA include *TYRP1* (OCA type 3), *SLC45A2* (OCA type 4), *SLC24A5* (OCA type 6), and *LRMDA/C10ORF11* (OCA type 7), *DCT* (OCA type 8). Unfortunately, not all individuals with OCA are provided with a complete molecular diagnosis after undergoing molecular diagnostic testing. This missing genetic heritability has ranged from 10-50% depending on an individual's residual pigmentation and the country of origin in which these large studies have been performed. We used high-depth custom capture sequencing to evaluate 352 OCA probands. This cohort was enriched for individuals who had undergone previous diagnostic testing and exhibit missing genetic heritability. The largest contributing TYR pathogenic allele was a rare haplotype which arose by recombination and defined by two common variants (p.S192Y and p.R402Q) in cis-orientation (cis-YQ). This rare cis-YQ allele comprised 19.1% (57/298) of OCA type 1 (*TYR*-associated) alleles. These two common alterations are positioned at the terminal ends of diametrically opposing alpha helices involved in copper molecule coordination at the TYR protein active site. Our protein modeling predicts a synergistic impact of the two variants. These results suggest that phasing of common variants p.S192Y and p.R402Q at the *TYR* locus is critical for diagnosis. In addition, we have identified 7 OCA probands who possess deleterious variants in two distinct OCA genes. Each proband harbors a rare predicted deleterious variant in OCA2 and an additional predicted deleterious rare variant in either *TYR* or *TYRP1*. This suggests that a digenic inheritance model for OCA should be considered. Our results highlight that molecular diagnostic screening pipelines need to be reassessed to better capture DNA variation that contributes to OCA. Such variation may occur in patients with other rare disorders that present with missing heritability.

Session Title: Mendelian Phenotypes Poster Session II

PB4891 Responsible and related genes for triplet repeat diseases are involved in the ubiquitin proteasome system

Authors:

H. Hohjoh[1], H. Shimizu[2]; [1]Juntendo Univ., Tokyo, Japan, [2]Natl. Ctr. of Neurology and Psychiatry, Kodaira, Japan

Fragile X syndrome (FXS) and Huntington's disease (HD) belong to the family of triplet repeat diseases caused by an abnormal expansion of trinucleotide repeats in the causative genes, *i.e.*, the *Fragile X mental retardation 1* (*FMR1*) and *Huntingtin (HTT)* genes, respectively. The *FMR1* gene encodes an RNA binding protein involved in the stabilization and translational regulation of messenger RNAs. In mutant *FMR1*, the CGG triplet is abnormally expanded, inhibiting the expression of *FMR1* and causing FXS. The *HTT* gene is thought to have a variety of functions in cells. Disease-causing *HTT* carries abnormal expansion of CAG repeat in exon 1, which is translated into polyglutamine and causes abnormal aggregation of HTT protein, and the aggregated HTT protein is harmful in the brain. Unwanted and defective proteins, including aggregates, are systematically degraded within the cell. Therefore, dysfunction of the intracellular proteolytic system is thought to be related to the accumulation and formation of disease-causing gene products. The ubiquitin-proteasome system (UPS) is the primary ubiquitin-dependent proteolytic pathway. We investigated the UPS in cells, where disease-causing genes and related genes were knocked down by RNA interference (RNAi). Mouse Neuro2a cells, a mouse neuroblastoma cell line, were subjected to gene silencing for *Fmr1*, *Fmr1 autosomal homolog 1 (Fxr1)* and *discs large MAGUK scaffold protein 4* (*Dlg4*) genes, which are associated with FXS, by RNAi, and UPS activity was examined. The results indicated that knockdown of *Fmr1*, *Fxr1* and *Dlg4* genes resulted in increased ubiquitination and that *Fmr1*- and *Fxr1*-knockdown further increased proteasome activity. Knockdown of these genes also affected process formation in N2a cells. Therefore, the findings suggest that these genes are involved in the UPS, and that UPS may contribute to neurite outgrowth. As for the *Htt* gene, *Htt*-knockdown studies are ongoing. We would like to discuss the results, including *Htt* data, at the conference.

Session Title: Mendelian Phenotypes Poster Session III

PB4892 Results from the PROPEL 2 dose-finding study: oral infigratinib leads to significant increases inheight velocity with good tolerability in children with achondroplasia

Authors:

R. Savarirayan[1], J. De Bergua[2], P. Arundel[3], J. Salles[4], V. Saraff[5], B. Delgado[6], A. Leiva-Gea[7], H. McDevitt[8], M. Nicolino[9], M. Rossi[10], M. Salcedo[11], V. Cormier-Daire[12], M. Skae[13], P. Kannu[14], M. Bober[15], J. Phillips[16], H. Saal[17], P. Harmatz[18], C. Burren[19], T. Candler[19], T. Cho[20], E. Muslimova[20], R. Weng[20], S. Raj[1], J. Hoover-Fong[21], M. Irving[22], D. Rogoff[20]; [1]Murdoch Children's Res. Inst., Melbourne, Australia, [2]Hosp. Vithas San José, Vitoria-Gasteiz, Spain, [3]Sheffield Children's NHS Fndn. Trust, Sheffield, United Kingdom, [4]Children Hosp., Toulouse Univ. Hosp., Toulouse, France, [5]Birmingham Women's and Children's NHS Fndn. Truste, Birmingham, United Kingdom, [6]Campus de Teatinos sin número, Hosp. Univ.rio Virgen de la Victoria, Malaga, Spain, [7]Inst. de Investigación Biomédica de Málaga – IBIMA, and Hosp. Univ.rio Virgen de la Victoria, Malaga, Spain, [8]NHS Greater Glasgow and Clyde, Glasgow, United Kingdom, [9]Hôpital Femme Mère Enfant, Hospices Civils de Lyon, Lyon, France, [10]Lyon Univ. Hosp., Bron, Lyon, France, [11]Hosp. Univ.rio La Paz, Madrid, Spain, [12]Necker Enfants Malades Hosp., Paris, France, [13]Manchester Univ. NHS Fndn. Trust, Univ. of Manchester, Manchester, United Kingdom, [14]Dept. of Med. Genetics, Univ. of Alberta, Edmonton, AB, Canada, [15]Nemours Children's Hosp., Wilmington, DE, [16]Vanderbilt Univ. Med. Ctr., Nashville, TN, [17]Cincinnati Children's Hosp. Med. Ctr., Individual, OH, [18]UCSF Benioff Children's Hosp. Oakland, Oakland, CA, [19]Univ. Hosp. Bristol and Weston NHS Fndn. Trust, Bristol, United Kingdom, [20]QED Therapeutics, San Francisco, CA, [21]Johns Hopkins Univ., Baltimore, MD, [22]Guy's and St Thomas' NHS Fndn. Trust, London, London, United Kingdom

Background: Achondroplasia (ACH), the most common short-limbed skeletal dysplasia, is characterized by impaired endochondral ossification resulting from gain-of-function pathogenic variants in the fibroblast growth factor receptor 3 (FGFR3) gene, a negative regulator of endochondral bone growth. People with ACH are at risk for several significant co-morbidities, including brainstem compression due to foramen magnum stenosis, sleep-disordered breathing, chronic otitis media with conductive hearing loss, and symptomatic spinal stenosis. Infigratinib is an oral, selective FGFR1-3 tyrosine kinase inhibitor being investigated for treating children with ACH.

Methods: PROPEL 2 (NCT04265651) is a phase 2 dose-finding, open-label study of infigratinib in children 3−11 years with ACH who participated for ≥6 months in PROPEL (NCT04035811), a non-interventional clinical assessment study. The PROPEL 2 dose-escalation (DE) phase includes 5 ascending dose cohorts from 0.016 to 0.25 mg/kg/day. Primary endpoints: safety; change from baseline in annualized height velocity (AHV); infigratinib pharmacokinetics.

Results: Children enrolled in PROPEL 2 DE completed ≥6 months of treatment at the assigned dose. Cohorts 1-3 (n=37; 0.016, 0.032, 0.064 mg/kg/day) did not show a significant increase in AHV and these doses were assessed as non-efficacious. Treatment at the cohort 4 dose (0.128 mg/kg/day) resulted in an increase in AHV from baseline of 1.52 cm/year in children ≥5 years old (n=11; p=0.02). Infigratinib at the cohort 5 dose (0.25 mg/kg/day, n=10, month 6) resulted in a significant mean increase from baseline of 3.03 cm/year (p=0.0022). In cohort 5, collagen X marker, a biomarker of endochondral ossification, showed a median increase of 28% from baseline at month 6 (n=6). Infigratinib was well tolerated with no serious AEs or AEs leading to study discontinuation, with most AEs mild/moderate in severity. At the cohort 5 dose, no grade 3 AEs or treatment-related AEs were reported.

Conclusion: Oral infigratinib in children with ACH was well tolerated and showed dose-dependent increases in AHV, with a significant mean change from baseline of +3.03 cm/year at the cohort 5 dose (0.25 mg/kg/day). The safety and efficacy of this oral, once-daily dose of infigratinib will be further explored in a phase 3 randomized controlled study. If these phase 2 data are confirmed, infigratinib could potentially offer children with ACH the first effective oral therapy to improve growth, enhance functionality and decrease medical complications.

Session Title: Mendelian Phenotypes Poster Session I

PB4893 Return of individual genetic results in 47,107 individuals with autism.

Authors:

P. Feliciano[1], J. Wright[1], I. Astrovskaya[1], S. Barns[1], A. J. R. Goler[1], J. Hall[1], B. Han[1], T. SPARK Consortium[1], A. Wong[1], N. Volfovsky[1], W. Chung[2]; [1]Simons Fndn., New York, NY, [2]Columbia Univ., New York, NY

SPARK (SPARKforAutism.org)— the largest recontactable cohort of people with autism spectrum disorder (ASD) — has performed whole-exome or genome sequencing and publicly released this data from 47,107 individuals with autism and 67,704 of their first-degree family members. A major goal of the SPARK cohort is to form a partnership with participants and return research results to them. SPARK returns ACMG likely pathogenic (LP) and pathogenic (P) variants from a pre-defined list of 586 genes, 48 recurrent CNVs and chromosomal aneuploidies associated with autism and neurodevelopmental conditions. Of the 47,107 people with autism, we identified 4,057 potentially LP/P variants in 333 unique genes, 193 unique chromosomal band locations and 6 aneuploidies in 3,872 people with autism (8.2%). Among the 3,872 people with putative genetic diagnoses, 13% (n = 520) were previously known, matching the diagnosis provided by the participant. We performed confirmatory testing in a CLIA/CAP-certified clinical lab on 2,013 variants and found that 87% of variants (n = 1,757) were confirmed and interpreted as Likely Pathogenic/Pathogenic (LP/P), 11% (n = 214) were analytically confirmed and interpreted as variants of uncertain significance, one variant was analytically confirmed and interpreted as likely benign, and 2% were not analytically confirmed. Of the 1,757 confirmed LP/P variants, 4% were already known to the participant. Of the remaining 1,682 new genetic diagnoses, 209 occurred in individuals who previously reported having had genetic testing, but no genetic diagnosis was made. The frequency of cognitive impairment was higher in individuals with a history of genetic testing (46%) compared to individuals with no history of previous genetic testing (34%). Among the new single-gene diagnoses, the most common were in: *SHANK3*, *SCN2A*, *NF1*, *PTEN*, and *CHD8*, with each explaining no more than 2% of the total new diagnoses. The most common CNVs in new diagnoses were: 16p11.2 del/dup (3.5%), 2p16.3 del (*NRXN1*) (1.6%), and 15q11 dup (1.2%). As expected, the genetic heterogeneity was high, with 15% of the new diagnoses occurring in just one or two individuals. Overall, our results show that an autism-related genetic result can be identified in at least 8% of individuals ascertained through an autism research cohort. All participants in SPARK—with and without genetic diagnoses—are available for recontact to the entire research community using SPARK's Research Match platform.

Session Title: Mendelian Phenotypes Poster Session II

PB4894 Risk of Alzheimer's disease associated with *ABCA7* tandem repeat expansions across multiple ethnicities.

Authors:

P. Valdmanis[1], E. Kaufman[2], S. Smukowski[1]; [1]Univ of Washington, Seattle, WA, [2]Univ. of Washington, Seattle, WA

Genome-wide association studies (GWAS) are plagued by an inability to locate the relevant functional variant at a given locus. Tandem repeats often have an outsized influence on gene expression and are implicated in over 50 disorders. One such repeat is a 25bp variable number tandem repeat (VNTR) that is present in ATP binding cassette subfamily A member 7 (*ABCA7*). *ABCA7* is a well-established risk factor for Alzheimer's disease (AD). We examined multiple GWAS SNPs associated with AD in *ABCA7* and by integrating long-read and short read data, we surprisingly found that risk alleles tag repeat length in an allele-specific manner. For example, an average of 87 repeat copies are present in homozygous carriers of the rs3764650 risk allele compared to 65 repeat copies in heterozygous individuals and 53 copies in individuals homozygous for the reference allele ($p<0.001$ by one-way ANOVA with multiple testing correction). Interestingly, this effect was specific to samples of European descent as the SNP had no influence on repeat length in samples of African ancestry. Nevertheless, an African-specific sentinel GWAS variant (rs115550680) is adjacent to the 25bp VNTR in its own mini-VNTR with implications for appropriate *ABCA7* splicing. Moreover, in a set of 7,968 cases and 6,792 controls from the Alzheimer's Disease Sequencing Project, we found increased repeat length in samples of both African and European ancestry. Of relevance, increases in *ABCA7* repeat length are associated with aberrant splicing leading to mRNAs with premature termination codons. Expanding our analysis to additional tandem repeats genome-wide revealed VNTRs with significantly expanded repeat length in Black, Hispanic, and non-Hispanic White individuals with AD, while re-enforcing the significant hit at *ABCA7*. Collectively, converging lines of genetic and functional evidence implicate *ABCA7* VNTR length as a risk for AD.

Session Title: Mendelian Phenotypes Poster Session III

PB4895 RNA sequencing for the diagnosis of scoliosis

Authors:

H. Du[1], Q. Li[2], Z. Zhou[1], S. Zhao[3], N. Wu[2]; [1]Peking Union Med. Coll., Beijing, China, [2]Peking Union Med. Coll. Hosp., Beijing, China, [3]Baylor Coll. of Med., Houston, TX

Scoliosis is the lateral curvature of the spine in the coronal plane. A fraction of scoliosis are caused by Mendelian diseases. In a previous study, the overall genetic diagnostic rate of early-onset scoliosis was 20.6%. The genetic causes of most scoliosis patients remain unclear. Recently, transcriptome sequencing combined with whole genome/exome sequencing (WGS/ES) has been increasingly utilized to facilitate the diagnosis of various Mendelian disorders and demonstrated significant gain in diagnostic power. No study has explored the effectiveness of transcriptome sequencing in facilitating the genetic diagnosis of scoliosis. In this study, we aim to explore the utility of skin-derived fibroblast transcriptome sequencing in the genetic diagnosis of scoliosis. We included 181 scoliosis patients who underwent surgical correction at Peking Union Medical College Hospital, including 68 congenital scoliosis (CS) patients, 66 idiopathic scoliosis (IS) patients, and 47 syndromic scoliosis patients. We performed transcriptome sequencing on cultured skin fibroblasts derived from patients' skin tissue. We also performed WGS (160/181)/ES (21/181) on DNA extracted from peripheral blood of each patient. We first analyzed the WGS/ES data from 181 patients and obtained a genetic diagnosis in 41 patients, with a diagnostic rate of 22.7%. After performing a combined analysis of WGS/ES data and transcriptome sequencing data, we further identified the genetic diagnoses in 11 patients, raising the diagnostic rate to 28.7%. The diagnostic rates for CS, IS, and syndromic scoliosis patients were 29.4% (20/68), 15.2% (10/66), and 46.8% (22/47), respectively. Transcriptome sequencing identified the aberrant transcriptional events in 32 patients out of the 52 diagnosed patients. Among the 11 patients who could only be diagnosed with the aid of transcriptome data, 6 had copy number variations, 2 had single-allele gene expression of pathogenic mutations for autosomal recessive inherited diseases, and 3 had low-quality variants that were filtered in WGS/ES-centered analysis. Our study indicates that transcriptome sequencing can aid the interpretation of WGS/ES data and improve the genetic diagnostic rate of scoliosis patients. Our study also highlights the heterogeneity of genetic causes in scoliosis.

Session Title: Mendelian Phenotypes Poster Session I

PB4896 *SGMS2*and Juvenile Open-Angle Glaucoma: A Novel Connection in a Multi-generational Filipino Family

Authors:

E. Collantes[1,2], J. Flores-Anotado[3], S. De Arrigunaga[2], A. Dumdum[3], M. Guevarra[4], M. de Guzman[5], B. Hidalgo[3], V. Guzman[3], M. Rivera-Francia[3], J. Wiggs[2]; [1]Manila Doctors Hosp., Manila, Philippines, [2]Massachusetts Eye and Ear, Boston, MA, [3]Western Visayas Med. Ctr., Iloilo, Philippines, [4]Cardinal Santos Med. Ctr., San Juan, Philippines, [5]Asian Hosp. and Med. Ctr., Muntinlupa, Philippines

SGMS2 (sphingomyelin synthase 2) pathogenic variants cause Calvarial Doughnut Lesions with Bone Fragility (CDL) (MIM 12655). A recurrent nonsense mutation (c.148 C>T, p.Arg50*) has been found in 4 of 6 reported families and interestingly affected members of one of these families also exhibit congenital glaucoma with variable expressivity. We report on a large Filipino family with CDL and juvenile-onset open-angle glaucoma (JOAG), expanding the phenotypic spectrum of *SGMS2*-related disease.

Seventeen members of a four-generation family underwent comprehensive eye exams, skull radiography, and DNA extraction. Whole genome sequencing was performed, using data pre-processing, germline short variant discovery, and joint variant calling pipelines developed by the Broad Institute's Genomics Platform. The seqr platform was used to screen genomic data for mutations in genes associated with early-onset glaucoma. If no such mutations were found, the data was filtered to retain only rare, pathogenic, protein-altering variants. Sanger sequencing was done to validate variant segregation.

Six family members diagnosed with JOAG presented with a mean diagnosis age of 25.5 years, with 83% blind in at least one eye, a mean intraocular pressure of 29.4 mmHg, and glaucomatous optic nerves. No mutations were identified in known disease-causing genes and after variant filtering, a known recurrent heterozygous stop gain mutation, *SGMS2* c. 148C>T (p.Arg50*), was identified in all subjects with JOAG. Ten family members had abnormal skull radiographs, including all 6 JOAG subjects. The *SGMS2* p.Arg50* variant was present in all family members with skull abnormalities.

We report the first Filipino family with CDL associated with the *SGMS2* p.Arg50* variant. Our study identifies a second CDL family with early-onset glaucoma, suggesting *SGMS2* genetic variation may contribute to glaucoma pathogenesis. Considering the role of *SGMS2* in ceramide, sphingomyelin, and diacylglycerol metabolism, it could contribute to trabecular meshwork dysfunction or retinal ganglion cell degeneration in glaucoma. Of interest, *SGMS2* is included in a loci associated with POAG (rs216197, 3.8e-10, beta= -0.06) and a rare variant burden test (LoF, SKAT-O) is also nominally associated with glaucoma in the UK Biobank (p=0.005). Further work is needed to verify a role for *SGMS2* in glaucoma development. This study expands the known genetic landscape of JOAG and the phenotypic spectrum of *SGMS2*-related disease.

Session Title: Mendelian Phenotypes Poster Session II

PB4897 *SHOX* duplication and biochemical findings of short-chain acyl-CoA dehydrogenase deficiency (SCADD) in a patient with Angelman syndrome.

Authors:

Y. Enchautegui-Colon[1], M. Napoli[1], C. Zimmerman[1], H. Pang[2], A. C. Tsai[1]; [1]Univ. of Oklahoma Hlth.Sci. Ctr., Oklahoma City, OK, [2]Univ. of Oklahoma Hlth.Sci. Ctr., Oklahoma City, OK

Angelman syndrome is a condition associated with severe developmental delay. Patients with Angelman syndrome typically have intellectual disability, ataxic gait and severe speech impairment. Patients also have an apparent happy disposition. We present a 3-year-old boy who was initially evaluated at 11 months old due to feeding difficulties, dysphagia, developmental delays and short stature. He has a history of respiratory failure, bronchiolitis and bradycardic arrest. Prenatal exposures to methamphetamines and marijuana were reported. A chromosomal microarray analysis revealed a 92 kb microduplication at Xp22.33 within pseudoautosomal region 1, classified as likely pathogenic. The Xp22.33 microduplication involves the entire *SHOX* gene. Microduplications at the *SHOX* locus have been reported in patients with autism spectrum disorder and neurodevelopmental conditions with low penetrance. Biochemical testing revealed elevated C4 in two separate acylcarnitine profiles and a small peak of ethylmalonic acid in urine organic acids, raising a concern for short-chain acyl-CoA dehydrogenase deficiency (SCADD). His newborn screen was normal. A molecular panel for elevated C4 was obtained, which revealed the variants c.542G>A (p.Gly181Glu) and c.625G>A (p.Gly209Ser) in *ACADS*, accounting for the biochemical findings. The patient's symptoms and developmental delays were felt not to be completely explained by SCADD or by the Xp22.33 microduplication and whole exome sequencing (WES) was performed. WES did not identify any known disease-causing variants to explain the extent of his features. Patient returned to follow-up at 3 years old, he continued to have developmental delays, had no words and was not walking independently. He also had short stature and excessive weight gain. His physical examination was remarkable for happy disposition, hand flapping and frequent smiling. Patient took steps with assistance and was noted to have a wide based gait. The happy disposition and abnormal gait raised concerns for Angelman syndrome. Methylation study for Angelman syndrome/Prader-Willi Syndrome was performed and showed loss of expression of the maternally derived *UBE3A*, confirming the diagnosis of Angelman syndrome. This case is of great significance as it notes the limitations of WES. It also supports that another underlying diagnosis should be considered in patients with biochemical findings of SCADD who have developmental delays and showcases the importance of physical examination in the diagnostic process. The happy disposition and behavioral phenotype that this patient developed later on provided the clues to make the diagnosis of Angelman syndrome.

Session Title: Mendelian Phenotypes Poster Session III

PB4898 Single nucleotide variations and structural variants affecting the *EYS* promoter cause autosomal recessive retinitis pigmentosa.

Authors:

D. Sharon[1], T. Hayman[1], J. Krishnan[2], D. Panneman[3], S. De Bruijn[3], S. Ovadia[4], S. Khateb[1], J. Nellissery[2], M. English[2], R. Ashery-Padan[4], T. Ben-Yosef[5], S. Roosing[3], F. Cremers[3], E. Banin[1], A. Swaroop[2]; [1]Div. of Ophthalmology, Hadassah Med. Ctr., Faculty of Med., The Hebrew Univ. of Jerusalem, Jerusalem, Israel, [2]Neurobiology, Neurodegeneration and Repair Lab., Natl. Eye Inst., NIH, Bethesda, MD, [3]Dept. of Human Genetics, Radboud Univ. Med. Ctr., Nijmegen, Netherlands, [4]Dept. of Human Molecular Genetics and Biochemistry, Sackler Faculty of Med. and Sagol Sch. of NeuroSci.s, Tel Aviv Univ., Tel Aviv, Israel, [5]Ruth and Bruce Rappaport Faculty of Med., Technion-Israel Inst. of Technology, Haifa, Israel

Purpose: Inherited retinal diseases (IRDs) are a group of heterogeneous diseases caused by variants in over 350 genes. Disease-causing variants can be identified in only 65-70% of cases, even when whole exome or whole genome sequencing are being used. Therefore, a large effort is recently directed to the identification of the remaining pathogenic variants, which led to the identification of deep intronic variants that affect splicing and structural variants affecting at least one exon. Untranslated regions (UTRs), on the other hand, are poorly studied, mainly due to lack of knowledge regarding the effect of variants in these regions on gene expression. The purpose of the current study is to characterize the effect of variants, both single nucleotide and structural, identified in the 5' untranslated region of *EYS* in patients with autosomal recessive retinitis pigmentosa (ARRP). **Methods:** DNA samples were screened for variants using smMIPs (targeted panel of 113 genes underlying RP and Leber congenital amaurosis), the Blueprint IRD panel, Sanger sequencing, whole exome sequencing, and whole genome sequencing. Variants were analyzed and annotated using the Franklin platform. Clinical analysis included electroretinography and retinal imaging. Electrophoretic mobility shift assay (EMSA) was performed using nuclear extracts from bovine retina. **Results:** We identified six Arab-Muslim index cases who harbored a 5' UTR variant (c.-453G>T; hg19, chr6:66,417,033C>A) in *EYS* noncoding exon #1. Three were homozygous and three were compound heterozygotes. In addition, we identified four index cases of various ethnicities who harbored biallelic *EYS* variants, at least one of which was a structural variant affecting the 5' UTR. Clinical phenotype of *EYS* RP patients harboring c.-453G>T is generally less severe than those with biallelic null *EYS* variants. EMSA analysis of c.-453G>T revealed a prominent low molecular weight band that is specific to the reference allele and is completely absent from the mutant allele, indicating that this variant affects binding of transcription factors. *In silico* transcription factor binding analysis revealed two transcription factors (YY-1 and RFX), whose binding to the *EYS* promoter is predicted to be affected by the presence of the c.-453G>T variant. **Conclusions:** Our analysis revealed both single nucleotide and structural variants in the *EYS* promoter as the cause of ARRP. We predict that promoter-affecting variants are responsible for some of the missing heritability in inherited disease, and such variants will be identified more frequently as whole genome sequencing becomes the major tool for mutation analysis.

Session Title: Mendelian Phenotypes Poster Session I

PB4899 Single-cell RNA-sequencing analysis reveals the role of endothelial cells in aneurysm pathogenesis in autosomal recessive cutis laxa type 1B syndrome.

Authors:

W. Espinoza Camejo, E. E. Bramel, L. Restrepo, E. G. MacFarlane; Johns Hopkins Univ., Baltimore, MD

Thoracic aortic aneurysms are localized dilation where the diameter of the blood vessel increases by 50% or more. If left untreated, this condition can lead to dissection or rupture. Autosomal Recessive Cutis Laxa type 1B (ARCL1B) is a connective tissue disorder caused by homozygous loss of function mutation in the *EFEMP2/FBLN4* gene. This disorder is associated with an increased risk of aneurysm in the ascending aorta. To better understand the mechanisms that predispose the ascending aorta to medial weakening and dilation, we examined the *Fbln4*$^{E57K/E57K}$ knock-in mouse model, which carries a homozygous mutation known to cause ARCL1B in humans. We conducted a single-cell RNA-sequencing (scRNA-seq) analysis identifying three major vascular smooth muscle cells (VSMC) subclusters in both control and *Fbln4*$^{E57K/E57K}$ aorta. We identified clusters enriched in the aortic root (VSMC3) and ascending aorta (VSMC1 and VSMC2) using transcripts previously shown to describe a gradient of expression along the length of the aorta. VSMC2, and to a lesser extent VSMC1, inherently expressed higher levels of Angiotensin-converting enzyme *(ACE)*, transcripts promoting proteoglycan accumulation, and those associated with a synthetic VSMC phenotype. All these features were exacerbated in the aortas of *Fbln4*$^{E57K/E57K}$ mice, indicating that VSMC2 may be primed and intrinsically more susceptible to disruptions caused by *Fbln4*$^{E57K/E5K7}$ mutation. We also observed increased expression of *Edn1*, the gene responsible for producing *Endothelin 1*, in endothelial cells (ECs) from *Fbln4*$^{E57K/E57K}$ aortas, suggesting that this ligand may drive some of the phenotypic changes observed in adjacent VSMCs. By validating these results, we can deepen our understanding of the intricate paracrine interactions between a specific subset of VSMCs and ECs, thereby elucidating their pivotal role in the heightened susceptibility to aneurysm formation within the ascending aorta of *Fbln4*$^{E57K/E57K}$ mice.

Session Title: Mendelian Phenotypes Poster Session II

PB4900 Single-cell sequencing analysis of heteroplasmy distribution in mitochondrial retinopathy.

Authors:

N. Mullin, A. P. Voigt, M. J. Flamme-Wiese, X. Liu, M. Riker, K. Varzavand, E. M. Stone, B. A. Tucker, R. F. Mullins; Univ. of Iowa, Iowa City, IA

Variants within the high copy number mitochondrial genome (mtDNA) can disrupt organelle function and lead to severe multi-system disease. The wide range of manifestations observed in mitochondrial disease patients may result in part from varying fractions of abnormal mtDNA molecules in different cells and tissues, a phenomenon termed heteroplasmy. However, the landscape of heteroplasmy across cell types within tissues and its influence on phenotype expression in affected patients remains largely unexplored. Here, we identify non-random, cell type-specific distribution of the most common pathogenic mtDNA variant (m.3243A>G) across a complex tissue using mitochondrial single-cell ATAC sequencing (mt-scATACseq), single-cell RNA sequencing (scRNAseq), and long-read mtDNA sequencing. We profile the transcriptome, chromatin accessibility state, and heteroplasmy in retinal cells of a patient with mitochondrial encephalopathy, lactic acidosis, and stroke-like episodes (MELAS) and of several healthy control donors. We found that the proportion of the m.3243G allele was neither evenly nor randomly distributed across diverse cell types. All neuroectoderm-derived neural cell types were near homoplasmic for the mutant variant (median m.3243G proportion of 1.0). However, a subset of mesoderm-derived cells, namely the vasculature of the choroid and T cells, were near homoplasmic for the wildtype allele (median m.3243G proportion beneath 0.05). Gene expression and chromatin accessibility profiles of cell types with high and low proportions of m.3243G implicate mTOR signaling in the cellular response to heteroplasmy. Using mt-scATACseq data, we observed additional heteroplasmic variants segregating in cells with the m.3243G allele. Linkage of these variants was validated using whole-mtDNA long-read sequencing on MELAS patient samples. Finally, by multimodal single-cell sequencing of MELAS retinal pigment epithelial cells we found that a high proportion of the pathogenic m.3243G variant was associated with dysregulation of genes with core roles for retinal metabolic homeostasis. Together, these findings show the non-random and cell type-specific nature of mitochondrial variant partitioning in human mitochondrial disease and underscore its implications for mitochondrial disease pathogenesis and treatment.

Session Title: Mendelian Phenotypes Poster Session III

PB4901 *SMAD4*-related Juvenile Polyposis - Hereditary Haemorrhagic Telangiectasia has a very high risk of gastrointestinal and pulmonary complications.

Authors:

M. Pearson[1], R. McGowan[2], L. Snadden[3], Z. Miedzybrodzka[4], J. Berg[1]; [1]Univ. of Dundee, Dundee, United Kingdom, [2]West of Scotland Ctr. for Genomic Med., Glasgow, United Kingdom, [3]NHS Grampian, Aberdeen, United Kingdom, [4]Univ. of Aberdeen, Aberdeen, United Kingdom

Background: *SMAD4*, the common-SMAD, is integral to mediating signal transduction for serine threonine kinase receptors, including those in the TGF-β, activin and bone morphogenic protein signalling pathways. Heterozygous germline variants that lead to loss of *SMAD4* function result in Juvenile Polyposis-Hereditary Haemorrhagic Telangiectasia Overlap Syndrome (JP-HHT). JP-HHT is associated with potentially life-threatening complications including gastrointestinal cancers, pulmonary arteriovenous malformations (PAVMs) and ascending aortic aneurysms. We investigated the risk of complications in a cohort of patients with JP-HHT in Scotland.
Methods: A retrospective multi-centre chart review of individuals with *SMAD4* loss of function pathogenic variants identified 21 patients from 11 families in three Scottish health boards.
Results: Colonoscopic screening was offered to 90% (19/21) of patients. Colonic polyps were identified in 84% (16/19) at a median age of 24.5 years (IQR 21-33.8). 42% (8/19) of patients required a colectomy at a median age of 24 years (IQR 20-42.5). Colorectal cancer occurred in 19% (4/21) of patients, diagnosed at a median age of 29.5 years (IQR 24.5-38.5). Gastric screening was performed in 71% (15/21) of patients, with gastric polyps present in 67% (10/15). 50% of those with known gastric polyps required a gastrectomy at a median age of 37 years (IQR 12.7-47). 1 patient was diagnosed with gastric cancer, at age 66 years. Epistaxis was investigated in 86% (18/21) of patients, of whom 78% (14/18) reported its presence, with onset universally in childhood. 62% (13/21) of patients were examined for telangiectasia, with its identification in 23% (3/12) of those examined. PAVM screening was performed in 81% (17/21) of patients and a PAVM identified in 47% (8/17) of those screened. Only 19% (4/21) of patients were screened for a thoracic aortic aneurysm, all of whom had negative screening.
Conclusions: All patients with loss of function pathogenic variants in *SMAD4* are at very high risk of life-threatening gastrointestinal disease, which includes very early onset bowel cancer and a high risk of gastric obstruction caused by non-malignant polyposis. The incidence of PAVMs was at least equivalent to that seen in HHT type 1, if not higher. All patients with LOF variants in *SMAD4* require upper and lower gastrointestinal screening in addition to screening and follow up for HHT.

Session Title: Mendelian Phenotypes Poster Session I

PB4902 Snijders Blok Campeau syndrome: Description of 18 additional individuals with variants in *CHD3* and literature review

Authors:

P. Lapunzina[1], P. Pascual[1], J. Tenorio[2], P. Arias[1], N. Gallego-Zazo[1], M. Cazalla[1], C. Silvan[1], M. Palomares[1], F. Ramos[3], J. Nevado[1], B. Almoguera[1], V. Srinivasan[4], V. K. Gowda[4], M. Morleo[5], C. Ayuso[6], V. Nigro[7], S. D'Arrigo[8], C. Ciaccio[8], D. Heron[9], C. Mignot[9], A. Afenjar[9], B. Keren[10], I. Popa[10], C. Martín-Mesa[11], G. Guillen-Fiel[11], A. Serrano[12], M. Dominguez[12], T. Consortium[1]; [1]INGEMM-HULP-CIBERER, Madrid, Spain, [2]INGEMM-HULP, ADNTRO Genetics and BITGENETIC, Madrid, Spain, [3]Univ. de Zaragoza, Zaragoza, Spain, [4]Dept. of Pediatric Neurology, Indira Gandhi Inst. of Child Hlth., Bangalore, India., Bangalore, India, [5]Telethon Inst. of Genetics and Med., Naples, Italy, [6]Inst. de Investigación Sanitaria-Fundación Jiménez Diaz Univ. Hosp., Madrid, Spain- CIBERER, Madrid, Spain, [7]Seconda Univ di Napoli, Napoli, Italy, [8]Dept. of Pediatric NeuroSci.s, Fondazione IRCCS Istituto Neurologico Carlo Besta, Milan, Italy, Milan, Italy, [9]APHP Sorbonne Université, Département de Génétique, Paris, France | Ctr. de Référence Maladies Rares Déficiences Intellectuelles de Causes Rares, Hôpital Trousseau et Groupe Hosp.ier Pitié-Salpêtrière, Paris, France., Paris, France, [10]APHP Sorbonne Université, Département de Génétique, Paris, France, Paris, France, [11]HM Hosp.es, Madrid Spain, Madrid, Spain, [12]Dept. of Med. Genetics, Hosp. Clínico Univ.rio Virgen de la Arrixaca, IMIB-Arrixaca, Murcia, Spain, Murcia, Spain

Snijders Blok-Campeau syndrome (SNIBCPS, OMIM# 618205) is an extremely infrequent disease with only approximately 60 cases reported so far. SNIBCPS belongs to the neurodevelopmental group of disorders. Clinical features of patients with SNIBCPS includes global developmental delay, intellectual disability, speech and language difficulties and behavioral disorders like autism spectrum disorder. In addition, patients with SNIBCPS exhibit typical dysmorphic features including macrocephaly, hypertelorism, sparse eyebrows, broad forehead, prominent nose and pointed chin. The severity of the neurological effects as well as the presence of other features is variable among subjects. SNIBCPS is caused by pathogenic variants in *CHD3 (Chromodomain Helicase DNA Binding Protein 3)* which seems to be involved in chromatin remodeling by deacetylating histones. Here, we report 18 additional patients with clinical features compatible with SNIBCPS from 15 unrelated families with confirmed pathogenic variants in *CHD3*. Patients were analyzed by either Sanger sequencing or by whole exome sequencing. Patients in this study showed different pathogenic variants affecting different functional domains of the protein. Besides, none of the variants described here were reported in control population databases and most computational predictors suggest that they are deleterious. The most common clinical features of the whole cohort of patients are global developmental delay (98%), speech disorder/delay (93%) and intellectual disability (73%). Other frequent features (51-74%) are hypotonia, hypertelorism, abnormality of vision, macrocephaly and prominent forehead, among others. This study expands the number of individuals with confirmed SNIBCPS due to pathogenic or likely pathogenic variants in *CHD3*. Furthermore, we add evidence of the importance of the application of massive paralleled sequencing for neurodevelopmental disorders in which the clinical diagnosis might be a challenge and in which deep phenotyping is extremely useful to accurate manage and follow up the patients.

Session Title: Mendelian Phenotypes Poster Session II

PB4903 Solving challenging titinopathy cases via multi-omics

Authors:

A. Javed, Y. Zhang, L. Xu, Y. Lei, S. H. S. Chan; The Univ. of Hong Kong, Hong Kong, China

Titinopathy, or skeletal muscle dystrophy due to autosomal recessive mutations in Titin (*TTN*) poses a unique diagnostic challenge due to significant variability in disease manifestation and our incomplete knowledge of the role of various domains within the gene. Many healthy individuals harbor rare and computationally predicted to be damaging variants in *TTN*. This confounds interpretation of *TTN* variants in cardiac and neuromuscular disease patients. Here we report six Titinopathy patients from four unrelated families. In each case, a confirmed diagnosis was reached by co-analyzing familial DNA and proband muscle-biopsy based RNA sequencing data together. This allowed direct window into confirming or ruling out the impact of potential deleterious variants. Comparison with in-house and public muscle transcriptomics datasets brought credence to the outlying isoform usage in each family. The phenotypic diversity amongst these patients is noted which highlights the difficulty of diagnosis. The first family included two siblings presented with early onset elbow and ankle contractures followed by adolescent onset mild limb girdle weakness. They had progressive spine rigidity with stable motor performance and elevated Creatine Kinase (CK). The second family is a girl with infancy onset ankle contractures and delayed walking. The third family is a boy with infancy onset elbow and ankle contractures, severe hypotonia and hyper CK in blood. The fourth family included two siblings with decreased fetal movement during pregnancy, and significant gross motor delay and hypotonia with mild facial weakness during early development. Both siblings had normal CK. All patients in this study had normal echocardiogram suggesting lack of cardiac involvement. Our multi-omics findings link individualized specificity of the patient condition to the specific transcriptomic impact of the observed mutations.

Session Title: Mendelian Phenotypes Poster Session III

PB4904 Somatic mutations in hereditary hemorrhagic telangiectasia arteriovenous malformations.

Authors:

E. DeBose-Scarlett[1], C. Gallione[1], C. Friday[2], S. Weinsheimer[3], H. Kim[3], G. Sapisochin Cantis[4], K. Prickett[5,6], J. Roberts[6], M. Faughnan[7,8], D. Marchuk[1]; [1]Molecular Genetics and Microbiol., Duke Univ. Med. Ctr., Durham, NC, [2]Cure HHT, Monkton, MD, [3]Dept. of Anesthesia and Perioperative Care, Univ. of California, San Francisco, San Francisco, CA, [4]Div. of Gen. Surgery, Dept. of Surgery, Univ. of Toronto, Toronto, ON, Canada, [5]Dept. of Otolaryngology, Dept. of Pediatrics, Emory Univ. Sch. of Med., Atlanta, GA, [6]Children's Hlth.care of Atlanta, Atlanta, GA, [7]Toronto HHT Ctr., St. Michael's Hosp. and Li Ka Shing Knowledge Inst., Toronto, ON, Canada, [8]Div. of Respirology, Dept. of Med., Univ. of Toronto, Toronto, ON, Canada

Vascular malformation syndromes, such as hereditary hemorrhagic telangiectasia (HHT), cause major complications including stroke, epilepsy, and hemorrhage. We recently identified biallelic loss of function of HHT genes (*ENG*, *ACVRL1*, or *SMAD4*) in skin telangiectasia, but a possible role for driver mutations in oncogenes in HHT arteriovenous malformation (AVM) pathogenesis has never been shown. Our objective is to identify the genetic mechanism of AVM genesis in brain and visceral HHT lesions. We sequenced DNA from HHT AVMs, including 3 brain (BAVM), 6 pulmonary (PAVM), 19 hepatic (HAVM), and 1 jaw AVM from HHT patients using a targeted panel of 189 genes implicated in vascular malformations on an Illumina platform. After data preprocessing, potential somatic variants were identified using custom software designed to identify somatic variants in tissue without a normal sample, implemented as part of the Gonomics platform. Sequences were then manually screened in IGV for potential variants. We identified potential second-hit somatic mutations in multiple HHT vascular malformation (VM) samples, including mutations in *ENG* or *ACVRL1* in 7 different HHT HAVMs and a *SMAD4* variant in 1 jaw AVM. We identified individual putative somatic mutations in individual HAVMs from individual patients: one patient harboring two distinct variants in *ENG* (c.1687_87-1delinsTT, p.Glu563* and c.174dupC, p.Asn59Glnfs*5), and another patient harboring three distinct variants in *ACVRL1* (c.881T>A, p.Leu294Gln, c.1277A>G, p.Tyr426Cys, and c.1461G>T, p.Lys487Asn), in addition to the germline mutations in the same genes that we also identified. We are currently validating and establishing the phase of these somatic mutations relative to the germline mutations. We are also further investigating potential oncogenic contributors to the pathogenesis of these lesions. We have identified potential new somatic mutations that may contribute to lesion pathogenesis in HHT AVMs. These novel findings further our understanding of the common genetic etiology underlying both visceral AVMs and skin telangiectasias in HHT and open new lines of investigation into previously unrecognized contributors to HHT pathology.

Session Title: Mendelian Phenotypes Poster Session I

PB4905 *SP7*/osterix neomorphic mutation S309W-zebrafish model recapitulates human high bone turnover disease

Authors:

S. Mumm[1], D. Sepich[1], G. Gottesman[1], W. McAlister[1], M. Whyte[2], L. Solnica-Krezel[1]; [1]Washington Univ. Sch. of Med., St. Louis, MO, [2]Washington Univ. Sch. of Med., St. Louis, MO

SP7 encodes the osteoblast-specific transcription factor osterix, which is critical for osteoblast development and function. In 2020, we reported (Whyte et al. Bone 137:115364) a de novo *SP7* missense mutation (c.926C-G, p.S309W) in a young woman with high bone turnover resembling juvenile Paget's disease (JPD). JPD features rapid bone remodeling throughout the skeleton, presents in infancy or early childhood with fractures and deformity, is characterized by elevated serum alkaline phosphatase (ALP), and is usually caused by recessive mutations in *TNFRSF11B* encoding osteoprotegerin (OPG). She began fracturing her lower extremity long bones during infancy, then developed skull deformity and hearing loss. At age 15, radiographs revealed generalized osteosclerosis, hyperostosis, scoliosis, and craniofacial/dental defects, and DXA showed elevated bone mineral density (Z-score +5.1 at lumbar spine and T-score +3.3 at wrist). Bone turnover markers (serum ALP and urine hydroxyproline) were elevated. Iliac crest biopsy was consistent with rapid skeletal remodeling. She responded to alendronate therapy. Sanger sequencing of the OPG gene was negative, whereas whole exome sequencing identified the *SP7* variant S309W. Independent publication of a second patient with the identical mutation and similar clinical characteristics supports the *SP7* neomorphic variant as causal. Herein, we created a zebrafish model of our patient's allele using CRISPR/cas9, called sp7stl838. The homozygous mutants and heterozygous siblings exhibit frequent scoliosis of caudal vertebra. Initial micro computed tomography analysis of homozygous mutant and WT sib showed altered vertebral morphology, teeth not mounted in bone, and fractures in the caudal (tail) fin. Additionally, the mutant fish have altered growth of the premaxilla, maxilla, and dentary and other bones in the head. Cranial sutures between the frontal plates of the neurocranium ridge-up rather than overlap as in WT. The mid-region of the face centered on the nose, appears to have reduced bone density, while scales (which are dermal bone) have increased bone density. Histologic staining for bone and cartilage shows early excessive mineralization due to overactive sp7. Thus, the zebrafish *sp7* allele recapitulates some of our patient's JPD, with additional features. Our findings confirm c.926C-G, p.S309W as the causal mutation for Type 4 JPD and provide a good model for further study of this *SP7* neomorphic mutation.

Session Title: Mendelian Phenotypes Poster Session II

PB4906 Spinal muscular atrophy 5q: molecular, clinical, and functional classification in a cohort of Colombian patients.

Authors:

F. Suarez[1,2], **M. Quiroga**[1], F. Ortiz[1,3], S. Castellar[1,3], E. Ruiz[1,3], M. Salcedo[1,3], D. Soto[1], M. Cediel[1], P. Anzola[1], O. Sanchez[1], C. Jaramillo[4]; [1]Inst. Roosevelt, Bogotá D.C., Colombia, [2]Inst. de Genética Humana. Pontificia Univ. Javeriana, Bogotá D.C., Colombia, [3]Univ. Natl. de Colombia, Bogotá D.C., Colombia, [4]Roche, Bogotá D.C., Colombia

Spinal muscular atrophy (SMA) is a genetic disorder that affects the nerve cells that control voluntary muscle movement. It is a rare disease that affects approximately 1 in 10,000 live births. SMA is caused by a mutation in the survival motor neuron 1 gene (SMN1, 5q13.2), resulting in an SMN protein deficiency. This protein is necessary for the survival and function of motor neurons responsible for sending signals from the spinal cord to the muscles throughout the body. SMN2 can partially compensate for the loss of SMN1, but it produces much less functional SMN protein due to a difference in how the gene is spliced. The number of copies of SMN2 that a person has can influence the severity of SMA and the age of onset. Rare diseases require information on functional phenotyping to understand the natural history of the disease. The molecular, clinical, and motor function description of a cohort of 85 Colombian patients with SMA is presented. The cohort is composed of 41 male patients and 44 female patients with a mean age of 13.6 years (Min: 1, Max: 58). The time to diagnosis had a mean of 8.9 years (Min: <1, Max: 53), SMA type 1: 8.3%; type 2A: 35.7%; type 2B 2.4%; type 3: 51%, and type 4: 2.4%. 18.8% of probands had relatives with SMA. 61% of patients had some comorbidity, 10.6% had fractures, and 41.2% required hospitalization related to SMS comorbidities. 40% of the cases received SMA Modifiers (Nusinersen); 98 of the patients had a homozygous deletion of exon 7 in SMN1. Only one case was compound heterozygous (SMN1: c.[835-?_*3+? del]; [724-2A>G]. 90.6% had Hammersmith scale evaluation; 41% of the cases had Motor Function Measurement (MFM); a 6-minute walk test (6MWT) was present in 34% of the patients. Some other motor function scales were applied to patients according to age and SMS type: Vignos, Brooke, CHOP-Intend, and RULM. More than 80% of the cases had multidisciplinary evaluation: rehabilitation medicine, neurology, genetics, pneumology, and pediatrics. Clinical correlation with SMS type and scale scores are discussed.

Session Title: Mendelian Phenotypes Poster Session III

PB4907 Spinocerebellar ataxia 27B: describing the phenotypic spectrum in an expanding founder French-Canadian cohort

Authors:

C. Ashton[1,2], **F. Villa**[1,3], D. Pellerin[1,4], M. Danzi[5], M-J. Dicaire[1], A-L. Lafontaine[1], M-F. Rioux[6], F. Evoy[6], R. La Piana[1], M. Renaud[7,8], A. Duquette[9,10], S. Zuchner[11], B. Brais[12,10]; [1]McGill Univ., Montreal, QC, Canada, [2]Royal Perth Hosp., Perth, Australia, [3]Montréal Neurological Inst., Montreal, QC, Canada, [4]Univ. Coll. London, London, United Kingdom, [5]Univ. of Miami, Dakota Dunes, SD, [6]Sherbrooke Univ., Sherbrooke, QC, Canada, [7]CHRU de Nancy, Nancy, France, [8]Univ. de Lorraine, Nancy, France, [9]Université de Montréal, Montreal, QC, Canada, [10]Ctr. de Réadaptation Lucie-Bruneau, Montreal, QC, Canada, [11]Univ. of Miami Miller Sch. of Med., Miami, FL, [12]Mc Gill Univ., Montreal, QC, Canada

Introduction: Spinocerebellar ataxia 27B (SCA27B) due to an autosomal dominant intronic GAA repeat expansion in *FGF14* is an increasingly recognised cause of late-onset cerebellar ataxia (LOCA). We performed in-depth clinical phenotyping of an expanding group of French-Canadian cases, including reassessment of the younger generation of three large families to better understand the early clinical manifestations and better define the pathogenic threshold. **Methods**: Clinical records were retrospectively reviewed for all individuals carrying an *FGF14* allele $(GAA)_{\geq 200}$. Individuals with $(GAA)_{200\text{-}250}$ were re-evaluated where possible. Demographics and clinical features, including Scale for the Assessment and Rating of Ataxia (SARA), were reviewed. **Results**: Data of 123 individuals from 52 families were assessed: 94 of 109 individuals with $(GAA)_{\geq 250}$ had sufficient clinical data available - 48 female (51%), median disease duration of 14y (range 1-41). Most (n=73, 78%) reported initial episodic symptoms at median age of 54y (10-87), including dysarthria, diplopia, gait ataxia and vertigo. These could be triggered by alcohol (n=40, 55%), physical exertion (n=36, 49%), as well as other stressors like bright lighting (n=7), emotional stress (n=10), fasting (n=3), caffeine (n=4) and excessive heat (n=3). Thirty-five individuals also suffered from migraine (37%) and three persons reported episodes of speech arrest. Permanent ataxia began at median age of 59y (30-88). Examination features were pancerebellar: axial ataxia (n=80, 85%), appendicular ataxia (n=75, 80%), dysarthria (n=63, 68%). Average SARA scores (n=54) were 8.2 after mean disease duration of 13 years. Only 11 (12%) cases used a wheelchair. Downbeat nystagmus was observed in 45 cases (48%), horizontal nystagmus in 40 (43%), chronic diplopia in 31 (33%). Vibration sensation was impaired in 30 cases (32%) but only seven had EMG proven neuropathy. Brain MRI showed cerebellar atrophy in 31 of 56 cases (55%). Alleles between $(GAA)_{219}$ and $(GAA)_{247}$ were found in 11 individuals: seven with clinical ataxia, allele size $(GAA)_{220\text{-}247}$ (preceded by episodic ataxia in five), one with episodic vertigo and oscillopsia only, $(GAA)_{219}$, and two unaffected on last review (aged 49 and 53, $(GAA)_{224}$ and $(GAA)_{235}$) but with affected family members with pathogenic alleles $(GAA)_{\geq 300}$. **Conclusions**: SCA27B is associated with a prodromal episodic ataxia syndrome, occurring, on average, a decade prior to onset of clinical ataxia. The expansion repeat threshold size for incomplete penetrance of disease will have to be lowered below $(GAA)_{250}$, with familial episodic symptoms associated with expansions as small as $(GAA)_{219}$.

Session Title: Mendelian Phenotypes Poster Session I

PB4908 Splice Mutations and Digital Anomalies Extend the Genotypic and Phenotypic Spectrum of Kim-Gusella Syndrome in PHF21A Patients

Authors:

A. Ben Mahmoud[1], V. Gupta[1], M. Scarano[2], S. Maas[3], M. Alders[4], I. Rouvet[5], E. Bijlsma[6], M. Towne[7], L. Cohen[8], S. Heide[9], B. Keren[9], K. Sørensen[10], A. Jønch[10], T. Hershkovitz[11], O. Josh[12], S-I. Hyun[13], J. Farris[14], A. Abou Tayoun[15], D. Lopergolo[16], D. Niyazov[17], J. Jacober[17], V. McNiven[18], F. Charlotte[18], H-G. Kim[1]; [1]Hamad Bin Khalifa Univ., Doha, Qatar, [2]Cooper Univ. Hosp., Philadelphia, PA, [3]Dept. of Clinical Genetics, Amsterdam, Netherlands, [4]Amsterdam UMC, Amsterdam, Netherlands, [5]Groupe Hosp.ier Est, Hospices Civils de Lyon, Lyon, France, [6]Dept. of Clinical Genetics, Leiden Univ. Med. Ctr., Leiden, Netherlands, [7]Ambry Genetics, California, CA, [8]Weill Cornell Med. Coll., New York, NY, [9]AP-HP Sorbonne Université, Paris, France, [10]Odense Univ. Hosp., Odense, Denmark, [11]Rambam Hlth.Care Campus, Haifa, Israel, [12]Cincinnati Children's Hosp. Med. Ctr., Cincinnati, OH, [13]Gangnam-gu, Seoul, Korea, Republic of, [14]Mayo Clinic, Rochester, MN, [15]Al Jalila Children's Specialty Hosp., Dubai, United Arab Emirates, [16]U.O.C. Neurologia e Malattia Neurometaboliche, Siena, Italy, [17]Ochsner Clinic Fndn., New Orleans, LA, [18]Univ. Hlth.Network & Mount Sinai Hosp., Toronto, ON, Canada

Kim-Gusella syndrome (KGS) is a rare neurodevelopmental disorder caused by heterozygous mutations in the *PHF21A* gene at 11p11.2. In 2012, *PHF21A* was identified as the causative gene for intellectual disability (ID) and craniofacial anomalies (CFA) through breakpoint mapping of balanced translocations and comparative deletion mapping. This discovery explained these two partial phenotypes observed in Potocki-Shaffer syndrome (PSS), a contiguous gene disorder resulting from the minimal 2.1 Mb interstitial deletion of 11p11.2. PSS also exhibits additional skeletal anomalies, including multiple exostoses caused by *EXT2* and parietal foramina caused by *ALX4*. These three genes in the PSS region at 11p11.2 manifest full spectrum of PSS phenotypes. Subsequent to the identification of *PHF21A* by positional cloning, 14 intragenic mutations in *PHF21A* were reported in KGS patients with additional clinical features, including autism, ADHD, and epilepsy. These mutations consisted of 11 frameshift, two nonsense, and one missense alterations. In our present study, we present 14 unrelated KGS patients with novel variants in *PHF21A*, including six frameshift (resulting from four nucleotide deletions, one nucleotide duplication, and one nucleotide insertion), three nonsense, two missense, and three splice mutations. Notably, the identification of splice variants in *PHF21A* is novel and further supports the loss-of-function mechanism associated with KGS. Most KGS patients exhibited developmental delay, intellectual disability, learning disabilities, and language/speech delays. Additionally, several patients displayed digital anomalies such as clinodactyly, syndactyly, and tapering fingers, confirming previously reported phenotypes, and establishing these digital anomalies as novel features associated with KGS. Our findings expand the genotypic and phenotypic spectrum of KGS and enhance our understanding of the role of *PHF21A* in the pathogenesis and potentially improve diagnostic and therapeutic strategies.

Session Title: Mendelian Phenotypes Poster Session II

PB4909 Splicing response QTLs identify putative osteoarthritis risk genes

Authors:

S. Byun, P. R. Coryell, E. Thulson, N. Kramer, S. D'costa, B. O. Diekman, R. Loeser, D. Phanstiel; Univ. of North Carolina at Chapel Hill, Chapel Hill, NC

Genome-wide association studies (GWAS) have identified over 100 genomic loci associated with osteoarthritis (OA) risk; however, mechanistic interpretation of these variants within these loci remains an enormous challenge. Differentially spliced transcripts have been observed in OA tissue, which suggests that some OA risk variants could act by altering splicing patterns; however, this possibility has not been extensively explored. Quantitative trait loci (QTL) mapping is a powerful tool to interpret disease-associated variants, but it must be carried out in the correct cell type and disease-related context. We have previously shown that OA risk variants are enriched in chondrocyte regulatory loci and that treating human chondrocytes with fibronectin fragment (FN-f), a matrikine present in OA tissue, stimulates an OA-like phenotype.

The overall objective of our work is to identify genetic variants that alter OA risk by influencing RNA splicing in either resting or stimulated chondrocytes. We performed RNA-seq in primary human chondrocytes isolated from 102 tissue donors treated with either PBS or FN-f. Using LeafCutter, we identified 422 differentially excised introns (affecting 303 genes) in response to FN-f treatment (FDR $P< 0.05$ and PSI > 0.2). Differentially spliced genes were enriched for many OA-relevant GO terms and KEGG pathways, including actin binding, cytoskeleton organization, EGFR, and Hedgehog signaling. Using QTLtools and condition-separated regression modeling, we identified 1137 sQTLs affecting 987 sGenes (FDR $P<0.05$) and 1073 sQTLs affecting 926 sGenes in PBS and FN-f treated cells, respectively. 54 sQTLs were in linkage disequilibrium ($r2 > 0.5$) with an OA risk variant, including NEK4 and ITIH1, which have been previously implicated in OA. This study highlights the relevance of genetic variants and transcript splicing in mediating the response to cartilage matrix damage and offers a valuable resource for further research and therapeutic development in OA.

Session Title: Mendelian Phenotypes Poster Session III

PB4910 Structural and non-coding variants increase the diagnostic yield of clinical whole genome sequencing for rare diseases

Authors:

J. C. Taylor[1,2], A. T. Pagnamenta[1], C. Camps[1], E. Giacopuzzi[3], J. M. Taylor[4], M. Hashim[1], E. Calpena[5], P. J. Kaisaki[1], A. Hashimoto[6], J. Yu[1], Y. Kesim[1], U. Kini[4], A. Wilkie[5], N. Popitsch[7], OxClinWGS Consortium; [1]Wellcome Ctr. for Human Genetics, Univ. of Oxford, Oxford, United Kingdom, [2]Oxford NIHR BioMed. Res. Ctr., Oxford, United Kingdom, [3]Human Technopole, Milan, United Kingdom, [4]Oxford Univ. Hosp. NHS Fndn. Trust, Oxford, United Kingdom, [5]Weatherall Inst. of Molecular Med., Univ. of Oxford, Oxford, United Kingdom, [6]Chugai, London, United Kingdom, [7]Dept Biochemistry and Cell Biology, Univ. of Vienna, Oxford, United Kingdom

Background
Whole genome sequencing (WGS) is increasingly being used for the diagnosis of patients with rare disease. However, the diagnostic yields of many studies, particularly those conducted in a healthcare setting where patients are referred from a wide range of clinical specialties, are often disappointingly low, at 25-30%. This is in part because although entire genomes are sequenced, analysis is often confined to examining in silico gene panels or coding regions of the genome.
Methods We undertook WGS on a cohort of 122 unrelated rare disease patients and their relatives (300 genomes) who had been pre-screened by gene panel sequencing or high-resolution arrays to exclude known disease genes or chromosomal abnormalities. Patients were recruited from a broad spectrum of clinical specialties. We applied a bioinformatics pipeline that would allow an in-depth interrogation of the genome to provide a comprehensive analysis of all variant types. We combined established bioinformatics tools for phenotypic and genomic analysis with our novel algorithms, SVRare, ALTSPLICE and GREEN-DB, to detect and annotate structural, splice site and non-coding variants.
Results Our diagnostic yield was 43/122 cases (35%), although 47/122 cases (39%) were considered solved when taking novel candidate genes with supporting functional data into account. Structural, splice site and deep intronic variants contributed to 20/47 (43%) of our solved cases. Five genes that are novel, or were novel at time of discovery, were identified (*POLR2A*, *MCM10*, *KMT2E*, *DOCK7*, *SAMD9L*), whilst a further three genes (*DHRS3*, *FOXD3*, *HDLBP*) are putative novel disease genes with evidence of causality. We identified variants of uncertain significance in a further fourteen candidate genes. The phenotypic spectrum associated with *RMND1* variants was expanded to include polymicrogyria. Two patients with secondary findings in *FBN1* and *KCNQ1* were confirmed to have previously unidentified Marfan and long QT syndromes, respectively, and were referred for clinical interventions. Clinical diagnoses were changed, further to WGS, in six patients and treatment adjustments made for eight individuals, which for five patients was considered life-saving.
Conclusions Genome sequencing is increasingly being considered as a first-line genetic test in routine clinical settings and is enabling rapid genetic diagnoses to be provided for many patients. We have demonstrated that structural, splice site and intronic variants make a substantial contribution to diagnostic yield and therefore, that comprehensive analysis of the entire genome is essential to maximise the value of clinical WGS.

Session Title: Mendelian Phenotypes Poster Session I

PB4911 Structural variants are rare sources of molecular diagnosis for secondary findings.

Authors:

M. Walkiewicz[1], E. Karlins[2], M. Sharma[2], S. Byers[1], J. Yan[1], M. Similuk[1], A. Oler[2], D. Veltri[2], R. Ghosh[1], B. Seifert[1], S. M. Holland[1]; [1]NIAID, Bethesda, MD, [2]BCBB/OCICB/NIAID, Bethesda, MD

The contribution of structural variants (SV) to molecular diagnosis of secondary findings (SF) remains largely unaddressed. Several genes (*EPCAM*, *GREM1*, and *HNF1B*) have been excluded due to a technical concern for reliable detection of duplications and deletions which make up a substantial percentage of disease-causing alleles. With the growing utility of genome sequencing in molecular diagnostics and more reliable algorithms to detect SVs, we sought to determine the contribution of SVs to diagnosis of SF.

We called SVs using the GATK-SV pipeline on the Terra platform on 3327 samples with genome sequencing data in subjects recruited for genomic evaluation at the National Institutes of Health. We focused on SVs less than 1% frequency in our cohort and in gnomAD and predicted to overlap coding, UTR, and promoter regions in one of the 78 genes on the SF list: *EPCAM*, *GREM1*, and *HNF1B***.** We utilized ClinGen's dosage sensitivity curation for establishing dose-sensitive gene list. We performed clinical chromosomal microarrays (CMA) for a subset of subjects.

Our analysis identified 18 unique copy number loss calls in 16 genes. Upon further evaluation, 3 calls were technical artifacts due to a very low coverage (*TTN* & *SMAD4*). Of the remaining 15 unique calls (14 genes), 6 occurred in genes with sufficient evidence for haploinsufficiency (*KCNQ1*, *PKP2*, *MAX*, *SMAD3*, *BRCA1*, *VHL*), 2 in genes with little evidence for haploinsufficiency (*ACTA2* & *TNNI3*), 4 in genes with no evidence for dosage sensitivity (*CACNA1S*, *MYH7*, *MYH11*, *TMEM43*0), and 2 in genes with recessive inheritance (*HFE* and *TRDN*; additional variant(s) were not detected). Of note, one copy number loss was detected in *HNF1B*. We also identified 20 unique copy number gain calls, including both whole gene and partial gene duplications, in 16 genes. There is no evidence for triplosensitivity in any of these genes. None of the gains were predicted to result in loss of function except for *TTN*. Interestingly, we identified a possible whole gene gain and a promoter region gain in *GREM1*. Lastly, 17 participants with SV calls by GATK-SV received a CMA: 4 arrays detected the same SV called by our algorithm, 6 arrays were reported as normal, 7 arrays included other findings. Orthogonal confirmations of our findings are pending.

In our cohort, SVs contribute to a small percentage of SF diagnoses. However, with improving algorithms and periodic updates to the SF gene list recommended for reporting the percentage might increase over time. Challenges include orthogonal confirmations as most of these variants were not reported on arrays for our participants, likely due to limited coverage and relatively small size of these variants.

Session Title: Mendelian Phenotypes Poster Session II

PB4912 Study of the cause of death in patients with ATR-X syndrome in Japan

Authors:

T. Wada[1], N. Okamoto[2], K. Kurosawa[3]; [1]Kyoto Univ. Graduate Sch. Med., Kyoto, Japan, [2]Osaka Women s and Children s Hosp., Izumi, Japan, [3]Kanagawa Children s Med Ctr, Yokohama, Japan

BACKGROUND ATR-X syndrome (MIM 301040), caused by loss-of-function variants of *ATRX* gene on X chromosome, is among X-linked intellectual disability syndromes and characterized by male patients, severe intellectual disabilities, alpha-thalassemia, peculiar facies, genital abnormality, skeletal abnormality, and characteristic posture and/or behavior. Since 1995, more than 100 patients have been diagnosed molecular genetically in Japan, and it is estimated that one out of 58,000 to 73,000 of newborn boys has ATR-X syndrome. ATRX gene encode an ATP-dependent chromatin remodeling factors, and ATRX protein binds to guanine-quadruplex, G4, and regulate the expression of genes nearby. (Law MJ. *Cell*, 2010) We reported that 5-aminollevulinic acid as a precursor of protoporphyrin IX and heme, which bind to G4 instead of abnormal ATRX protein and improve expressions of ATRX-targeted genes, could be a treatment for cognitive dysfunction of Atr-x syndrome mice. (Shioda N. *Nature Medicine*, 2018) Now we have been conducting an investigator-initiated clinical trial for ATR-X syndrome patients in Japan since July 2022. Here we report the cause of death of ATR-X patients, because we have few reports about natural history of adult patients or cause of death of ATR-X patients.**PATIENTS and METHOD** 50 ATR-X patients, confirmed molecular genetically, are registered in our ATR-X syndrome database in Japan, including 12 adult patients who are more than 18 years old. We have collected clinical information of these patients from their registered data, clinical records and interviews from their families and/or doctors.**RESULTS** Here we review 10 patients who have passed away. Six died as infants, and the others as adults. In the former group, one died during MRI examination under anesthesia at three years old, one of status epilepticus at one, one of choking on vomit at six, and three of ileus at eight, nine, and 13. In the latter group, one of colon cancer at 61, one of acute renal failure at 46, one of epilepsy or gastro-esophagus regurgitation at 23, and one of myocardial infarction at 37. No cases underwent autopsy. Intriguingly, all patients except two had few complications, and had been living in relatively good health condition. Some infection-like episode triggered their onset, and they got worse rapidly to death.**CONCLUSION** Our data suggests that their health condition can change rapidly after some infectious triggers, even if they seem in relatively good condition as ATR-X syndrome patients, and all adult patients should be examined regularly at least once a year, because some patients were noticed to have renal hypoplasia or heart disease in adults, which were missed in childhood.

Session Title: Mendelian Phenotypes Poster Session III

PB4913 Systematic phenotype and genotype characterization of Moebius syndrome

Authors:

B. Webb[1,2], J. A. Jurgens[3,4,5], N. Narisu[6], Z. Zhang[2], B. J. Barry[3,5,7], W-M. Chan[3,4,7], S. Di Gioia[3,4,5], S. MacKinnon[3,4], C. D. Robson[3,4], D. Oystreck[3], T. A. Hadlock[4,8], D. G. Hunter[3,4], M. Erazo[2], T. Frempong[2], J. C. Rucker[9], T. P. Naidich[2], L. Bonnycastle[6], P. S. Chines[6], A. J. Swift[6], T. Yan[6], C. Van Ryzin[6], F. Facio[6], B. P. Brooks[10], C. Pierpaoli[11], E. J. FitzGibbon[10], Moebius Syndrome Research Consortium, E. W. Jabs[2], F. S. Collins[6], E. C. Engle[3,4,5,7], I. Manoli[6]; [1]Univ. of Wisconsin Sch. of Med. and Publ. Hlth., Madison, WI, [2]Icahn Sch. of Med. at Mount Sinai, New York, NY, [3]Boston Children's Hosp., Boston, MA, [4]Harvard Med. Sch., Boston, MA, [5]Broad Inst. of MIT and Harvard, Cambridge, MA, [6]Natl. Human Genome Res. Inst., NIH, Bethesda, MD, [7]Howard Hughes Med. Inst., Chevy Chase, MD, [8]Massachusetts Eye and Ear Infirmary, Boston, MA, [9]NYU Grossman Sch. of Med., New York, NY, [10]Natl. Eye Inst., NIH, Bethesda, MD, [11]Natl. Inst. of BioMed. Imaging and Bioengineering, NIH, Bethesda, MD

Background: Moebius syndrome (MBS) is defined by congenital, neurogenic nonprogressive facial weakness with limited abduction of one or both eyes with absence of ptosis and vertical gaze limitations, with or without additional manifestations such as lower cranial nerve involvement, limb anomalies, pectoralis major hypoplasia (MBS-Poland), other organ dysmorphogenesis, and neurocognitive impairment. While some MBS cases have been attributed to teratogen exposure or maternal homocystinuria, germline genetic etiologies have been posited for others. These include *de novo* changes in *PLXND1* or *REV3L* which were reported in five probands with MBS-like phenotypes (Tomas-Roca et al., 2015). **Objectives**: The Moebius Syndrome Research Consortium was formed to establish a large research-based MBS cohort for whom we performed systemic phenotypic characterization and germline genetic analyses using exome and genome sequencing (ES, GS). **Results:** 153 probands were ascertained after excluding subjects with overlapping congenital facial weakness syndromes. Subsets received deep phenotyping at the NIH Clinical Center (n = 32) and germline ES or GS (n = 70). All probands were sporadic/simplex cases with no known horizontal or vertical transmission, with 17 unaffected offspring and 178 unaffected full siblings across the cohort. Common co-occurring phenotypes included clubfoot (43.1%); intellectual disability (37.7%); major limb anomalies (30.7%); hearing loss (29.6%); autism (17.9%); MBS-Poland (12.4%); and seizures (11.9%). MR imaging revealed hypoplasia or absence of the abducens and facial cranial nerves. Filtering for rare *de novo* or autosomal recessive SNVs, indels, and SVs yielded 180 variants in 113 genes. While none were definitively causal and no convincing recurrent variants were identified, we prioritized 7 candidate genes with *de novo* variants (*USP15, POU2F1, CNTRL, MINDY1, KPNA3, STMN3, PALM*) and 5 candidate genes with biallelic variants (*ZNF407, KCNAB2, ZRANB1, MORC2,* and *PBXIP1).* Putative matches in GeneMatcher were made for 11 candidate genes, but none had MBS phenotypes. We did not identify pathogenic variants in *PLXND1* or *REV3L*, nor a substantial contribution of variants in maternal thrombophilia genes. **Conclusions:** MBS is phenotypically heterogeneous. We did not identify a strong or unifying germline genetic etiology. Alternative genetic etiologies to be explored in future studies include noncoding or somatic variants, as well as oligogenic, polygenic, epigenetic, or complex modes of inheritance. These, alone or in combination with environmental factors, could alter brainstem and organ embryogenesis.

Session Title: Mendelian Phenotypes Poster Session I

PB4914 The genetic architecture and discovery of novel candidate causative genes of visual impairment in consanguineous families from Iran and Pakistan.

Authors:

I. Anwar[1], S. Oreshkov[1], A. Zafar[2], H. Frederiksen[1], M. Choung[1], A. Rashid[1], M. Ullah[3,4], T. Millo[5], L. Zelinger[5], A. Munir[6], A. Umer[6], S. Afsar[6], A. Murtaza[7], K. Ashfaq[7], F. Batool[7], S. Khakwani[7], S. Javed[8], J. Raja[2], A. Arshad[2], H. Ahmadieh[9], H. Sabbaghi[10], L. Javanparast[9], A. Moghaddasi[9], S. Zulfiqar[7], H. Kausar[7], F. Suri[9], A. Rehman[6], D. Sharon[5], S. Antonarakis[11], M. Ansar[1,12]; [1]Dept. of Ophthalmology, Univ. of Lausanne, Jules Gonin Eye Hosp., Fondation Asile Des Aveugles, Lausanne, Switzerland, [2]Dept. of Biology, PMAS-Arid Agriculture Univ., Shamsabad, Rawalpindi, Pakistan, [3]Inst. of Molecular and Clinical Ophthalmology Basel, Basel, Switzerland, [4]Dept. of Ophthalmology, Univ. of Basel, Basel, Switzerland, [5]Ophthalmology, Hadassah Hebrew Univ. Med. Ctr., Jerusalem, Israel, [6]Dept. of Zoology, Faculty of Biological and Hlth.Sci., Hazara Univ., Mansehra, Pakistan, [7]Dept. of Biotechnology, Kinnaird Coll. for Women, Lahore, Pakistan, [8]Dept. of Biotechnology, Lahore Coll. For Women Univ., Lahore, Pakistan, [9]Ophthalmic Res. Ctr., Res. Inst. for Ophthalmology and Vision Sci., Shahid Beheshti Univ. of Med. Sci., Tehran, Iran, Islamic Republic of, [10]Ophthalmic Epidemiology Res. Ctr., Res. Inst. for Ophthalmology and Vision Sci., Shahid Beheshti Univ. of Med. Sci., Tehran, Iran, Islamic Republic of, [11]Dept. of Genetic Med., Univ. of Geneva Med. Faculty, Ctr. Médicale Univ.ire, Geneva, Switzerland, [12]Advanced Molecular Genetics and Genomics Disease Res. and Treatment Ctr., Dow Univ. of Hlth.Sci., Karachi, Pakistan

Consanguinity facilitates the discovery of novel genes for autosomal recessive disorders because of the extent of genomic homozygosity. In the children of consanguineous parents, the average homozygosity extends to approximately 253 Mb, which is ten-fold higher than in outbred individuals. Iran and Pakistan, are countries where consanguineous marriages account for up to 39% and 70%, respectively. We have studied 218 consanguineous families (817 patients) from Iran and Pakistan, mostly having more than one affected individual with visual impairment. The most common phenotype was Retinitis pigmentosa, followed by Leber congenital amaurosis and Stargardt disease. Whole exome sequencing was performed on affected individuals and genotyping of potential variants was carried out in all family members (parents, all affected and unaffected siblings). The overall diagnostic yield was 72% (156 families). The most frequent causative genes were *CRB1*, *ABCA4* and *AIPL1* identified in 15, 11 and 9 families, respectively. This includes five families with pathogenic mutations in *RPE65* that can be treated, thanks to the advancements in gene therapy. In the 62 undiagnosed families, we identified several novel candidate genes, including *ACACB*, *NOC4L*, *LIX1* and *MOB1A*. In the case of *ACACB* gene, we identified a homozygous splice variant (c.2295+1G>A) in three affected individuals of a consanguineous family from Pakistan. Through gene matching, we identified another case in which a homozygous missense mutation (c.4469G>A, p.Arg1490His) was observed in five affected individuals of an Iranian Jewish family. The phenotype of all the eight patients was a similar form of Retinitis pigmentosa (RP). In the *NOC4L* gene, we identified homozygosity for a splice variant (c.1235-2A>G) in two individuals with RP. Mouse model for this gene have shown optic disc atrophy. Homozygous missense mutations (c.348C>A, p.Phe116Leu) in *LIX1* in three affected individuals and (c.503C>G, p.Ser168Cys) in *MOB1A* in two affected individuals were identified in two Pakistani consanguineous families. Both genes have been shown to regulate *YAP1* activity which has previously been linked with visual impairment. Additional candidate variants/genes will be presented at the conference. The identification of novel genes for autosomal recessive genetic disorders is accelerated by studying large consanguineous cohorts, this improves the molecular diagnosis and genetic counseling for the affected families, and provides opportunities for precision treatments.

Session Title: Mendelian Phenotypes Poster Session II

PB4915 The genomic landscape of malformations of cortical development across 431 surgically accessible epileptogenic human brain lesions.

Authors:

C. Bosselmann[1,2], C. Leu[1,2,3], L. Hoffmann[4], S. Baldassari[5], R. Coras[4], K. Kobow[4], H. Hamer[6], S. Brandner[7], K. Rössler[7,8], C. Bien[9], T. Kalbhenn[9,10], T. Pieper[11], T. Hartlieb[11], K. Becker[12], J. Altmüller[12,13,14], L. Ferguson[15], R. Busch[2,1], S. Baulac[16], P. Nürnberg[12,17], I. Najm[2], I. Blümcke[4,2], D. Lal[1,2,12,18]; [1]Genomic Med. Inst., Lerner Res. Inst., Cleveland Clinic, Cleveland, OH, [2]Epilepsy Ctr., Neurological Inst., Cleveland Clinic, Cleveland, OH, [3]Dept. of Clinical and Experimental Epilepsy, Inst. of Neurology, Univ. Coll. London, London, United Kingdom, [4]Dept. of Neuropathology, Partner of the European Reference Network (ERN) EpiCARE, Univ.sklinikum Erlangen, FAU Erlangen-Nürnberg, Erlangen, Germany, [5]Brain and Spine Cord Inst. (ICM) - INSERM, Paris, France, [6]Epilepsy Ctr., EpiCARE Partner, Univ.sklinikum Erlangen, FAU Erlangen-Nürnberg, Erlangen, Germany, [7]Dept. of Neurosurgery, EpiCARE Partner, Univ.sklinikum Erlangen, FAU Erlangen-Nürnberg, Erlangen, Germany, [8]Dept. of Neurosurgery, EpiCARE Partner, Med. Univ. of Vienna, Vienna Gen. Hosp., Vienna, Austria, [9]Dept. of Epileptology (Krankenhaus Mara), Med. Sch., Bielefeld Univ., Bielefeld, Germany, [10]Dept. of Neurosurgery (Evangelisches Klinikum Bethel), Med. Sch., Bielefeld Univ., Bielefeld, Germany, [11]Ctr. for Pediatric Neurology, Neurorehabilitation, and Epileptology, Schoen-Clinic, Vogtareuth, Rosenheim, Germany, [12]Cologne Ctr. for Genomics (CCG), Univ. of Cologne, Faculty of Med. and Univ. Hosp. Cologne, Cologne, Germany, [13]Berlin Inst. of Hlth.at Charité, Univ.smedizin Berlin, Core Facility Genomics, Berlin, Germany, [14]Max Delbrück Ctr. for Molecular Med. in the Helmholtz Association (MDC), Berlin, Germany, [15]Epilepsy Ctr., Neurological Inst., Cleveland Clinic,, Cleveland, OH, [16]ICM, Paris, France, [17]Ctr. for Molecular Med. Cologne (CMMC), Univ. of Cologne, Faculty of Med. and Univ. Hosp. Cologne, Cologne, Germany, [18]Stanley Ctr. for Psychiatric Res., Broad Inst. of Harvard and M.I.T., Cambridge, MA

Rationale: Understanding the molecular mechanisms involved in the etiology of epileptogenic malformations of cortical development (MCD) is essential to improve the treatment of drug-resistant focal epilepsy. Methods: We aggregated 431 MCD across multiple centers in Europe and the Cleveland Clinic, USA. 224 samples were previously deep (>350x) exome sequenced and 148 samples were deep (>1500x) panel sequenced for 122 candidate genes. Results were pooled with 59 external deep WES samples. All pathology samples were independently reviewed and classified according to the 2022 ILAE consensus classification of focal cortical dysplasia. We tested for somatic variant enrichment with dNdScv, a model that provides quantitative estimates for driver mutations. We used stringent Bonferroni correction after the number of genes captured by the whole-exome screens in each pathology to identify genome-wide significant somatic variant-enriched genes. Gene list enrichment analysis was done with Enrichr. Results: We observed that 122/431 (28.3%) samples had >=1 variant in known causal genes for MCD, and a further 35/431 (8.1%) samples had validated functional cancer driver variants from the Catalogue Of Somatic Mutations In Cancer (COSMIC v98). MTOR was the most commonly affected gene (N=34 samples). Interestingly, multiple variants in shared pathways (e.g., MTOR, TSC2, DEPDC5) co-occurred in 15 samples. Somatic variant enrichment analysis confirmed two genes previously reported to be enriched with somatic variants in individuals with MCD: MTOR in 26.4% of individuals with focal-cortical dysplasia type 2 (FCD II; $P=6.45x10^{-8}$) and SLC35A2 in 30.6% of individuals with mild malformation of cortical development with oligodendroglial hyperplasia in epilepsy (MOGHE; $P=2.22x10^{-9}$). Pathway analyses across all nominal enrichments showed an overrepresentation of the PI3K-Akt signaling (KEGG; $Padj=7.49x10^{-7}$) and autophagy pathways (GO:0016242; $Padj=1.12x10^{-6}$) in FCD II. Conclusions: Few genes contribute to large proportions of specific pathology-defined epilepsies in somatic variant enrichment analyses of MCD. Our study suggests that the remaining unexplained lesional epilepsies are highly heterogeneous. We hypothesize that second-hit somatic variants may potentially occur at pathway-level, pending further confirmation.

Session Title: Mendelian Phenotypes Poster Session III

PB4916 The p.C759F variant in USH2A is a pathogenic mutation: systematic literature review and meta-analysis of 667 genotypes

Authors:

J. Han[1,2], F. Cancellieri[1,2], I. Perea-Romero[3,4], M. Quinodoz[1,2,5], C. Ayuso[3,4], C. Rivolta[1,2,5]; [1]Inst. of Molecular and Clinical Ophthalmology Basel, Basel, Switzerland, [2]Univ. of Basel, Basel, Switzerland, [3]Univ. Autónoma de Madrid, Madrid, Spain, [4]Inst. de Salud Carlos III, Madrid, Spain, [5]Univ. of Leicester, Leicester, United Kingdom

Background: The p.C759F variant in the Usherin gene (USH2A) has been reported several times in the literature as the molecular cause of recessive retinitis pigmentosa, a form of retinal degeneration. However, its pathogenicity has been questioned once by the publication of two discordant genotypes from a single family. Methods: To elucidate its role as a disease-causing mutation, we performed a meta-analysis on 667 individuals carrying this variant, according to data we collected by reviewing 437 research articles published in the last 22 years. Results: We performed three independent statistical tests and ascertained with a very high degree of confidence that p.C759F is: i) enriched ~11x in patients with respect to healthy individuals (Chi-square = 718, *p*-value = 3.45 x 10-158); ii) enriched ~45x in patients compared to healthy individuals when in trans with a pathogenic mutation in *USH2A*, indicating it is a recessive mutation (Chi-square = 15,681, *p*-value < 5.0 x 10-324); iii) enriched ~2400x in patients with respect to healthy individuals when in homozygosis (Chi-square = 47149, *p*-value < 5 x 10-324). Conclusion: In summary, our results unambiguously confirm that p.C759F is a Mendelian recessive mutation, leading to retinal blindness. Grant References: This work was supported by the Swiss National Science Foundation (Grants # 176097 and 204285).

Session Title: Mendelian Phenotypes Poster Session I

PB4917 The phenotypic spectrum of the Cornelia de Lange-like “Alazami-Yuan Syndrome”: A case report of the 7th diagnosed patient and review of the literature.

Authors:

A. Pappas[1], K. Kohnen[1], M. Mooney[2], J. Owens[3], W. Zhang[4], A. Shillington[3]; [1]Univ. of Cincinnati, Cincinnati, OH, [2]Univ. of Cincinnati Coll. of Med., Cincinnati, OH, [3]Cincinnati Children's Hosp., Cincinnati, OH, [4]Cincinnati Children's Hosp. Med. Ctr., Cincinnati, OH

Alazami-Yuan Syndrome is an autosomal recessive Cornelia de Lange-like syndrome caused by TAF6 loss of function variants. TAF6 encodes a TATA-binding protein associated factor that mediates interactions in early steps of gene transcription. Cornelia de Lange syndrome (CdLS) and Alazami-Yuan Syndrome have significant phenotypic overlap, though no Alazami-Yuan syndrome patient has had enough symptoms to qualify for a clinical diagnosis of classic CdLS. We present a 17-year-old female who had a clinical diagnosis of CdLS in early childhood, but genetic testing later revealed compound heterozygous variants in TAF6 consistent with Alazami-Yuan syndrome.

In addition to CdLS-like features including thick arched eyebrows, long eyelashes, widely spaced teeth, sensorineural hearing loss, growth restriction, and microcephaly, the patient also presented with features not typically seen in CdLS, including premature adrenarche, hair loss, palindromic rheumatism, and a transaminitis. A developmental gene panel revealed a paternally inherited likely pathogenic c.323T>C (p.Ile108Thr) variant and a maternally inherited likely pathogenic c.1974dup (p.Thr659HisfsTer?) variant in TAF6.

This patient represents the 8th reported case of Alazami-Yuan syndrome. She further supports phenotypic overlap between Alazami-Yuan syndrome and CdLS while maintaining that these are two clinically distinct conditions. According to recently developed CdLS clinical scoring systems, all Alazami-Yuan syndrome patients have enough symptoms to merit CdLS testing, but none meet enough criteria for a clinical diagnosis of classic CdLS. This is potentially related to overlapping transcription disruption in both conditions. TAF6 plays a significant role in transcription through stabilization of the TFIID complex, while CdLS’s disruption of cohesins also leads to disrupted transcription. However, cohesion disruption also negatively impacts additional cellular processes such as DNA repair and gene expression. As other CdLS-like genes also have direct roles in gene transcription, we suspect that the “transcriptomopathies” lead to a similar, but more restricted, phenotype than the classical CdLS “cohesinopathies.”

Session Title: Mendelian Phenotypes Poster Session II

PB4918 The use of Chinese version of hyperphagia questionnaire to assess the hyperphagia status of Prader Willi syndrome patients in Taiwan.

Authors:

L-P. Tsai[1], G. Liu[2], S-P. Lin[3], P-C. Chang[4], W-S. Lai[2], P-S. Tsai[4]; [1]Dept. of Pediatrics and Med. Genetic Canter, Taipei Tzu Chi Hosp., Taipei, Taiwan, [2]Dept. of Psychology, Natl. Taiwan Univ., Taipei, Taiwan, [3]Dept. of Pediatrics and Med. Res., MacKay Mem. Hosp, Taipei, Taiwan, [4]Sch. of Nursing, Coll. of Nursing, Taipei Med. Univ., Taipei, Taiwan

Background and Purpose The hyperphagia behavior is one of the core symptoms of Prader-Willi syndrome(PWS) patients. Recently, there are many clinical trials aiming the improvement of satiety disorder for PWS. All these clinical trials using a modified hyperphagia questionnaire(HQ) to evaluate the improvement of hyperphagia condition after treatment. In order to understand the hyperphagia status of Taiwan PWS patients and prepare for the coming of satiety treatment, we started the HQ assessment in Taiwan. **Method** Basing on the HQ designed by Elisabeth M. Dykens et al. at year 2007, we translated it into Chinese in the way of back-translation which was approved by a translation committee. Then we enrolled 84 PWS main caregivers with patient age above 4 years for HQ assessment and compared our result with Dykens et al. report to evaluate the difference between these 2 groups. Thirty three caregivers received second assessment 1.5 to 2 month later for reliability evaluation of this questionnaire. **Result** Overall, the Chinese version of HQ showed acceptable test-retest reliability over a 1.5- to 2-month interval with a coefficient of 0.827 for all subjects. In the three dimensions of HQ, behavior, drive and severity, the average scores are 11.85, 10.32 and 4.19, respectively. Compared to the Dykens et al. report with 13.57, 12.29 and 4.61 on the corresponding dimension, our score were all lower with statistic significance. For the subgroup analysis, the study number for patient age 4-10 years, 11-19 years, 20-29 years and above 30 years is 26, 34, 19 and 5, respectively. The score for behavior dimension was higher with age in the trend of 9.96, 11.88, 13.95 and 15.4. The score for hyperphagia drive was also increasing with age in the way from 9.70 to 9.90 to 11.27 and 13.2, which was consistent with clinical observation .However, on the severity dimension, the score from 3.57 to 4.04 to 5.42 and 5.2, revealed lower for patients above 30 years of age. Since only 5 patients more than 30 years in our study, we have to enroll more patients in this age group to confirm the presence of declining severity for older PWS patients. **Conclusion** Our data suggested that this back-translation Chinese version HQ had a good consistency reliability and was available for the clinical evaluation of PWS hyperphagia status. And, the hyperphagia condition in Taiwan PWS patients seemed less severe than Dykens report 10 years ago. Further enrolling the older patients more than 30 years of age is necessary to make sure the idea of presence of declining satiety for older PWS patients.

Session Title: Mendelian Phenotypes Poster Session III

PB4919 The usefulness of molecular diagnosis to identify a patient with atypical MPS IVA phenotype suggesting Legg-Calve-Perthes disease and skeletal dysplasia.

Authors:

N. Davalos[1,2], A. R. Rincón-Sánchez[3], S. A. Ramírez-García[4], V. C. Ledezma-Rodriguez[5], C. Carbajal-Villaseñor[6], R. Corona-Rivera[7], M. A. Alcantara-Ortigoza[8], D. García-Cruz[1]; [1]Inst. of Human Genetics, Dept. of Biol. Mol and Genomics CUCS UdeG, Guadalajara, Jalisco, Mexico, [2]Genetics, Regional Hosp. "Dr. Valentin Gomez Farías" ISSSTE., Zapopán, Jalisco, Mexico, [3]IBMMTG, Dept. of Biol. Mol and Genomics CUCS, UdeG, Guadalajara, Jalisco, Mexico, [4]MePiel Dermatological Ctr., Guadalajara, Jalisco, Mexico, [5]Dept. of Plastic and Reconstructive Surgery: Hosp. de Especialidades CMNO, IMSS., Guadalajara, Jalisco, Mexico, [6]Pediatric Orthopedics HR "Dr. Valentin Gomez Farías" ISSSTE., Zapopán, Jalisco, Mexico, [7]Genetics, Civil Hosp. of Guadalajara "Juan I Menchaca"., Guadalajara, Jalisco, Mexico, [8]Molecular Biology Lab.. Natl. Inst. of Pediatrics. DNA-GEN SC,, Ciudad de México, Mexico

Mucopolysaccharidosis type IVA (MPS IVA or Morquio Syndrome, MIM #253000) is caused by biallelic pathogenic variants in *GALNS* gene (16q24.3) and characterized by an important skeletal and articular involvement. **Case Report**: The propositus, the 2nd child of non-consanguineous parents, was referred to our institution at 7 years of age, because of frequent falls, abnormal gait, hip, and feet pain, but without intellectual disability. His Apgar score and anthropometric parameters at birth were normal. At 11 years of age, was referred for genetic evaluation by multiple epiphyseal dysplasia and Legg-Calvé-Perthes disease with a height of 138 cm (40th-50th percentile), weight of 30 kg (20th-30th percentile), and head circumference of 54 cm (80th-90th percentile). No facial dysmorphism was noted. A gene panel for skeletal dysplasias (Mendelics, Brazil), revealed a homozygous genotype for the pathogenic *GALNS* variant NM_000512.5(GALNS): c.181C>T or p.(Arg61Trp) [rs145798311], which is associated to attenuated forms of MPS IVA. Later, whole *GALNS* gene sequencing performed in the healthy 15-year-old male brother, and the probably affected 4-year-old female sister (DNA-GEN SC, México), revealed a normal and heterozygous p.(Arg61Trp) *GALNS* genotypes, respectively. We are re-analyzing the family samples to corroborate genotype results. *A posteriori*, evaluation of N-acetylgalactosamine-sulfate sulfatase activity in the *propositus* showed decreased enzymatic activity in dried blood spot samples and leukocytes. Urinary GAGs analysis by metachromatic staining was negative. **Discussion:** The present case illustrates the clinical utility of the DNA sequencing approach to diagnosing atypical phenotypes unsuspected by clinical examination. Our patient does not show the classic dysmorphological and skeletal phenotype of MPS IVA, such as short stature, coarse face, prognathism, short trunk, prominent sternum, inguinal hernia; and evident ophthalmologic, respiratory, cardiac and/or hepatic involvements. According to The International Skeletal Dysplasia Registry database, atypical forms of MPS IVA are often radiographically confused with other skeletal dysplasias, more common Legg- Calvé-Perthes disease, spondylo-epiphyseal dysplasia, and multiple epiphyseal dysplasias, as described herein. **Conclusion:** The diagnosis of MPS IVA, including its atypical forms, must be considered in the clinical approach to bone dysplasias. Molecular diagnosis could be considered a useful tool to achieve a definitive diagnosis in these patients.

Session Title: Mendelian Phenotypes Poster Session I

PB4920 Therapeutic effects of Antisense Oligonucleotide for Treatment of VCP Multisystem Proteinopathy.

Authors:

V. Kimonis[1], P. Pallabi[1], M. Youssef[1], L. Weiss[2], C. Cheng[3], Y. Fazeli[1], A. Shmara[4], M. Carrer[5]; [1]Univ CA Irvine, Irvine, CA, [2]Univ. of California, Irvine, Irvine, CA, [3]UC Irvine, Irvine, CA, [4]UCI, Irvine, CA, [5]Ionis Pharmaceuticals, Inc., Carlsbad, CA

Valosin-containing protein (VCP) disease is an autosomal dominant disease caused by gain-of function pathogenic variants in the *VCP* gene. The disease is associated with inclusion body myopathy with early-onset Paget's disease of the bones, frontotemporal dementia and familial amyotrophic lateral sclerosis, also known as multisystem proteinopathy 1 (MSP1). We hypothesize that regulating VCP hyperactivity to normal levels can reduce the disease pathology. We propose this could be achieved through the reduction of expression in VCP with the use of antisense oligonucleotides (ASOs), a method already approved for spinal muscular atrophy. In this study, we utilized the transgenic mouse model of VCP disease which overexpresses the humanized VCP gene with the severe A232E mutation. After establishing the validity of the mouse model, we screened three ASOs specifically targeting the human VCP gene by administrating ASOs in the mice harboring wildtype VCP transgene weekly for 8 weeks. ASO #1 showed over 50% knockdown of VCP in quadriceps and diaphragms at the mRNA level and a similar knockdown effect at the protein level. At the end-point toxicity analysis, ASO #1 demonstrated a non-significant elevation in organ-to-body weight ratio indicating minimal toxicity. We focused on ASO #1 to continue with a subsequent long-term treatment regimen to test whether VCP ASOs can ameliorate VCP disease pathology for symptomatic mice. We treated VCP A232E mice starting from 4 months of age for 3 months and performed monthly motor tests. Interestingly, the VCP A232E mice showed improvements in the inverted screen tests across 3-month of ASO treatment compared to mice treated with control ASO. We also found reduction at the protein level in quadriceps and diaphragm upon treatment with ASO #1 as compared to control ASO in VCP, and autophagy markers. These results suggest that knockdown of the mutant VCP allele early in asymptomatic mice could be beneficial in preventing progression of the myopathy, and has the potential of improving the clinical features of MSP1 in patients.

Session Title: Mendelian Phenotypes Poster Session II

PB4921 Three KINSSHIP syndrome patients with mosaic and germline *AFF3* variants.

Authors:

Y. Inoue[1], N. Tsuchida[1,2], N. Okamoto[3], S. Shuichi[4], K. Ohashi[5], S. Saitoh[5], A. Ogawa[6], K. Hamada[7], M. Sakamoto[1], N. Miyake[1,8], K. Hamanaka[1], A. Fujita[1], E. Koshimizu[1], S. Miyatake[1], T. Mizuguchi[1], K. Ogata[7], Y. Uchiyama[1,2], N. Matsumoto[1]; [1]Dept. of Human Genetics, Yokohama City Univ. Graduate Sch. of Med., Yokohama, Japan, [2]Dept. of Rare Disease Genomics, Yokohama City Univ. Hosp., Yokohama, Japan, [3]Dept. of Med. Genetics, Osaka Women's and Children's Hosp., Izumi, Japan, [4]Dept. of Pediatrics, Osaka Med. and Pharmaceutical Univ. Hosp., Osaka, Japan, [5]Dept. of Pediatrics and Neonatology, Nagoya City Univ. Graduate Sch. of Med. Sci., Nagoya, Japan, [6]Dept. of Pediatrics, Chikushi Hosp., Fukuoka Univ., Fukuoka, Japan, [7]Dept. of Biochemistry, Yokohama City Univ. Graduate Sch. of Med., Yokohama, Japan, [8]Dept. of Human Genetics, Res. Inst., Natl. Ctr. for Global Hlth.and Med., Tokyo, Japan

AFF3 at 2q11.2 encodes the nuclear transcriptional activator AF4/FMR2 Family Member 3. AFF3 constitutes super elongation complex like 3, which plays a role in promoting the expression of genes involved in neurogenesis and development. The degron motif in AFF3 with nine highly conserved amino acids is recognized by E3 ubiquitin ligase to induce protein degradation. Recently, six *AFF3* missense variants in this region and two variants featuring deletion including this region were identified and shown to cause KINSSHIP syndrome, named after characteristic clinical features, horseshoe KIdney, Nievergelt/Savarirayan type of mesomelic dysplasia, Seizures, Hypertrichosis, Intellectual disability (ID), and Pulmonary involvement. The variant outside the degron is also reported as causing a developmental delay (DD) and causing skeletal abnormalities. In this study, we identified two novel and one previously reported missense variants in the degron of AFF3 in three unrelated Japanese patients. These variants are not registered in the public control database. Notably, novel two of these three variants exhibited mosaicism in the examined tissues and have not been registered in the Catalogue of Somatic Mutations in Cancer. Although our three affected patients differed in the presence or absence of the typical phenotype and the degree of the phenotype differed among them, all patients consist with KINSSHIP syndrome. This study suggests that mosaic variants also cause KINSSHIP syndrome, showing various phenotypes.

Session Title: Mendelian Phenotypes Poster Session III

PB4922 Three patients with Wiedemann-Steiner syndrome

Authors:

K. Takano[1,2], T. Kawai[3], T. Yamaguchi[1,2,4], K. Wakui[2,1], K. Nakabayashi[3], K. Hata[3,5], S. Saitoh[6], T. Kosho[2,1,4]; [1]Ctr. for Med. Genetics, Shinshu Univ. Hosp., Matsumoto, Japan, [2]Dept. of Med. genetics, Shinshu Univ. Sch. of Med., Matsumoto, Japan, [3]Dept. of Maternal-Fetal Biology, Natl. Res. Inst. for Child Hlth.and Dev., Tokyo, Japan, [4]Div. of Clinical Sequencing, Shinshu Univ. Sch. of Med., Matsumoto, Japan, [5]Dept. of Human Molecular Genetics, Gumma Univ. Sch. of Med., Gunma, Japan, [6]Dept. of Pediatrics and Neonatology, Nagoya City Univ. Graduate Sch. of Med. Sci., Nagoya, Japan

Background: Wiedemann-Steiner syndrome (MIM605130: WDSTS)is an autosomal dominant neurodevelopmental disorder caused by pathogenic variants in the *KMT2A* gene. *KMT2A* encodes a histone methyltransferase which regulates chromatin-mediated transcription of multiple genes. WDSTS is characterized by developmental delay (DD), intellectual disability (ID), facial features, hypertrichosis, and additional clinical features. The features associated with WDSTS overlap those of other genetic syndromes and genetic testing is required for accurate diagnosis. In April 2014, the "ID clinic" was established at the Center for Medical Genetics, Shinshu University Hospital. We provide clinical diagnosis, systematic genetic evaluation, and genetic counseling to patients with DD/ID. **Patients:** To date, 281 patients with DD/ID have visited the "ID clinic". ID-multigene panel testing detected heterozygous *KMT2A* variants in three patients (3/281=1%). Two variants (nonsense: 1, missense: 1) have been already known as pathogenic, and the other nonsense variant occurring in the last exon of *KMT2A* is predicted to escape nonsense-mediated mRNA decay and classified as uncertain significance by ClinVar. Comprehensive DNA methylation analysis of this patient's genomic DNA reveals WDSTS-specific episignatures and leads to a definitive diagnosis. All patients presented with mild to moderate ID, short stature, characteristic facial features, and dental anomalies. Two of three patients had hypertrichosis, failure to thrive, and hypotonia in infancy. The patient with a missense variant in the CXXC DNA-binding domain presented with moderate ID and behavior problems. **Conclusion:** Comprehensive DNA methylation analysis is useful for reclassifying a variant of uncertain significance that could lead to an accurate diagnosis. As with previous reports, the patient with a missense variant in the CXXC DNA-binding domain had more severe neurodevelopmental issues.

Session Title: Mendelian Phenotypes Poster Session I

PB4923 *TNNI1* variants disrupt sarcomere contractility resulting in hypo- and hypercontractile muscle disease

Authors:

S. Donkervoort[1], M. van de Locht[2], D. Ronchi[3], J. Reunert[4], C. McLean[5,6], M. Zaki[7], R. Orbach[1], J. de Winter[8], S. Conijn[8], A. Viaene[9], F. Malik[10], N. Acquaye[1], V. Bolduc[1], M. Davis[11], M. Essawi[7], S. Gorokhova[1], R. Bryson-Richardson[6], F. Magri[12], S. Zanotti[12], N. ElBagoury[7], P. Kekenes-Huskey[13], L. Napoli[12], J. Park[14], G. Pietro Comi[3,12], W. Sharaf-Eldin[7], T. Marquardt[4], G. Ravenscroft[11], C. Bönnemann[1], C. Ottenheijm[8], TNNI1 study group; [1]NIH, Bethesda, MD, [2]Amsterdam UMC (location VUmc), Amsterdan, Netherlands, [3]Univ. of Milan, Milan, Italy, [4]Univ. of Münster, Münster, Germany, [5]Alfred Hosp., Melbourne, Australia, [6]Monash Univ., Melbourne, Australia, [7]Natl. Res. Ctr., Cairo, Egypt, [8]Amsterdam UMC (location VUmc), Amsterdam, Netherlands, [9]Univ. of Pennsylvania Perelman Sch. of Med., Philadelphia, PA, [10]Cytokinetics Inc, South San Francisco, CA, [11]QEII Med. Ctr., Nedlands, Australia, [12]Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, Milan, Italy, [13]Loyola Univ., Chicago, IL, [14]Univ. Hosp. Münster, Münster, Germany

Troponin I (TnI) regulates thin filament activation and muscle contraction. Two isoforms, TnI-fast (*TNNI2*) and TnI-slow (*TNNI1*), are predominantly expressed in fast- and slow-twitch myofibers respectively. *TNNI2* variants are a rare cause of arthrogryposis, while *TNNI1* variants have not been conclusively established to cause skeletal myopathy. We identified both recessive loss-of-function *TNNI1* variants, as well as dominant gain-of-function *TNNI1* variants as a cause of muscle disease, each with distinct physiological consequences and disease mechanisms. For the loss-of-function scenario, we report three families with biallelic *TNNI1* variants (F1: p.R14H/c.190-9G>A, F2 and F3: homozygous p.R14C), manifesting with early onset progressive muscle weakness and rod formation on muscle histology. For the gain-of-function scenario, we report two families with a dominantly acting heterozygous *TNNI1* variant (F4: p.R174Q, F5: p.K176del), manifesting with muscle cramping, myalgias, and rod formation in F5. In zebrafish, TnI proteins with either of the missense variants (p.R14H; p.R174Q) incorporate into thin filaments. Molecular dynamics simulations suggest that the loss-of-function p.R14H variant decouples TnI from TnC, which was supported by functional studies showing a reduced force response of sarcomeres to submaximal [Ca^{2+}] in patient's myofibers. This contractile deficit was reversed by a novel slow skeletal muscle troponin activator. In contrast, patient's myofibers with the gain-of-function p.R174Q variant showed an increased force to submaximal [Ca^{2+}], which was reversed by the small-molecule drug *mavacamten.* Our findings demonstrate that *TNNI1* variants cause muscle disease with variant-specific pathomechanisms, manifesting as either a hypo- or a hypercontractile phenotype, suggesting rational therapeutic strategies for each mechanism.

Session Title: Mendelian Phenotypes Poster Session II

PB4924 Translating the effects of CMT2S variants with human-on-a-chip neuromuscular junction systems.

Authors:

C. Johnson[1], S. Smieszek[1], B. Przychodzen[1], C. Tyner[1], A. Kaden[1], C. Polymeropoulos[1], G. Birznieks[1], W. Hagan[2], L. Lenkiu[2], C. Niccum[2], H. Cannon-Patron[2], R. Brighton[2], X. Guo[3], K. Hawkins[3], N. Akanda[3], J. Hickman[2,3], M. Polymeropoulos[1]; [1]Vanda Pharmaceuticals Inc., Washington, DC, [2]Hesperos, Inc., Orlando, FL, [3]NanoSci. Technology Ctr., Univ. of Central Florida, Orlando, FL

Charcot-Marie-Tooth disease Type 2S (CMT2S) is a rare Charcot-Marie-Tooth disease subtype caused by immunoglobulin mu-binding protein 2 (*IGHMBP2*) variants that result in abnormal RNA processing leading to alpha-motor neuron degeneration. A patient was reported with consequential variants within *IGHMBP2*. Whole genome sequencing revealed a paternally inherited cryptic splice site non-coding variant (c.1235+894 C>A) deep in intron 8, which leads to nonsense-mediated decay resulting in haploinsufficiency. Our objective was to characterize the neuromuscular phenotype of this CMT2S patient by analyzing the morphology and physiology of diseased motoneurons compared to healthy motoneurons, and their subsequent neuromuscular junction (NMJ) differences. CMT2S motor neurons (CMT2S-MNs) were differentiated from an induced pluripotent stem cell (iPSC) cell line generated from the patient's fibroblasts. Patch clamp electrophysiology, phase imaging, and immunocytochemistry experiments were performed to characterize CMT2S-MNs and to identify differences in morphology. To determine NMJ defects, CMT2S-MNs and control iPSCs (WT-MNs) were integrated into a dual-chamber NMJ platform with wild-type (WT) PSC-derived skeletal muscle myofibers. NMJ functional defects were analyzed by NMJ number per chamber, NMJ fidelity, and NMJ fatigue index (FI).

Patch clamp electrophysiological analyses revealed hyperexcitability and spontaneous firing in CMT2S-MNs. Similar characteristics can be found in amyotrophic lateral sclerosis derived motoneurons (ALS-MNs), though ALS-MNs showed hyperexcitability at Day 20, whereas hyperexcitability was seen as early as Day 5 in CMT2S-MNs. Further, analysis of Na+ current normalized by membrane capacitance showed a significant decrease of membrane capacitance and membrane potential in CMT2S-MNs compared to WT-MNs. An NMJ FI functional readout revealed a low FI in CMT2S compared to WTs.

Patient-derived CMT2S-MNs revealed a hyperexcitable phenotype with spontaneous firing, comparable to amyotrophic lateral sclerosis (ALS). However, hyperexcitability in CMT2S-MNs had a much earlier onset as compared to ALS-MNs. These data confirm the observed, clinical phenotype of CMT2S that develops earlier in life than ALS. The hyperexcitability observed in CMT2S-MNs may be caused by the observed reduced resting membrane potential, reduced membrane capacitance, and thus reduced Na+ current density. NMJ FI showed quick fatigue which also correlates with a CMT2S clinical phenotype. We are further analyzing this patient-specific model to continue phenotyping CMT2S caused by *IGHMBP2* variants.

Session Title: Mendelian Phenotypes Poster Session III

PB4925 Trio exome sequencing identifies increased incidence of clonal hematopoiesis of indeterminate potential in Bloom Syndrome probands and carriers

Authors:

I. Lin, A. Wei, T. Gebo, P. Boutros, V. Arboleda, V. Chang; Univ. of California, Los Angeles, Los Angeles, CA

Bloom Syndrome (BSyn) is an autosomal recessive disorder caused by recessive or compound heterozygous variants in BLM. BLM is a RecQ helicase, a family of helicase enzymes that maintain genomic stability. The genomic instability caused by biallelic germline variants in BLM drives the incidence of early-onset cancer, particularly hematologic malignancies, in patients with BSyn. The malignancy risk in BSyn carriers (heterozygous for deleterious BLM variants), remains controversial and understudied. The underlying mechanisms leading to increased cancer risk may involve increased somatic mutation rate, de novo mutations that are not inherited from parents, aberrant epigenetic alterations which can be caused by mutations in DNA methylation (DNAm) genes, or through tissue-specific somatic mutations that drive clonal expansion. For example, in blood, clonal hematopoiesis of indeterminate potential (CHIP) is caused by somatic mutations in leukemia-related "CHIP" genes in individuals without overt evidence of leukemia but associated with increased risk of leukemia.

In this study, we performed exome sequencing on BSyn probands (n=10), BSyn carriers (n=19), control children (n=19) and control adults (n=38). We obtained these 19 age- and sex-matched control trios from the publicly available dbGAP study phs001272.v1.p1 deposited by the Broad Institute Center for Mendelian Genomics. We analyzed the prevalence of CHIP, the presence of somatic variants in DNAm genes, and the occurrence of de novo germline mutations using a multinomial model.

We discovered significantly increased occurrence of low-frequency, putative somatic CHIP variants in both BSyn probands (p-value = 0.0017) and BSyn carriers (p-value = 1.4093E-06) compared to age-matched controls. We also identified an increased occurrence of somatic variants in DNAm genes in BSyn carriers. These findings suggest that the presence of even just one germline pathogenic variant in BLM could be sufficient to increase the risk of clonal hematopoiesis, which may be used as a biomarker of aging, cancer, cardiovascular disease, morbidity and mortality. We find statistically-significant, increased CHIP and somatic DNAm gene variants in BSyn probands and Bsyn carriers compared to age-matched controls. Our findings contribute to growing literature that carriers for genes important in maintaining genomic integrity also have a higher somatic mutation rate and increased cancer risk. Further studies are warranted to confirm these findings in population-scale cohorts and to follow up on the impact of BLM loss-of-function variants on cancer risk.

Session Title: Mendelian Phenotypes Poster Session I

PB4926 Triple molecular diagnosis of Wolf Hirschhorn syndrome,20p duplication syndrome and Frontotemporal dementia and/or amyotrophic lateral sclerosis

Authors:

Y. Ben Moshe, V. Maria Claudia, N. Owen, J. E. Posey; Baylor Coll. of Med., Houston, TX

Background: Wolf Hirschhorn syndrome (WHS) is a chromosomal disorder involving a deletion of the short arm of chromosome 4. Although dual and multiple molecular diagnoses have been reported to occur in at least 5% of individuals for whom exome sequencing (ES) is diagnostic, no individuals with WHS have been reported to have multiple molecular diagnoses unrelated to the WHS genomic rearrangement. Case presentation: A 37-year-old white female with WHS was referred for episodic general mental and functional deterioration. At 3y of age, she had been found to have an unbalanced translocation [der(4)t(4;20)(p16.2;p13)], establishing molecular diagnoses of WHS and 20p duplication syndrome. From 34 to 37 y/o, she had ~6 dementia-like episodes during concurrent infectious illness. Each week-long episode was characterized by disorientation, decline in her ability to manage activities of daily living, and severe sleep disturbances. Between episodes, she would improve but never quite reach her previous baseline level of functioning. Her physical examination did not reveal new neurologic insults. A repeat chromosomal analysis and chromosomal microarray confirmed and refined the breakpoints of her unbalanced translocation der(4)t(4;20)(p16.2;p13). Proband ES showed two pathogenic variants, including a deletion of *NSD2* in the WHS critical region (expected), and a heterozygous variant in *SQSTM1* c.1175C>T(p.Pro392Leu). Notably, both WHS and 20p duplication syndrome have been independently associated with severe developmental delay and intellectual disability, dysmorphic features, as well as other anomalies. Thus, it is challenging to isolate the phenotypic contribution of the 20p13 duplication. The pathogenic variant in *SQSTM1* uncovers a third condition of Paget disease of the bone (OMIM#167250) and Frontotemporal dementia and/or amyotrophic lateral sclerosis (OMIM#616437). A follow-up evaluation revealed that the current mental and functional status of the proband had returned almost to baseline with no clinical signs of myopathy. A recent brain MRI did not show degenerative findings. Alkaline phosphatase level was 82 U/L (ref 40-112). The potential contribution of *SQSTM1* to the proband's episodic neurological decline is unclear, and routine follow-up is warranted to monitor for any neurological or skeletal changes. Conclusions:To the best of our knowledge, there are no other similar individuals with WHS and a second, unrelated molecular diagnosis. This case underscores the potential role of ES in identifying multiple molecular diagnoses, particularly when evolving phenotypes do not fit with the expected disease course of the primary disorder.

Session Title: Mendelian Phenotypes Poster Session II

PB4927 Two additional cases of GM3 synthase deficiency in non-Amish patients

Authors:

J. Muskett[1], J. M. Chambers[1], M. Franchi[2], P. Trapane[3], M. Y. Platt[1]; [1]Quest Diagnostics, Marlborough, MA, [2]UF Hlth.- Jacksonville, Jacksonville, FL, [3]UF Coll. of Med. – Jacksonville, Jacksonville, FL

The *ST3GAL5* gene encodes an enzyme responsible for the formation of ganglioside GM3, which is an important precursor in the production of many biologically important glycosphingolipids. Pathogenic variants in *ST3GAL5* result in low plasma GM3 ganglioside levels, which manifests clinically as global developmental delay, progressive microcephaly, failure to thrive, visual impairment, involuntary movements, and epilepsy. GM3 synthase deficiency was initially described as an autosomal recessive disorder in the Amish population in which there is a founder variant, p.Arg288*, in *ST3GAL5*. With the increase in whole exome sequencing, researchers have recently begun to identify more cases of GM3 synthase deficiency in non-Amish populations, including a recent review that described several families from 3 distinct regions, and included 3 new founder alleles. Our laboratory recently identified 2 cases of this disorder in non-Amish individuals through trio whole exome sequencing. In the first case, the pathogenic homozygous nonsense variant, p.Arg288*, was detected in a patient with Palestinian ancestry. This is the third example in the literature of this variant occurring in a non-Amish family. This patient's clinical features included global developmental delay, microcephaly, encephalomalacia, cortical visual impairment, and an abnormal electroencephalogram (EEG). In the second case, a different pathogenic homozygous nonsense variant, p.Arg334*, was detected in an African American patient. This variant was previously reported in Italian siblings as compound heterozygous with a missense variant. This patient's clinical features included global developmental delay, sensorineural hearing loss, and a skin rash described as ichthyosis. Together, the 2 patients described here illustrate the phenotypic variability of this disorder, while also sharing many core features. Additionally, these cases contribute to the growing body of literature of non-Amish patients with GM synthase deficiency, showing the disorder is present in diverse populations. Finally, it provides additional evidence that p.Arg288* and p.Arg334* are recurrent variants in this gene.

Session Title: Mendelian Phenotypes Poster Session III

PB4928 Two distinct genetic loci contribute to the genetic etiology of a complex syndromic neurodevelopmental disorder.

Authors:

V. Maduro[1], K. Dixon[2,3], S. Huynh[2,4], Y. Shen[3,2], S. Jones[3,2], C. Boerkoel[2,4], M. Malicdan[1]; [1]NIH, Bethesda, MD, [2]Univ. of British Columbia, Vancouver, BC, Canada, [3]BC Cancer Agency, Vancouver, BC, Canada, [4]Women's Hosp. and Hlth.Ctr., Vancouver, BC, Canada

Background: Despite the advancements in molecular and next generation sequencing, establishing a precise diagnosis in complex syndromic disorders remains a significant challenge. This study presents the case of a 17-year-old male who presented with hypotonia, severe failure to thrive, developmental delay, poor brain myelination, cortical visual impairment, feeding difficulties, dysmorphic features and intractable seizures causing hemiparesis. Given the complexity of the phenotype, we hypothesize a digenic etiology to explain the observed manifestations. **Methods:** The patient underwent an extensive clinical and biochemical consultation at the NIH Undiagnosed Diseases Program. To investigate potential genetic causes, we employed targeted sequencing, genome and exome sequencing, as well as long-read sequencing (Oxford Nanopore Technologies, ONT) to identify candidate genes. Sanger sequencing was utilized for validation and disease segregation analysis. Additionally, functional studies were conducted using the patient's primary cells, induced pluripotent stem cells (iPSCs), and neural progenitor cells. **Results:** Genome sequencing identified a de novo missense mutation in *BRAF*, resulting in an Asp638Glu amino acid change. Notably, some phenotypic features overlapped with *BRAF*-associated Noonan syndrome type 7. Further, chromosomal analysis and sequencing show a de novo translocation 46,XY,t(6; 12)(q24.3.1;q12) that occurred within intron 4 of the *SLC2A13* gene, which encodes a membrane channel that localizes in the Golgi apparatus, and intron 6 of the STXBP5-Antisense RNA 1, which encodes a non-coding mRNA. The breakpoints of the translocation were confirmed through ONT and Sanger sequencing. Absolute quantification of gene expression demonstrated amplification of both wild-type and mutant transcripts in iPSC-derived neural progenitor cells from the patient. **Conclusions:** Our findings support a digenic etiology involving two potential disease loci contributing to the complex syndromic manifestation observed in our proband: a de novo translocation disorder that creates a *SLC2A13-STXBP5* fusion transcript, and a likely pathogenic variant in *BRAF*. Our case illustrates the complexity of establishing diagnosis, especially when considering the possibility of a digenic or multigenic inheritance. Advancements in genetic testing technologies, functional characterization of mutations, and collaborative efforts will contribute to improving diagnostic accuracy and enhancing our understanding of the underlying molecular mechanisms of these complex genetic disorders.

Session Title: Mendelian Phenotypes Poster Session I

PB4929 UBA5-related Epilepsy: From Cellular Models to Novel Therapies

Authors:

H. Chen[1], E. Almanza Fuerte[1], M. Wang[1], H. Mefford[2]; [1]St Jude Children s Res. Hosp., Memphis, TN, [2]St. Jude Children`s Res. Hosp., Memphis, TN

UBA5 encodes an E1 activating enzyme in the ubiquitin-fold modifier 1 (UFM1) pathway, regulating ER stress response. Biallelic mutations in *UBA5* cause a rare neurodevelopmental disorder characterized by epilepsy, severe developmental delay and growth failure. In ~75% of patients, one of the pathogenic alleles is a missense variant, p.A371T, that produces a hypomorphic protein, while the other pathogenic allele encodes a nonfunctional protein. Interestingly, there have been at least three healthy adults with homozygous p.A371T variants reported, suggesting that despite the decreased enzymatic activity of the p.A371T allele, the combination of two hypoactive alleles provides enough enzyme activity to prevent disease. This suggests that augmenting the expression of p.A371T may be an effective therapy. To test this, we characterize the cellular phenotypes in engineered and patient-derived cells with disease-causing UBA5 variants and in cells with homozygous p.A371T variants. We then increase the level of the p.A371T protein using CRISPRa-dCas9 or long noncoding antisense RNA modality called SINEUP to rescue cellular defects. We generated U87-MG cell lines with compound heterozygous pathogenic alleles seen in patients (p.A371T/p.R55H and p.A371T/p.F292*). We observed decreased enzymatic activity and conjugation of UFM1 onto substrates. We also noted elevated base-level ER stress, demonstrated by an increase in ER stress proteins and ER swelling. We generated a cell line with homozygous p.A371T allele and demonstrated that it behaves similarly to wildtype, with no cellular defects. To investigate neuronal phenotypes, we engineered cortical organoids from UBA5 patient-derived iPSC. Patient organoids are significantly smaller than those from unaffected controls and exhibit an increase in ER stress markers. Multi-electrode array analyses show aberrant neuronal firing in patient organoids. Moreover, we noted patient organoids showed reduced expression of GABAergic markers, suggesting a reduction in inhibitory neurons compared to controls. We are currently performing single-cell RNAseq experiments to determine neuronal composition and developmental trajectory. Finally, using UBA5-specific SINEUPs or CRISPRa, we can augment the expression of p.A371T allele in both engineered and patient-derived cells, which rescues the aforementioned defects. These findings suggest that augmenting the expression of the hypomorphic allele, p.A371T by SINEUP or CRISPRa can be a potential therapeutic outlet for UBA5 patients. Moreover, these therapeutic modalities may be extended to other neurodevelopmental disorders caused by haploinsufficiency.

Session Title: Mendelian Phenotypes Poster Session II

PB4930 Unbiased plasma proteomics analysis uncovers potential biomarkers for renal disease and pulmonary hypertension in patients with sickle cell disease.

Authors:

M. Garrett[1], M. Foster[2,3], M. Telen[3], A. Ashley-Koch[1]; [1]Duke Molecular Physiology Inst., Durham, NC, [2]Duke Proteomics and Metabolomics Core Facility, Duke Univ. Sch. of Med., Durham, NC, [3]Dept. of Med., Duke Univ. Med. Ctr., Durham, NC

Sickle cell disease (SCD) is characterized by red blood cell sickling, vaso-occlusion, hemolytic anemia, damage to multiple organ systems, and, as a result, shortened life expectancy. Sickle cell disease nephropathy (SCDN) and pulmonary hypertension (pHTN) are common and frequently co-occurring complications of SCD; both are associated with accelerated mortality. To identify potential biomarkers for SCDN and pHTN, we performed a mass spectrometry-based proteomic analysis of plasma samples from 189 adult patients from the OMG-SCD cohort. Among 1190 quantified proteins, there were 107 proteins that associated with SCDN (false discovery rate [FDR] $q<0.05$), most prominently β-1,4-GalT1 ($p=1.3x10^{-10}$), cystatin C ($p=4.4x10^{-9}$), and IGFBP6 ($p=1.5x10^{-8}$). No proteins were associated with pHTN by itself ($q>0.05$). However, 42 proteins were associated with the co-occurrence of SCDN and pHTN (SCDN+pHTN; $q<0.05$), the most prominent of which were β-1,4-GalT1 ($p=1.4x10^{-6}$) and PCPE-1 ($p=2.3x10^{-6}$). Of interest, the APOL1 peptides were not associated with SCDN ($p>0.05$). This is consistent with our recent GWAS, showing that SNPs in *APOL1* were not strongly associated with SCDN outcomes. Next, we conducted network analysis to identify modules of co-regulated proteins associated with SCDN and pHTN. Cystatin C was removed prior to network analysis as it can be used to estimate SCDN and is highly correlated with serum creatinine in our cohort ($r=0.72$, $p<0.0001$). One module of co-regulated proteins, termed MEblue, was significantly associated with SCDN ($p=8.0x10^{-6}$) and SCDN+pHTN ($p=1.9x10^{-3}$). Pathway analysis of the 134 proteins in MEblue identified many FDR-significant pathways, including enrichment of glycoproteins, insulin-like growth factor binding, those involved in cell adhesion (such as collagen and basement membrane cellular components), EGF-like calcium binding, and several cadherin family members. Finally, we refined our previously reported renal risk score by including the MEblue module and observed an improved odds ratio: for each one-point increase in the renal risk score, the odds of low eGFR increased 2.195-fold ($p<0.0001$). In a subset of this cohort with longitudinal eGFR measurements (N=67), the odds of rapid renal decline increased with the renal risk score as well. In summary, we identified many proteins associated with SCDN and pHTN, including a network of co-regulated proteins that improves disease prediction and could be utilized as potential biomarkers to facilitate early detection of SCDN and pHTN and help forestall disease progression and early mortality in SCD.

Session Title: Mendelian Phenotypes Poster Session III

PB4931 Uncover Phenotypic Networks through Mutation Carriers at Mendelian Disease Genes: a case Study of Autosomal Dominant Polycystic Kidney Disease (ADPKD)

Authors:

W. Zheng[1], Y. Xie[1], J. Gu[1], G. Zhou[1], S. Somlo[2], H. Zhao[1]; [1]Yale Univ. Sch. of Publ. Hlth., New Haven, CT, [2]Yale Univ Sch Med., New Haven, CT

Although thousands of chromosomal regions and genes have been implicated for thousands of traits, and many statistical methods have been developed to identify traits with shared genetic effects, it is still challenging to infer the phenotypic relationship networks with these methods. Mendelian diseases are disorders where patients for a given disorder often carry rare deleterious genetic mutations in specific genes. Analyzing phenotypes from individuals carrying loss-of-function mutations in Mendelian disease genes offers a new angle to infer relationships among phenotypes. In this project, we focus on Autosomal Dominant Polycystic Kidney Disease (ADPKD), a distinct genetic disorder mainly caused by mutations in the PKD1 and PKD2 genes. We aim to investigate the phenotypic interrelationships among ADPKD rare variant carriers.

We identified rare (MAF<0.01) loss-of-function variants on the PKD1 and PKD2 genes using the whole-exome sequencing data in the UK biobank subjects. We defined carrier and non-carrier groups based on the carrying status of the selected rare variants. We used the hierarchical structure information of ICD-10 codes to explore the phenotypic relationship, and constructed phenotype networks for the two groups. We further compared the intra-group phenotypic similarity scores for carriers and non-carriers and identified phenotypes with significant differences between the two groups.

Our preliminary results show 1) many traits are associated with ADPKD rare variant carrier status. The top significant phenotypes are highly related to kidney and cystic, such as ADPKD, chronic kidney disease, acute renal failure, and so on. Intriguingly, some significantly different phenotypes not directly related to ADPKD-affected organs were also observed, including disorders of the atrial fibrillation and flutter, acute lower respiratory infection and so on. Also, although ADPKD shows significantly different pattern in the two groups, more than 60% carriers do not have ADPKD diagnosis. 2) In comparing the intra-group phenotypic similarity scores of the two groups, a notable difference was found, and it was enhanced when phenotypes were selected by their carrier status and ADPKD patient status, suggesting that specific phenotypes primarily drive these differences. Our study is among the first to integrate genetic data and structured phenotypic data to discover genotype-phenotype relationship of ADPKD in the UK Biobank. These preliminary findings provide new insights into the phenotypic variations among ADPKD variant carriers and lay a foundation for further exploration of this approach to other Mendelian disease genes.

Session Title: Mendelian Phenotypes Poster Session I

PB4932 Understanding heterogeneity of POLG related disorders for drug discovery

Authors:

y. liu[1], S. Yoo[1], E. Lee[1], Q. Chen[1], S. McGee[2], L. Wang[1], J. Zhu[1]; [1]GeneDx, Stamford, CT, [2]GeneDx, Gaithersburg, MD

POLG encodes the catalytic subunit of mtDNA polymerase that is critical for integrity of the mitochondrial genome. POLG-related conditions are autosomal recessive genetic diseases including a spectrum of severity with clinical features including neurological features, epilepsy, liver dysfunction, and myopathy. Variations in POLG interfere with mtDNA replication and result in mtDNA deletion/depletion. The disease onset and severity are correlated in POLG-related disorders with early onset associated severe diseases. Oxidative phosphorylation (OXPHOS) in mitochondria is essential for all cells. Brain, liver, and muscle cells have the highest dependency on OXPHOS. The molecular pathogenesis of POLG-related disorders is still poorly understood, and there is no efficient treatment for curing the diseases.

Genes transcriptionally co-regulated with POLG in 60 human tissues were significantly overlapped, suggesting common molecular mechanisms in all cells. We then investigated cell type-specific POLG expression and inferred OXPHOS activity in diverse human cells. Across 79 human cell types, brain cells have the lowest OXPHOS activities and are among cells expressing the lowest level of POLG. Furthermore, during development, OXPHOS activities in brain cells increased from fetal to adult. These results indicate that disturbing POLG function or OXPHOS activity in brain cells, especially in early development stage, may have the highest impact on cell function.

It has been suggested that early and late onset of POLG-related diseases are dictated by variations in POLG. However, the same pair of POLG variants (needed for autosomal recessive diseases) occurred in patients with early and late onset diseases, suggesting that pairs of POLG variants are not sufficient to predict POLG-related disease onset and severity. Other genetic and environmental factors may contribute to early or late onset of POLG-related diseases. For variants in calmodulin binding protein were likely associated with early onset of the disease. Similarly, a recent drug screen in zebrafish shows Colfilium Tosylate (CT), a potassium channel blocker, can rescue POLG-related disease phenotypes. We compared gene expression changes in human cells treated with low dose of CT, and identified a set of up-regulated genes which were enriched for genes in multiple stress response pathways, such as HEME deficiency response (Fisher exact test $p=1.9x10^{-10}$). Thus, whole genome sequencing or whole exome sequencing beyond sequencing POLG is necessary to understand heterogeneity of POLG-related disease onset and critical to develop personalized medicine for these patients.

Session Title: Mendelian Phenotypes Poster Session II

PB4933 Unraveling the Intrafamilial Phenotypic Variability in Sibling Pairs with Neurodevelopmental Diseases

Authors:

T. Bozkurt Yozgatli[1,2], D. Pehlivan[3,4,5,6], R. A. Gibbs[7,8], U. Sezerman[2], J. R. Lupski[3,6,7,8], J. E. Posey[7], Z. Coban Akdemir[1,7]; [1]Human Genetics Ctr., Dept. of Epidemiology, Human Genetics, and Environmental Sci., Sch. of Publ. Hlth., Univ. of Texas Hlth.Sci. Ctr. at Houston, Houston, TX, [2]Dept. of Biostatistics and Bioinformatics, Inst. of Hlth.Sci., Acibadem Mehmet Ali Aydinlar Univ., Istanbul, Turkey, [3]Dept. of Pediatrics, Baylor Coll. of Med., Houston, TX, [4]Section of Pediatric Neurology and Dev.al NeuroSci., Dept. of Pediatrics, Baylor Coll. of Med., Houston, TX, [5]Jan and Dan Duncan Neurological Res. Inst. at Texas Children's Hosp., Houston, TX, [6]Texas Children's Hosp., Houston, TX, [7]Dept. of Molecular and Human Genetics, Baylor Coll. of Med., Houston, TX, [8]Human Genome Sequencing Ctr., Baylor Coll. of Med., Houston, TX

Introduction: Mendelian studies have focused on the impact of single-gene mutations; however, multi-locus pathogenic variants (MPVs) can be a key contributor to phenotypic variability between individuals diagnosed with the same disease. Here, we identified potential MPVs causing intra-familial phenotypic variability in a group of Turkish sibling pairs previously diagnosed with rare neurodevelopmental disorders (NDD), supporting our previous hypothesis that the 'absence of heterozygosity rare deleterious variant burden drives MPV' (PMID: 29790871, 31230720, and 34582790). Method: We reanalyzed exome sequencing (ES) data from 40 affected sibling pairs with shared molecular diagnoses of a wide variety of NDDs recruited from an admixed Turkish population, which exhibited a high level of consanguinity. We detected ROH regions from ES data using BafCalculator and classified them into three size categories (short, medium, and long) based on the previously defined size cut-offs for the Turkish population. Results: In addition to the initial molecular diagnosis, we identified homozygous variants at a second locus, driven by long ROH regions, that are relevant to the phenotypic expansions in one of the siblings in each of four families. In family HOU1842, we revealed additional known variant G61E in *CYP1B1* (OMIM *601771, PMID: 9497261) reported as pathogenic in ClinVar (Variation ID: 7730) in the sibling BAB4134. In pedigree HOU2437, we reported a novel variant R519C in *MFN2* (OMIM *608507), classified as likely pathogenic based on the American College of Medical Genetics and Genomics (ACMG) criteria associated with Charcot-Marie-Tooth disease type 2A2B in BAB6511. In family HOU2280, we identified a novel variant G671E in *ECEL1* (OMIM *605896) classified as likely pathogenic in ACMG associated with Arthrogryposis type-5D in sibling BAB6025. In family HOU4131, we detected a previously reported variant R6C in *PLA2G6* (OMIM *603604, PMID: 33547378) associated with infantile neuroaxonal dystrophy 1 in the sibling BAB11388. It is noteworthy that the siblings with MPVs exhibited significantly larger total size of long ROH regions in all families (Paired t-test, p-value = 6.9e-3). Conclusion: We found MPVs in 10% of families (4/40). Based on our findings, MPVs may exist in families with affected sibling pairs, and ROH can be utilized as an adjuvant tool to uncover an MPV wherein the second locus parsimoniously explains intra-familial phenotypic differences. Overall, this study sheds light on the significance of considering MPVs and the role of ROH in explaining phenotypic variability within families, which includes individuals affected by rare disorders.

Session Title: Mendelian Phenotypes Poster Session III

PB4934 Unveiling the Genetic Landscape of Inherited Retinal Diseases: Insights from a Large Pakistani Cohort

Authors:

M. Ullah[1,2], A. U. Rehman[3], A. Rashid[4], A. Iqbal[5], S. Javed[6], M. Dawood[7], H. Baig[1], S. Saleha[7], S. Naz[5], H. Kausar[8], A. Waryah[9], M. Quinodoz[1,2,10], A. Superti-Furga[11], M. Ansar[12,13], C. Rivolta[1,2,10]; [1]Inst. of Molecular and Clinical Ophthalmology Basel, Basel, Switzerland, [2]Dept. of Ophthalmology, Univ. of Basel, Basel, Switzerland, [3]Dept. of Zoology, Faculty of Biological and Hlth.Sci., Hazara Univ., Khyber Pakhtunkhwa, Pakistan, Mansehra, Pakistan, [4]Dept. of Higher Ed., Archives and Libraries Peshawar, Khyber Pakhtunkhwa, Basel, Pakistan, [5]Dept. of Zoology, Lahore Coll. for Women Univ., Lahore, Pakistan, [6]Dept. of Biotechnology, Lahore Coll. for Women Univ., Lahore, Pakistan, [7]Dept. of Biotechnology and Genetic Engineering, Kohat Univ. of Sci. and Technology, Kohat, Pakistan, [8]Dept. of Biotechnology, Kinnaird Coll. for Women, Lahore, Pakistan, [9]Molecular Biology and Genetics Dept., Med. Res. Ctr., Liaquat Univ. of Med. and Hlth.Sci., Jamshoro, Pakistan, [10]Dept. of Genetics and Genome Biology, Univ. of Leicester, Leicester, United Kingdom, [11]Lausanne Univ. Hosp., Lausanne, Switzerland, [12]Jules-Gonin Eye Hosp., Univ. of Lausanne, Lausanne, Switzerland, [13]Advanced Molecular Genetics and Genomics Disease Res. and Treatment Ctr., Dow Univ. of Hlth.Sci., Karachi, Pakistan

Background: Inherited retinal diseases (IRDs) are a group of rare Mendelian disorders that causes degeneration of retinal cells, resulting in progressive vision loss and potential blindness. They are primarily caused by mutations in any of the many genes that are essential for the development, the maintenance, and the normal function of the retina. Pakistan, as the fifth most populous country globally, presents a unique population for genetic studies due to its distinctive features, including a high prevalence of consanguinity and diverse ethnic groups. In this study, we investigated a large cohort from this country, with the aim of understanding the genetic architecture of IRDs in this population.**Methods:** A total of 722 patients were ascertained from three provinces of Pakistan: Khyber Pakhtunkhwa (155 families), Sindh (29 families), and Punjab (28 families). Inclusion criteria for family recruitment was to have at least two affected siblings exhibiting night blindness or progressive vision loss. Genomic DNA was extracted from blood or saliva samples and analyzed using both whole exome sequencing and Sanger sequencing.**Results:** The analyzed cohort comprised 212 indexed cases, each representing a unique family. Of these, 117 (55%) were female and 95 (45%) were male. In terms of inheritance patterns, 201 families (95%) exhibited autosomal recessive, 6 families (3%) showed autosomal dominant, and 5 families (2%) had X-linked inheritance. Overall, the post-analysis diagnostic success rate was 87%. Clinical phenotypes encompassed retinitis pigmentosa (29%), Leber congenital amaurosis (20%), Stargardt (8%), Usher (7%), and other IRDs. Exome sequencing revealed a total of 140 unique variants across 62 different genes, with the most frequently mutated genes being *ABCA4* (10%), *CRB1* (8%), *MYO7A* (6%), and *PDE6B* (5%). Among these unique variants, 60 were unreported in clinical databases. The most common variants observed in the study were: *ABCA4* c.214G>A, p.Gly72Arg (15 alleles from 8 families); *CRB1* c.1459T>C, p.Ser487Pro; (14 alleles from 7 families); *MYO7A* c.4838del, p.Asp1613fs (8 alleles from 4 families); and *PDE6A* c.1444T>C, p.Cys482Arg; c.304C>A, p.Arg102Ser; *PDE6B* c.1921-20_1921-3del; and *PDE6C* c.633G>A, p.Glu=, each present as 6 alleles from three unrelated families. Missense variants were the most frequently observed category of pathogenic variants (42%), followed by indels (25%) and stopgain variants (21%).**Conclusion:** These findings contribute to our understanding of the genetic basis of IRDs in Pakistan and have implications for both diagnosis and future development of therapies.

Session Title: Mendelian Phenotypes Poster Session I

PB4935 Using long-read sequencing to identify complex damaging variants in familial Parkinson's disease

Authors:

K. Daida[1,2], M. Funayama[2], K. Billingsley[1], L. Malik[1], A. Miano-Burkhardt[1], M. Ishiguro[2], H. Yoshino[2], Y. Li[2], K. Nishioka[2], C. Blauwendraat[1], N. Hattori[2]; [1]NIH, Bethesda, MD, [2]Juntendo Univ. Sch. of Med., Tokyo, Japan

Monogenic Parkinson's disease (PD) accounts for 5~10% of PD. Monogenic PD has autosomal dominant hereditary forms and autosomal recessive (AR) hereditary forms. PRKN and PINK1 are the most common causative genes of ARPD. The significance of biallelic variants of ARPD genes is well established as the cause of ARPD, however, the significance of heterozygous variants is conflicting. For example, in PRKN, the most frequent causative gene in ARPD, the role of heterozygous PRKN variants in PD is controversial; where some reports show an increased risk of carrying a single damaging variant, and others report no effect. There are patients where only one variant is identified although they fit in the typical PRKN phenotype and have young onset PD. We hypothesize that there are unrevealed second damaging variants in ARPD genes, which short-read sequencing and other conventional methods cannot identify, especially in young onset PD cases. To identify the causative variant in ARPD genes from PD subjects, we used targeted resequencing for PD-related genes, multiplex ligation-dependent probe amplification (MLPA), and/or short-read whole exome sequencing (WES). PD subjects only harbor heterozygous PRKN or PINK1 causative variants, without other known causative variants in PD-related genes, were performed long read whole genome sequencing. To confirm the identified variants by long-read sequencing, Sanger sequencing was performed. Long-read sequencing identified large structural variants which MLPA and short-read WES could not determine from ARPD subjects. This study shows that long-read sequencing is a useful tool to identify structural variants in ARPD genes which cannot be identified by conventional short-read sequencing methods. Long-read sequencing may expand the role of the ARPD genes.

Session Title: Mendelian Phenotypes Poster Session II

PB4936 † Using precision animal models to support the discovery of a new AXIN2-related disorder.

Authors:

L. Burrage[1], D. Lanza[1], P. Marcogliese[2], D. Lu[1], C-W. Hsu[1], N. Aceves[1], M. Gonzalez[1], A. Christiansen[1], T. Rasmussen[1], A. Gaspero[1], J. Seavitt[1], M. Dickinson[1], B. Shayota[3], S. Pachter[4], D-L. Day-Salvatore[5], O. Kanca[1], M. Wangler[1], L. Potocki[6], J. Rosenfeld[1], B. Lee[1], Undiagnosed Diseases Network, Baylor College of Medicine Knockout Mouse Production and Phenotyping Project, Baylor College of Medicine Center for Precision Medicine Models, S. Yamamoto[1], H. Bellen[1], J. Heaney[1]; [1]Baylor Coll. of Med., Houston, TX, [2]Univ. of Manitoba, Winnipeg, MB, Canada, [3]Univ. of Utah Hlth., Salt Lake City, UT, [4]Atlantic Hlth.System, New Brunswick, NJ, [5]Saint Peter's Univ. Hosp., New Brunswick, NJ, [6]Baylor Col Med/TX Child Hosp, Houston, TX

In the BCM Undiagnosed Diseases Network (UDN), ~66% of variants associated with new disorders or phenotypic expansion are *de novo* and heterozygous. Modeling these variants in mice with traditional approaches is challenging if the variant causes a lethal, early-onset phenotype that prevents production of the line. Consequently, time-consuming strategies, such as conditional knock-in approaches, are needed to model these variants in mice. In the BCM Center for Precision Medicine Models, we hypothesized that the generation of low-level germline mosaic founder mice that are viable and fertile with prime editing might be a complementary, more rapid approach for assessing variant pathogenicity. We tested this hypothesis with a *de novo* variant in *AXIN2* [NM_004655.4:c.196G>A (p.Glu66Lys)] identified by the UDN in a patient with phenotypes, such as global developmental delay, microcephaly, ectodermal dysplasia, limb anomalies, bilateral cystic renal dysplasia, eye anomalies, and short palate, that only partially overlap with AXIN2-associated oligodontia-colorectal cancer syndrome (ODCRS). With prime editing, we made a single mosaic (20%) founder with the variant. The founder male was viable and fertile, but no living offspring (N1) had the variant. However, the heterozygous variant was detected in 48% of embryonic day (E)18.5 embryos, a subset of which had tail anomalies (36%), limb anomalies (7%), and small size (21%). Moreover, all heterozygous E18.5 embryos imaged with microCT had a short palate like the patient. Heterozygous p.Glu66Lys embryos did not phenocopy an *Axin2* knockout mouse generated and characterized by BCM KOMP2. Homozygous knockout, E18.5 embryos had a broader range of anomalies and heterozygous knockout adult mice only had abnormal skin coloration. Thus, loss-of-function is unlikely to be the pathogenic mechanism of this variant. In line with this, overexpression of human *AXIN2* p.Glu66Lys variant cDNA in the *Drosophila* wing showed a gain-of-function phenotype compared to reference that also produced different phenotypes than truncations observed in ODCRS. In contrast, the p.Glu66Lys acted as a loss-of-function in the developing eye, likely demonstrating context-specificity of the opposing role of Wnt-signaling in the development of these tissues. Our model organism data and identification of two other patients with similar phenotypes in Clinvar show that p.Glu66Lys is a unique variant in *AXIN2* associated with phenotypic expansion. Our work demonstrates that a prime editing N1 screen may be a useful strategy for quickly testing pathogenicity of heterozygous *de novo* variants that cause severe pediatric phenotypes in mice.

Session Title: Mendelian Phenotypes Poster Session III

PB4937 † Utilizing multiplexed assays of variant effects to solve VOUS and screen for small molecule therapeutics in Alagille syndrome.

Authors:

M. Gilbert[1,2], E. Keefer-Jacques[1], T. Jadhav[1], G-W. Shaw[1], N. Valente[1], G. Matwijec[1], K. M. Loomes[1,2], R. Rajagopalan[1,2], T. J. Hayeck[1,2], N. B. Spinner[1,2]; [1]The Children's Hosp. of Philadelphia, Philadelphia, PA, [2]The Univ. of Pennsylvania, Philadelphia, PA

A major disparity in genomic diagnostics exists between the rate of obtaining sequencing data and the ability to interpret sequence variants, with nearly half of all entries submitted to ClinVar classified as variants of uncertain significance (VOUS). Scalable functional assays are essential for removing this bottleneck. We developed a high-throughput system to classify the disease-causing potential of missense variants in Jagged1 (*JAG1*), an autosomal dominant cause of the multi-system disorder, Alagille syndrome (ALGS). Pathogenic variants in *JAG1*, a Notch signaling ligand, are the primary cause of ALGS. Although functional haploinsufficiency is the established disease mechanism, up to 15% of *JAG1* variants are missense, with only a subset (14%) mechanistically understood. These variants show defects in protein folding, membrane localization, and ability to bind the Notch signaling receptor, NOTCH2. Diagnostic rates vary depending on reporting center and clinical presentation, with VOUS rates as high as 51% for patients with broad cholestatic phenotypes. Additionally, over 400 *JAG1* missense variants are reported in ≤5 individuals in gnomAD, confusing diagnostics. We have developed Multiplexed Assays of Variant Effects (MAVEs) as a scalable, high-throughput assay to allow for the simultaneous study of thousands of *JAG1* variants. We created a library containing every nucleotide permutation across the NOTCH2-binding domain of *JAG1*(exons 1-7). Prior studies of variants in this region show that most result in proteins that fail to reach the cell membrane. Using flow cytometry, we measured the cellular localization of JAG1 as a readout of protein function, sorting our cells into two populations representing membrane (WT-like) or intracellular (mutant-like) expression. Using a pool of >600 synonymous variants included within our library we employed an established variant scoring system to translate raw counts into benign or pathogenic predictions on protein function. Up to 17% of previously uncharacterized variants (373/2144) were scored as having defective cellular localization and validation is in progress using patient samples. Further characterization, such as analysis of NOTCH2 binding ability, is planned to resolve the remaining uncertain variants. Additionally, we show that this assay can be repurposed to characterize mutants for their ability to respond to small molecule treatment. Application of this approach to a disease with a well-characterized genetic etiology efficiently reduces the VOUS burden and provides true diagnostic value while offering a unique setting to test the therapeutic effect of small molecules.

Session Title: Mendelian Phenotypes Poster Session I

PB4938 Variant pathogenicity - support from companion animal genomes

Authors:

L. Lyons; Univ. of Missouri, Columbia, MO

A limiting factor of precision medicine is variants of uncertain significance. Whole genome and whole exome sequencing can be used to rapidly identify candidate variants for maladies and diseases, however, determining the pathogenicity of candidate variants is tedious. For Mendelian disorders, the American College of Medical Genetics and Genomics have developed criteria to assist the interpretation of sequence variants, providing a process and recommendations for the use of specific standard terminology: 'pathogenic', 'likely pathogenic', 'uncertain significance', 'likely benign', and 'benign' to describe variants. Genetic and genomic datasets from diverse animal species have become more extensive, and now include thousands of genomes and millions of variants, particularly as part of variant discovery for diseases and traits. Hence, comparative genetic approaches may be more useful in supporting variant classifications. Commercial genetic testing laboratories for domestic animals are mainly direct - to - consumer and more efforts are being explored to maintain standards, similarly to those established for humans, to interpret sequence variants for diseases and traits. Similar nomenclature for gene names and the unambiguous designation of a variant as set by the Human Genome Variation Society are being employed by domestic cats and other species to help translate animal data to homologous loci and variants in humans. Commonly, reports of variants in domestic cats and other animal species are supported by the same *in silico* approaches, segregation analyses and population datasets as suggested by the established criteria. In addition, variants in animals are usually supported by the conservation of the variant and or amino acid across species, as well as indications if the variant is pathogenic in humans and present in human population datasets. The Online Mendelian Inheritance in Animals curates the variants in animals, other then primates and rodent species. Over 170 variants are associated with diseases and traits is cats, overall, more than 1500 for diverse species. Often, innocuous, and desired aesthetic traits in a cat, such as curly fur, can be in the same gene causing a severe ectodermal dysplasia in humans. The cat research community is reviewing each of the known cat variants to provide interpretations towards pathogenicity following the human standards. Examples of new variants for Duchene muscular dystrophy, Ehlers - Danlos syndrome and dwarfism that have been discovered in the domestic cat will be presented, which may lead to new opportunities in defining undiagnosed patients and variants of uncertain significance.

Session Title: Mendelian Phenotypes Poster Session II

PB4939 Variant specific phenotypes for *SLC6A1*-related disorders.

Authors:

K. Jay[1,2], J. C. Andrews[1,2], R. German[1,2], S. V. Jangam[1,2], K. Pham[1], The National Brain Gene Registry, H. Dierick[1], H. Bellen[1,2], M. Wangler[1,2]; [1]Baylor Coll. of Med., Houston, TX, [2]Jan and Dan Duncan Neurological Res. Inst., Houston, TX

The *solute carrier family 6, member 1 (SLC6A1)* gene encodes the gamma aminobutyric acid (GABA) transporter type 1 (GAT1) protein and contributes to maintaining homeostasis of excitatory and inhibitory neurotransmitters. GAT1 mediates the reuptake of the inhibitory neurotransmitter GABA at the synapse which is essential to protect from seizure activity. *SLC6A1* has recently been implicated in a spectrum of neurological disorders. The *SLC6A1* clinical phenotype presents as a dominant disorder but is pleiotropic and shows variable expressivity of symptoms including developmental delay, epilepsy, intellectual disability, motor dysfunction, and autism spectrum disorder. Seizure types are also highly variable, ranging from absence seizures to myoclonic atonic seizures that are resistant to standard therapeutics. Furthermore, GAT1 is a potential therapeutic target for other forms of epilepsy. In this study, we hypothesize that genotype phenotype correlation underlies the variability of *SLC6A1* disease. We recruited 15 patients through the *SLC6A1 Connect Foundation*, Undiagnosed Diseases Network, and National Brain Gene Registry with rare *de novo* variants in *SLC6A1*. Importantly, seizures were reported in 9/10 patients and the average number of therapeutics attempted was 4.6 per patient. To evaluate variant specific effects *in vivo*, I created patient specific personalized constructs and inserted them into the *Drosophila* genome, effectively replacing the fly homolog with a variant form of the gene. This approach results in genome editing at the endogenous locus and allows us to evaluate dominant heterozygous variants. We quantified variant specific effects in flies using electroretinogram recordings and bang sensitivity assays, which are analogous to seizure activity in humans. The expression of several variants resulted in deficits in “on” and “off” transient signaling and photoreceptor depolarization in electroretinogram recordings. We observed bang sensitivity in only a single variant, suggesting electrophysiological signaling changes can be detected prior to seizure phenotype onset. Our short-term goal is to evaluate phenotypes associated with each variant and correlate this with our patient phenotypes. Our long-term goal is to apply these personalized patient specific models to predict variant response to therapeutics. Identifying therapeutics that will be most effective in suppressing seizure phenotypes by genotype will eliminate several rounds of trial and error for the patients and result in improved outcome for those living with *SLC6A1*-related disorders.

Session Title: Mendelian Phenotypes Poster Session III

PB4940 Variants in *EFCAB7* underlie nonsydromic postaxial polydactyly.

Authors:

T. Bharadwaj[1], M. Bilal[2,3], H. Khan[4,1], M. Javed Khan[5], T. B. Haack[3], R. Buchert[3], K. Liaqat[1], K. Ullah[2], S. Ahmed[5], A. Acharya[1], S. Peralta[3], N. Najumuddin[6], H. Ali[4], M. S. Hasni[5], I. Schrauwen[1], A. Ullah[7], W. Ahmad[2], S. M. Leal[1,8]; [1]Ctr. for Statistical Genetics, Gertrude H. Sergievsky Ctr., and the Dept. of Neurology, Columbia Univ., New York, NY, [2]Dept. of Biochemistry, Faculty of Biological Sci., Quaid-i-Azam Univ., Islamabad, Pakistan, [3]Inst. for Med. Genetics and Applied Genomics, Univ. of Tübingen, Tübingen, Germany, [4]Dept. of BioSci.s, COMSATS Univ., Islamabad, Pakistan, [5]Inst. of Biochemistry, Univ. of Balochistan, Quetta, Pakistan, [6]Dept. of Bioinformatics, Quaid-i-Azam Univ., Islamabad, Pakistan, [7]Novo Nordisk Fndn. Ctr. for Basic Metabolic Res., Faculty of Hlth.and Med. Sci., Univ. of Copenhagen, Copenhagen, Denmark, [8]Taub Inst. for Alzheimer's Disease and the Aging Brain, Columbia Univ. Med. Ctr., New York, NY

Polydactyly is the most common limb malformation that occurs in 1.6-10.6 per one thousand live births, with incidence varying with ancestry. The underlying gene has been identified, for many of the ~100 syndromes that include polydactyly. While for the more common form, nonsydromic polydactyly, eleven candidate genes have been reported. We investigated the underlying genetic cause of autosomal recessive nonsyndromic postaxial polydactyly in four consanguineous Pakistani families. Some family members with postaxial polydactyly also presented with syndactyly, camptodactyly, or clinodactyly. Analysis of the exome sequence data revealed two novel homozygous frameshift deletions in EFCAB7: [c.830delG; p.(Gly277Valfs*5); in three families and [c.1350_1351delGA;p.(Asn451Phefs*2)] in one family. Sanger sequencing confirmed that these variants segregated with postaxial polydactyly, i.e., homozygous in the two to three members with postaxial polydactyly and heterozygous or wild type in the unaffected members. EFCAB7 displays expression in the skeletal muscle and on the cellular level in cilia. IQCE-EFCAB7 and EVC-EVC2 are part of the heterotetramer EvC complex, which is a positive regulator of the Hedgehog pathway, that plays a key role in limb formation. Depletion of either EFCAB7 or IQCE inhibits induction of Glis1, a direct Hedgehog target gene. Variants in IQCE and GLI1 have been shown to cause nonsyndromic postaxial polydactyly, while variants in EVC and EVC2 underlie Ellis van Creveld and Weyers syndromes, which include postaxial polydactyly as a phenotype. This is the first report of the involvement of EFCAB7 in limb anomalies.

Session Title: Mendelian Phenotypes Poster Session I

PB4941 Variants of uncertain significance: Characterization of a cohort from the Brain Gene Registry

Authors:

A. Sveden[1], J. M. Savatt[2], E. Riggs[3], A. Morgan[4], S. Mahida[5], V. Lanzotti[6], I. Oh[7], A. Gupta[8], H. Hazlett[9], D. Baldridge[10], M. Sahin[1], C. Gurnett[11], J. Constantino[12], P. Payne[7], M. Chopra[5], Brain Gene Registry Consortium; [1]Boston Children's Hosp., Boston, MA, [2]Geisinger, Danville, PA, [3]Geisinger, Huntsville, AL, [4]Geisinger, Kingston, PA, [5]Boston Children s Hosp., Boston, MA, [6]Washington Univ. Sch. Of Med., St. Louis, MO, [7]Washington Univ. in St. Louis, St. Louis, MO, [8]Washington Univ. in St Louis, St. Louis, MO, [9]Univ. of North Carolina, Chapel Hill, NC, [10]Washington Univ. in St. Louis, Saint Louis, MO, [11]Washington Univ. in St Louis, Saint Louis, MO, [12]Children's Hosp. of Atlanta, Atlanta, GA

As the utilization of genetic testing for neurodevelopmental disorders (NDDs) has expanded, the return of variants of uncertain significance (VUS) has increased. VUS can pose emotional challenges for patients and prove difficult to interpret for healthcare providers hoping to tailor care to a genetic diagnosis. VUS in NDDs present particular interpretation difficulties given the broad, and often non-specific, spectrum of neurocognitive manifestations for many monogenic disorders, lack of complete phenotyping in published cohorts and rapid pace of gene discovery. The Brain Gene Registry (BGR) is a consortium of 13 Intellectual and Disability Research Centers (IDDRC's) that was established to address these challenges, creating a repository of paired genotypic and phenotypic data derived from a standardized neurobehavioral battery and the electronic health record. Uniquely, the BGR accepts participants with clinical reports demonstrating VUS in genes implicated in NDDs, in addition to variants classified as pathogenic/likely pathogenic (P/LP) regardless of clinical presentation.

Here, we characterize the first 91 VUS from individuals co-enrolled in the BGR and the Clinical Genome Resource (ClinGen) patient registry, GenomeConnect. Of the 91 VUS, the majority had unknown inheritance (52%). The most common variant type was missense (73%), followed by copy number (9%), intronic non-canonical splice site (7%), truncating (5%), in-frame-indel (3%), and synonymous with potential splicing impact (3%). Variants were primarily identified through exome sequencing (44%) followed by panel testing (29%) and genome sequencing (14%). There were 65 genes represented, of which 35 (38%) had never been evaluated by a ClinGen NDD Gene Curation Expert panel, and 13 (14%) had a Gene-Disease Validity classification of Limited or Moderate. The proportion of Black/African/African American participants in the VUS cohort (11%) was over fivefold that of the Black/African/African American participants in the cohort with P/LP variants (2%).

Our data suggest the importance of family member testing for resolution of inheritance and the need for improved training, tools and assays for interpretation of missense variants. These data also highlight how the underrepresentation of Black/African/African American in genomic studies results in increased uncertain results. Further, we demonstrate the need for more evidence-backed gene curations for NDD-related genes. Expert review of clinically ascertained genotypic and phenotypic data from the BGR will enrich and accelerate the resolution of VUS.

Session Title: Mendelian Phenotypes Poster Session II

PB4942 Vitamin B12-deficient mice lacking the transcobalamin-vitamin B12 receptor, CD320, exhibit modest changes in balance, locomotor activity, peripheral sensation and vision.

Authors:

D. Bernard[1], M. L. Esposito[1], S. P. Bradley[2], D. J. Walsh[1], L. C. Brody[3]; [1]NHGRI, NIH, Bethesda, MD, [2]NIMH, NIH, Bethesda, MD, [3]NHGRI (NIH), Bethesda, MD

Background: Prolonged cobalamin deficiency in humans results in peripheral neuropathy. Cognitive decline has also been associated with deficiency suggesting an impact on CNS function. Previously, we established a mouse model of severe cobalamin deficiency via a targeted deletion of the transcobalamin receptor (encoded by the *Cd320* gene) and dietary restriction of vitamin B12 (1). After 10-12 months on a vitamin B12-deficient diet, Cd320 $^{-/-}$ mice develop severe macrocytic anemia and succumb without exhibiting overt peripheral or CNS neuropathology. We have extended the phenotyping analysis our mouse model of vitamin B12-deficiency through physical and behavioral experiments aimed at testing for subtle neuropathology prior to the onset of anemia. **Methods:** *Cd320* $^{-/-}$ mice and controls were evaluated at 37-51 weeks old (mean of 42.6 weeks) after being continuously fed a vitamin B12-deficient or replete diet starting at 3 weeks old. We subjected these mice to a battery of behavioral (novel object recognition, open field, and balance beam) and physical (hot plate and electroretinogram) assays. All tests were performed prior to the onset of macrocytic anemia. **Results:** *Cd320* mutant mice fed vitamin B12-deficient diet do not differ in their response to novel objects as compared to control isogenic mice. The gene-nutrition interaction in our *Cd320* $^{-/-}$ mice fed a vitamin B12-deficient diet has a negative effect on their ability to traverse a balance beam, as measured by increased falls, perhaps as a result of impaired proprioception. In support of this, *Cd320* $^{-/-}$ mice fed the vitamin B12-deficient diet exhibit a delayed response to peripheral footpad sensations (hot plate assay, $p = 0.03$). Open field testing indicates mild reductions in vertical and horizontal movements. Electroretinogram (ERG) assays, measuring the retinal response to light flashes, indicate that the *Cd320* $^{-/-}$ mice fed vitamin B12-deficient diet have an impaired retinal response ($p < 0.001$). *Cd320* $^{-/-}$ mice fed a replete diet (50µg vitamin B12/kg of diet) do not demonstrate abnormal neuropathology with exceptions in the balance beam assay, where these mice lose balance more frequently than controls ($p < 0.03$), and in the open field testing where they travel less distance and have fewer bout of rearing than controls. **Conclusions:** Dietary vitamin B12 restriction in *Cd320* null mice results in mildly impaired performance in open field, balance beam, and hot plate assays indicating a subtle neuropathology. Furthermore, these mice have reduced ERG responses indicating impaired retinal responses.

Session Title: Mendelian Phenotypes Poster Session III

PB4943 WGS reveals secondary modifiers contribute to parkinsonism in patients with *GBA1* variants.

Authors:

N. Tayebi; NIH/NHGRI, Bethesda, MD

Gaucher disease (GD) is an autosomal recessively inherited disorder resulting from biallelic mutations in *GBA1*, encoding the lysosomal enzyme glucocerebrosidase. In addition, variants in *GBA1* are the most common known genetic risk factor for Parkinson disease (PD), although most patients with GD do not develop PD. Since the frequency of parkinsonism among *GBA1* carriers is low, other genetic or non-genetic factors must work in concert to cause the parkinsonian symptoms in these patients. To identify genetic modifiers leading to or protecting against the development of PD, a cohort of 96 patients (92 GD - with or without PD and four *GBA1* carriers) was evaluated by short-read whole genome sequencing for insertions and deletions (SNP) as well as structural variants (SV). Reads were aligned to hg19 using bwa, while SNPs were called via GATK and SVs by CNVnator. Variants were annotated using ANNOVAR. Of the 96, 28 had both GD and PD (GD/PD), and 26 had a family history of PD. Initially, we focused on eight families with discordant sibling pairs, where one sibling has GD and PD, while the other one has GD alone. We then extended the evaluation to the entire cohort with and without PD. Using minimal differences between the number of patients with GD and GD/PD affected by a specific variant (>25% for sibling pairs and >40% in the rest of the cohort), we detected 41 candidate modifiers. These genes and variants are being explored through functional and pathway analyses. The chromosome 1q21 locus is complicated by the presence of the two pseudogenes, *GBAP1* and *MTX1P1*, and their intergenic region, which both increase the chance of creating complex recombinant alleles. More than twenty different recombinant *GBA1*-associated alleles have been described and generate SV, including deletions, conversions, inversions, and duplications of contiguous genes and pseudogenes, which may function as disease modifiers. Among our cohort, four carried a recombinant allele validated with Sanger sequencing, two Rec*NciI* (1 carrier and 1 GD/PD), one *RecTL*+55bp (GD/PD), and one Rec7 (GD/PD). Two additional cases (1 carrier and 1 GD) with increased copy numbers were identified, which could indicate possible recombination events. We are currently using long-read sequencing to further validate these recombination events and to investigate whether the resulting disruptions may act as modifiers in our GD/PD cases, which could provide insight into why only a subset of patients with GD develop PD.

Session Title: Mendelian Phenotypes Poster Session I

PB4944 Whole exome and mitochondrial genome sequencing identified heterozygous variants in PLOD1, SLC22A5, and SDHD in a Syrian family with a history of sudden unexplained deaths.

Authors:

J. Silas[1], C. Noushad[1], S. Dineshkumar[1], L. Loganathan[2]; [1]Micro Hlth.Lab., Doha, Qatar, [2]Micro Hlth.Lab., Kozhikode, India

Sudden unexplained death (SUD) refers to cases of sudden death where no cardiac or extracardiac underlying cause is identified during autopsy. In less than one-third of SUD cases, standard genetic testing following guidelines can identify a disease-causing mutation. In this study, a consanguineous Syrian couple experienced the loss of two children to SUD. Additionally, the family had a history of five miscarriages, and their third child died on the 14th day after birth due to a cardiac arrest caused by septic shock. Laboratory reports indicated elevated levels of pyruvate, lactate, ammonia, and low blood sugar in the infant, suggesting the presence of a metabolic disease. However, a definitive diagnosis could not be established. Therefore, we conducted whole exome and mitochondrial genome sequencing on samples obtained from the parents with the aim of identifying the carrier status of pathogenic genes responsible for SUD. Here, we report that WES and mitochondrial genome sequencing analysis identified the father and mother as carriers of a heterozygous pathogenic variant in the PLOD1 gene (p.Asp372AlafsTer19), a heterozygous likely-pathogenic variant in the SLC22A5 gene (p.Tyr124Cys), and a heterozygous variant of uncertain significance in the SDHD gene (p.Arg17Gln). The PLOD1, SLC22A5, and SDHD genes have been reported to be associated with SUD. Our data strongly supports the inclusion of WES and mitochondrial sequencing in the evaluation of SUD.

Session Title: Mendelian Phenotypes Poster Session II

PB4945 Whole exome sequence analysis reveals candidate disease genes for lichen sclerosus.

Authors:

K. Lovrenert[1,2], I. Patel[1,2], A. Tomaszewski[3], A. Myers[2], J. Marino[2], F. R. Schumacher[4], J. Eastman[5], M. Drumm[5], C. Farr[6], R. Pope[7], C-H. Wu[1,2]; [1]Dept. of Genetics and Genome Sci., Case Western Reserve Univ. and Univ. Hosp., Cleveland, OH, [2]Dept. of Urology, Case Western Reserve Univ. and Univ. Hosp., Cleveland, OH, [3]Case Western Reserve Univ. Sch. of Med., Univ. Hosp., Cleveland, OH, [4]Dept. of Population and Quantitative Hlth.Sci., Case Western Reserve Univ. Sch. of Med., Cleveland, OH, [5]Dept. of Genetics and Genome Sci., Case Western Reserve Univ., Cleveland, OH, [6]Ctr. for Global Hlth.and Diseases, Dept. of Pathology, Case Western Reserve Univ., Cleveland, OH, [7]Urology Inst., Univ. Hosp. Cleveland Med. Ctr., Case Western Reserve Univ., Cleveland, OH

Lichen sclerosus (LS) is a chronic inflammatory mucocutaneous dermatosis most commonly affecting anogenital skin, leading to recalcitrant itching, scarring, sexual dysfunction, and increased risk of local skin malignancy. LS onset can occur at any age in both sexes, but presents a bimodal distribution in incidence, primarily affecting pre-pubertal children and middle-aged adults, and typically affecting more females than males. Accumulating evidence suggests a positive association between LS and autoimmune disease in women, however the exact pathogenesis of LS remains unknown. Here we performed a whole exome sequencing-based gene burden analysis using VAAST2 to profile potential pathogenic variants associated with LS using both a local female cohort of 17 biopsy-proven LS cases and 3 healthy controls and data from the 1000 Genomes Project. We identified candidate variants in 12 genes, all affecting the protein coding regions, with the highest genetic burden among genes regulating HLA class I and II antigens (*HLA-C*, *HLA-DRB5*), those involved in JNK or NF-KappaB signaling (*MAP3K1*, *MAP3K4*), rRNA processing (*KRI1*, *RPL14*), and those associated with tumor suppression (*HNF1A*) and cilium motility (*CFAP251*). Gene ontology analysis of the affected gene set yields “antigen processing and presentation”, “MAP kinase kinase kinase activity”, “MHC protein complex”, and “luminal side of endoplasmic reticulum membrane” as the most significantly enriched terms, while pathway analysis against KEGG and Reactome reveals “interferon-gamma signaling” and “allograft rejection” as top hits. We find that the gene cluster, *HLA-C*/*HLA-DRB5*, is consistently enriched across the most significant ontology and pathway terms. Our results suggest that an enhanced mutational burden in gene sets regulating key immunologic pathways could be associated with LS pathogenesis in women.

Session Title: Mendelian Phenotypes Poster Session III

PB4946 Whole exome sequencing of families from Senegal reveals known and candidate hearing impairment genes

Authors:

Y. Dia[1], R. Ndiaye Diallo[1], I. Schrauwen[2], T. M. Bharadwaj[2], S. M. Adabey[3], E. T. Abogye[3], K. Popel[3], J. P. D. Diop[1], M. Jonas[3], S. S. Mbacké[1], C. De Kock[3], S. A. Ba[1], P. D. Sarr[1], A. R. G. Sène[1], C. A. T. Ly[1], B. K. Diallo[1], S. M. Leal[2], A. Wonkam[4]; [1]Univ. Cheikh Anta Diop, Dakar, Senegal, [2]Columbia Univ., New York, NY, [3]Univ. of Cape Town, Cape Town, South Africa, [4]Dept. of Genetic Med., JHU, Baltimore, MD

Hearing impairment (HI) remains the most disabling sensory deficit worldwide. Without early diagnosis and treatment, HI impedes language acquisition and cognitive development. 60% of congenital HI have a genetic etiology, of which 70% are non-syndromic. *GJB2* is the major cause of non-syndromic HI (NSHI) in Europeans and Asians. However, its contribution to HI in most sub-Saharan African populations is close to zero. This study aimed to investigate the genetic etiology of HI in Senegal. 157 hearing impaired individuals belonging to 66 families with suspected genetic origin were recruited from 13/14 administrative regions of Senegal. Careful clinical examination was performed, and hearing threshold was assessed by pure tone audiometry or auditory brainstem response. Families segregating autosomal recessive NSHI (ARNSHI) were screened for variants in *GJB2* exon2. Furthermore, Whole Exome Sequencing (WES) was used to identify other HI genes involved in *GJB2*-negative families. Pedigree analysis suggested autosomal recessive inheritance in 48 families. Consanguinity rate was estimated at 70%. Clinically, 59 families exhibited NSHI and 7 families showed SHI. The majority (151/157 patients) displayed prelingual HI. Of the 44 multiplex families segregating ARNSHI, *GJB2* pathogenic variants were identified in 15 families. *GJB2*: c.94C>T was the most frequent variant (11/15 families) and may have a founder effect in Senegal. 51 *GJB2*-negative families underwent WES. We identified variants in 15 known NSHI genes [21/51 families (41.2%)], 5 genes associated with either SHI or NSHI [7/51 families (13.7%)] and 6 SHI genes [6/51 families (11.9%)]. Variants in *CLDN14* and *CIB2* contributed the most to HI [7/51 families (13.7%)]. *CIB2*: c.409C>T p.(Arg137Trp) was the most prevalent variant among *GJB2* negative families and was identified in 3 unrelated consanguineous families with the same ethnic background (Serer). Post lingual HI associated with type 1 diabetes was linked to a variant in *LOXDH1*. *KCNE1,* known to cause Jervell Lange-Nielsen syndrome segregates with profound HI associated with a normal QT interval in a family. Digenic inheritance was suspected for *GJB2/USH2A* in 2 families. We also identified, 3 novel candidate genes associated with NSHI (*OPA3*, *DDX10* and *FST*). Half of the genetic variants identified in this study have not been previously associated with HI (25/50). 10 families were unsolved by WES and will go for Whole Genome Sequencing.This study showed that *GJB2* is a major cause of NSHI in Senegal. WES reveals locus heterogeneity despite a high consanguinity rate, and suggests discovery of novel genes which will foster our understanding of HI biopathology.

Session Title: Mendelian Phenotypes Poster Session I

PB4947 Whole exome sequencing reveals known and novel variants in two Malians families with neuronal ceroid lipofuscinosis

Authors:

O. Traoré[1], D. COULIBALY[2], Z. Kone[1], S. Bamba[3], A. Maiga[4], A. Yalcouyé[5], A. Toure[1], M. Kotioumbé[1], L. Jeffries[6], J. Weizhen[6], S. Diarra[3], S. Lakhani[6], G. Landoure[7]; [1]Univ. of Sci. Techniques et Technologies of Bamako, USTTB, Bamako, Mali, [2]Ctr. Hosp.ier Univ.ire Mère enfant le Luxambourg, Bamako, Mali, [3]USTTB, Bamako, Mali, [4]Univ. of Sci., Techniques and Technologie, Bamako, Mali, [5]USTTB, Mali, Bamako, Mali, [6]Yale Univ., New Haven, CT, [7]CHU du Point G, Bamako, Mali

Introduction: Neuronal ceroid lipofuscinoses (NCL) are a group of severe neurodegenerative disorders, characterized by accumulation of autofluorescent ceroid lipopigments in cells with a global incidence of 1:100†000. Common clinical features in NCL include retinopathy, motor abnormalities, epilepsy and dementia. To date, eight genes are reported to cause 13 subtypes of NCL. Here, we report the case of two consanguineous Malian families with a novel variant in two genes known to cause NCL. Aim: To clinically characterize patients with NCLs and identify the underlying genetic defects. Methods: After full consent, patients were thoroughly examined by Neurologists. Electroencephalogram (EEG) and brain imaging were performed. DNA was extracted from peripheral blood for whole exome sequencing (WES), in trios. Segregation analysis was done in all available family members. In addition, in silico analysis was done for deleteriousness. Results: Two boys aged 7 and 9 respectively, and their healthy consanguineous parents, were enrolled. Disease started at the age of 5 and 6 by generalized seizure. The average age of onset was 5 years. Clinical examination found psychomotor delay and visual loss, and EEG found an abnormal electrical pattern. Brain imaging showed cortical and subcortical atrophy in both patients. WES identified a known stop gained variant (c.622C>T; p. Arg208Ter) in *TPP1* gene in family 1 and a novel missense variant (c.G784C:p.Asp262His) in *CLN8* gene in family 2. These variants segregate with the phenotypes in the respective families and were shown to be deleterious by several in silico prediction tools, including CADD: 36 and 27, respectively. These genes are known to cause NCL type 1 and type 8, respectively. Functional studies are needed to further assess the pathogenicity of the novel variant in *CLN8* gene, likely using cell and/or frog models. Conclusion: This study reports the first molecularly diagnosed cases of ceroid lipofuscinosis in the Malian population, using WES. It expands the genetic spectrum of NCL and reveals the need of including understudied populations in genetic studies for the global benefit. It also prepares our families for future clinical trials.

Session Title: Mendelian Phenotypes Poster Session II

PB4948 Whole Exome Sequencing Uncovers the Genetic Complexity of Bicuspid Aortic Valve in Families with Early Onset Complications

Authors:

S. Prakash[1], S. Mansoorshahi[1], A. Yetman[2], M. M. Bissell[3], Y. Y. Kim[4], H. Michelena[5], D. S. Hui[6], A. Caffarelli[7], M. G. Andreassi[8], I. Foffa[8], D. Guo[1], R. Citro[9], M. De Marco[10], J. Tretter[11], S. A. Morris[12], S. C. Body[13], J. X. Chong[14], M. J. Bamshad[14], University of Washington Center For Rare Disease Research, EBAV Investigators; [1]Univ. of Texas Hlth.Sci. Ctr. at Houston, Houston, TX, [2]Univ. of Nebraska Med. Ctr., Omaha, NE, [3]Univ. of Leeds, Leeds, United Kingdom, [4]Univ. of Pennsylvania, Philadelphia, PA, [5]Mayo Clinic, Rochester, MN, [6]Univ. of Texas Hlth.Sci. Ctr. at San Antonio, San Antonio, TX, [7]Hoag Mem. Hosp. Presbyterian, Newport Beach, CA, [8]CNR Inst. of Clinical Physiology, Pisa, Italy, [9]Univ. Hosp. Salerno, Salerno, Italy, [10]Univ. Degli Studi Di Salerno, Salerno, Italy, [11]Cleveland Clinic, Cleveland, OH, [12]Baylor Coll. of Med., Houston, TX, [13]Boston Univ., Dover, MA, [14]Univ. of Washington, Seattle, WA

Introduction: Bicuspid aortic valve (BAV), the most common adult congenital heart malformation, causes aortic disease requiring valve replacement and thoracic aortic aneurysms predisposing to acute aortic dissections. The spectrum of BAV ranges from serious valve or aortic complications in childhood to incidentally discovered late onset disease. We hypothesized that BAV cases with early onset complications requiring intervention prior to age 30 (EBAV) may be enriched for rare, highly penetrant, causal genetic variants that predict valve or aortic outcomes.
Methods: We analyzed whole exome sequences of 350 individuals in 215 EBAV families, including 34 trios, using seqr (10.1002/humu.24366). The mean age of probands was 18 years. Individuals with known mutations or syndromes were excluded. We selected high quality, rare ($<1x10^{-4}$ in gnomAD and TOPMed) loss of function, missense, or indel variants with *de novo* or autosomal dominant inheritance. Missense variants and indels were additionally filtered to select variants with REVEL > 0.5, CADD > 20, and damaging by at least 3 of SIFT, Polyphen, Mutation Taster, or FATHMM. We prioritized candidate genes that are implicated in dominant phenotypes and were mutated in at least three EBAV families, with validation by Sanger sequencing.
Results: We identified predicted pathogenic variants of causal genes in 47 families (22%), including genes that cause BAV (9, 4%) or thoracic aortic disease (TAD, 22, 10%), and dominant congenital heart disease genes (24, 11%). The most frequently mutated BAV and TAD genes were FBN2 (5), COL1A2 (4), MIB1 (3), EVC2 (3), and SMAD6 (3). We did not observe any variants in GATA genes or other TGF-beta pathway genes that passed our filters. TTN (9), MYH6 (8), KCNH2 (6) and PKP2 (4) were the most frequently mutated CHD genes. We observed loss of function variants in PKP2 (4), SMAD6 (3), KCNH2 (1), EVC2 (1), and MYH6 (1). MIB1, TTN, and KCNH2 variants segregated in EBAV families.
Conclusion: Our findings highlight the genetic heterogeneity of BAV, even with similar clinical presentations, and implicate alterations of novel candidate genes in the pathogenesis of BAV.

Session Title: Mendelian Phenotypes Poster Session III

PB4949 Whole genome sequencing provides high diagnostic rate and improves patient care for under-served adults with intellectual disability

Authors:

A. Sabo[1], P. Trotter[1], H. Gu[1], S. Dugan[1], M. Gingras[2], D. Murdock[3], B. Yuan[1], K. Glinton[1], S. Pereira[2], R. Gibbs[2]; [1]Baylor Coll. of Med., Houston, TX, [2]Baylor Coll. Med., Houston, TX, [3]The Univ. of Texas Hlth.Sci. Ctr. at Houston, Houston, TX

Background: Establishing a genetic diagnosis for individuals with intellectual disability (ID) has numerous benefits for patients and their families. It can provide valuable information about prognosis, guide appropriate therapy and monitoring, and facilitate access to supportive services. Although exome and genome sequencing have become part of routine care for diagnosing pediatric patients, there is a gap in the genetic evaluation of adults with ID, particularly among families who are unable to access these tests due to socioeconomic position, lack of insurance coverage or lack of awareness regarding new testing opportunities.
Methods: Our recruitment strategy involved partnering with service and education providers and clinical sites. We employed whole genome sequencing (WGS) and analysis to identify SNVs, indels, and CNVs, and performed clinical variant interpretation according to ACMG standards. Where WGS analysis identified pathogenic variants, the results were discussed with a board-certified clinical geneticist, prior to further medical follow-up. We formally evaluated clinical utility related to diagnosis, patient clinical management, and familial and psychosocial implications, using the C-GUIDE questionnaire.
Results: We established a molecular diagnosis in 39% (14/36) of individuals. Our results include de novo variants in MED13L, POGZ, BRAF, EHMT1, SATB2, and DNMT3A, homozygous or compound heterozygous variants in TUSC3 and VRK1, and heterozygous variants with unknown inheritance in KANSL1, SATB1, ASH1L, CHD4, AHDC1 and TBL1XR1. Clinicians completed a C-GUIDE questionnaire for nine individuals. The mean score of utility of the genetic testing was 17.1 (SD=3.6; on a scale of -2 - 32, higher scores indicating more utility). When asked whether the genetic testing prompted better care for the patient or their family, clinicians reported it prompted better care or somewhat better care in 78% of cases (7/9). Clinicians also indicated that no psychosocial concern was experienced by the patient or family in eight cases.
Conclusion: Our study indicates that WGS is a valuable diagnostic tool for adult patients with ID who have not received comprehensive sequencing based genetic testing. Furthermore, it demonstrates that WGS has a high level of clinical utility and minimal psychosocial concerns for patients and their families. Research studies like ours, which include assessment of clinical utility, could play a crucial role in evaluating the practical value of genetic testing in cases where medical insurance providers rarely cover the costs of genome sequencing due to perceived limited clinical utility.

Session Title: Mendelian Phenotypes Poster Session I

PB4950 Whole genome sequencing study of consanguineous Pakistani families reveals a set of potentially causative variants in various male infertility phenotypes.

Authors:

A. Akbari[1], M. Khan[2], A. Shah[2], D. Conrad[1]; [1]Oregon Hlth.& Sci. Univ., Portland, OR, [2]Univ. of Malakand, Malakand, Pakistan

Introduction:Last decade witnessed a revolution in studies of human genetic diseases with Mendelian inheritance as a result of emergence and wide use of High-Throughput Sequencing technologies. However, the most widely used method of targeted sequencing has had limited success in discovering large genetic changes due to the intermittent coverage of the genome. In this study, we recruited a set of seven consanguineous families with severe male infertility phenotypes including non-obstructive azoospermia (NOA), oligozoospermia and asthenozoospermia for whole genome sequencing and functional validation assays. Each family includes at least two infertile siblings and, in some cases, a fertile brother which makes for a suitable case study.Methods:Genomic DNA of a total of 26 samples from 7 families were subjected to whole genome sequencing on illumina NovaSeq. The resulting FASTQ files were checked for quality by FASTQC 0.12.0. Raw reads were aligned to the human genome (hg38) using BWA 0.7.17 followed by preprocessing, sorting, marking of duplicate reads, base recalibration and short variant calling by GATK 4.4. In addition, copy number variations were called using Manta 1.6 and smoove 0.2.7. Candidate variants were visualized using IGV 2.16.Results:Three NOA brothers were homozygous for a novel 10 kb deletion in *M1AP* encompassing exon 2 of the canonical transcript whereas their fertile brother was heterozygous. In another family, again three NOA brothers were homozygous for a novel 152 bp deletion downstream to *THORLNC* - A long non-coding RNA whose knock-out mouse has been described as infertile. The unmarried brother is also homozygous for the deletion but we cannot confirm his phenotype. One NOA patient and his oligozoospermic brother were hemizygous for a novel *TKTL1* in-frame deletion. This gene is specifically expressed in spermatogonia but its role in infertility has not been thoroughly explored. In addition, two sets of two asthenozoospermic brothers were homozygous for a rare *SPAG6* and a rare *CCDC9* missense mutation, respectively. Both of these genes have been previously described in asthenozoospermic cases. In total, we managed to discover a potential causative variant in 5 families out of 7 where two of the variants were deletions bigger than 150 bp.Conclusion:This study reaffirms the power of family analysis in revealing the genetic mutations behind Mendelian diseases. Now with whole genome analyses more feasible than ever before, our knowledge about large events affecting disease causing genes both in coding and non-coding regions can be expanded.Keywords:male infertility, whole genome sequencing, CNV analysis, family study

Session Title: Cancer Poster Session I

PB4951 3DIV update for 2021: a comprehensive resource of genome and 3D cancer genome

Authors:

H. Sung, I. Jang, B. Lee, S. Sung; Korea Res. Inst. of BioSci. and Biotechnology, Daejeon, Korea, Republic of

Three-dimensional (3D) genome organization plays a critical role in regulating gene expression during various biological processes and diseases. In cancer, disruptions in the 3D genome structure caused by large-scale genomic rearrangements can lead to abnormal gene expression patterns with oncogenic implications. However, understanding the pathogenicity of the 3D cancer genome has been challenging due to limited tools capable of analyzing disorganized higher-order chromatin structures.
To address this issue, we have updated a comprehensive resource called 3DIV, which is a 3D-genome Interaction Viewer and database. This update aims to expand our understanding of the 3D cancer genome. The key enhancements of 3DIV include:
1. Expanded Data Collection: We have added 401 samples to the database, including 220 cancer cell line/tumor Hi-C data, 153 normal cell line/tissue Hi-C data, and 28 promoter capture Hi-C data. This diverse collection allows researchers to explore both normal and cancer 3D genomes.
2. Interactive Manipulation: Users can now interactively manipulate the 3D cancer genome to simulate the impact of structural variations. This feature enables researchers to investigate how genomic rearrangements affect gene regulation in a dynamic and customizable manner.
3. User-Defined Chromosome Order: We have introduced the ability to reconstruct Hi-C contact maps based on user-defined chromosome order. This feature facilitates the exploration of complex genomic rearrangements and their influence on the 3D genome.
With these updates, 3DIV serves as the most comprehensive resource for studying the gene regulatory effects of the 3D genome in both normal and cancer contexts. Researchers can freely access and utilize 3DIV at http://3div.kr. These advancements in 3DIV empower researchers to gain deeper insights into the intricate relationship between 3D genome organization and gene regulation in cancer and other biological processes.

Session Title: Cancer Poster Session II

PB4952 9p21.3 microdeletion cancer susceptibility syndrome: Case series, systematic literature review, and continued clinical challenges.

Authors:

C. Wyman, G. V. George, A. L. Schroeder, D. Korones, C-T. Fong, M. A. Iqbal, A. J. Cox; Univ. of Rochester Med. Ctr., Rochester, NY

Families with germline deletions of 9p21.3 involving the *CDKN2A* gene have been reported previously, and appear to have a cancer susceptibility syndrome overlapping *CDKN2A*-related melanoma-pancreatic cancer and melanoma-neural system tumor syndromes. Clinical similarities to Li Fraumeni syndrome and Neurofibromatosis have also been noted. Additional associated tumor types and cancer risk levels have yet to be well characterized for this multigene deletion, leaving affected families without concrete screening and management recommendations to mitigate cancer risk. Herein, we report two cases of germline 9p21.3 deletion including *CDKN2A*.
In the first case, we describe four generations of a family with a germline 783 kilobase interstitial deletion at 9p21.3. Ten family members presented with various neoplasms including ampullary adenocarcinoma, liposarcoma, astrocytoma, glioma, schwannoma, neurofibroma, sigmoid adenocarcinoma in situ, and melanocytic nevi. In the second case, we describe a pediatric patient with a germline de novo 1.7 megabase interstitial deletion at 9p21.3 who presented with a malignant histiocytic neoplasm. In both cases, following identification of *CDKN2A* gene deletion via next-generation sequencing, we performed confirmatory array comparative genomic hybridization (aCGH) to characterize the extent of the deletion.
These cases demonstrate the association between deletions at 9p21.3 encompassing *CDKN2A* with nerve sheath tumors, gliomas, melanocytic lesions, and sarcomas, and raise additional possible tumor associations including ampullary adenocarcinoma and colorectal cancer. Ultimately, we seek to provide tailored recommendations for clinical screening based on the specific family history of cancer and ages of diagnosis, while also taking into consideration prior groups' broader recommendation for the Li Fraumeni surveillance regimen. We support the ongoing incorporation of *CDKN2A* into hereditary cancer panels given its broad association with various tumor types and clinical similarity to other hereditary cancer predisposition syndromes. While the molecular diagnosis provided an explanation for the cancers in these families, the medical literature concerning germline 9p21.3 multigene deletion remains lacking to inform adequate screening recommendations and more precise characterization of tumor and cancer risks. A systematic review of the literature and characterization of two additional cases is presently underway to further define the phenotypic spectrum of this new cancer susceptibility syndrome.

Session Title: Cancer Poster Session III

PB4953 A cost effective whole genome sequencing method identifies clinically relevant germline variants in patients recruited in a breast cancer clinic.

Authors:

A. Cruickshank[1,2,3], R. Li[1,2], A. Lovato[3], C. Serruya[3], N. Palmour[4], J. St-Cyr[1,2], D. Fleiszer[5,6], Y. Joly[4], S. Meterissian[5,6], I. Ragoussis[1,2], P. Tonin[1,3,7]; [1]Dept. of Human Genetics, McGill Univ., Montreal, QC, Canada, [2]McGill Genome Ctr., McGill Univ., Montreal, QC, Canada, [3]Cancer Res. Program, The Res. Inst. of the McGill Univ. Hlth.Ctr., Montreal, QC, Canada, [4]Ctr. of Genomics and Policy, McGill Univ., Montreal, QC, Canada, [5]Dept. of Surgery, McGill Univ., Montreal, QC, Canada, [6]Dept. of Oncology, McGill Univ., Montreal, QC, Canada, [7]Dept. of Med., McGill Univ., Montreal, QC, Canada

Introduction: Whole genome sequencing (WGS) is increasingly used to capture all possible biologically and clinically relevant germline variants in the human genome from cancer patients. However, the cost, large size, and complexity of this data presents a major challenge when considering implementing WGS technology as part of routine care. To assess the efficacy of integrating WGS in oncological settings, we have investigated WGS data from the germline of at least 100 breast cancer (BC) patients recruited at a local clinic.
Materials and Methods: Consenting participants with a history of invasive BC that were recruited at a Montreal Canada hospital-based breast clinic provided blood samples for WGS and completed questionnaires. Clinical metrics were collected from medical charts. DNA was sequenced at the McGill Genome Centre using the cost effective MGI DNBSEQ-T7 technology. Data was processed using Digital Research Alliance of Canada's high performance computing clusters and the Canadian Centre for Computational Genomics pipelines. Clinically relevant variants harboured by 11 known carriers were investigated, followed by the investigation of variants in 19 clinically actionable BC susceptibility genes (BCSG) according to NCCN guidelines, using ClinVar. Genetic data was aggregated with personal clinical data and family history.
Results: In all, 74% of participants were prioritized for analysis and have completed sequencing. WGS data available from 98 participants exhibited high quality. All SNP/INDEL pathogenic variants were validated in the 11 known carriers. The investigation of 19 BCSG's identified 6 pathogenic/likely pathogenic variants in participants not previously known to be carriers based on clinical reports. A review of clinical data provided information concerning lack of *a priori* BCSG testing for some cases.
Conclusions: Results indicate that WGS can detect variants from prior genetic screens and identify new carriers of actionable variants that for whatever reason where not known to be carriers.
Funding Sources: MUHC Foundation, CIHR, FRQS Tumour Bank of the RRCancer Network, Genome Canada, Canada Foundation for Innovation, Compute Canada, McGill University Department of Human Genetics.

Session Title: Cancer Poster Session I

PB4954 A genome-first approach to characterize *DICER1*pathogenic variant prevalence, penetrance and phenotype in adults from two population-scale cohorts

Authors:

J. Kim[1], J. Haley[2], J. Hatton[1], U. Mirshahi[3], H. Rao[4], D. Smelser[4], G. Urban[4], Geisinger-Regeneron DiscovEHR Collaboration, K. Schultz[5], D. Carey[3], D. Stewart[1]; [1]Natl. Cancer Inst., Rockville, MD, [2]Geisinger, Danville, PA, [3]Geisinger Clinic, Danville, PA, [4]Geisinger Hlth.System, Danville, PA, [5]Children's Minnesota, Minneapolis, MN

DICER1 is an RNaseIII endonuclease essential for processing pre-microRNA into mature, microRNA (miRNA). Loss-of-function germline variants in *DICER1* underlie an autosomal dominant disorder, *DICER1*-related tumor predisposition, with an increased risk of numerous neoplasms, macrocephaly, and thyroid-related disease. The availability of large-scale, exome-sequenced cohorts with linked electronic health records (EHR) from the DiscovEHR study and UK Biobank cohort permits genome-first exploration of phenotype in individuals with a pathogenic germline *DICER1* variant. In this study, we investigated the prevalence of pathogenic germline *DICER1* variants in exomes of 175,503 individuals from the Geisinger DiscovEHR study and 469,787 individuals from the UK Biobank cohort. *DICER1* variants were classified using ClinGen *DICER1* variant curation expert panel (VCEP) criteria; non-carriers were defined as harboring B/LB or canonical *DICER1* varation. For the UK Biobank cohort, we queried for diagnoses using cancer and death registry, and EHR. For the DiscovEHR study, we queried the Geisinger Cancer Registry and EHR for similar diagnoses.We identified 57 individuals (1:8,119) and 77 individuals (1:8,242) with a *DICER1* pathogenic variant in the DiscovEHR study and the UK Biobank cohort, respectively. From *DICER1* heterozygotes, we identified 5 individuals with 6 cancers in DiscovEHR and 12 individuals with 14 cancers in UK Biobank, although this was not significantly enriched compared with non-carriers. Next, to investigate thyroid phenotypes, we analyzed ICD10 codes E01-E07. We found in both cohorts, E01-E07 was significantly enriched in *DICER1* heterozygotes compared to non-carriers. In addition, cumulative incidence showed statistically significant difference in age at diagnosis with thyroid phenotypes in context of DICER1 diagnosed earlier than in non-carriers. Lastly, we applied an agnostic approach and investigated whether any ICD10 code with >10% prevalence in *DICER1* heterozygotes were enriched compared with non-carriers. There were 18 and 6 ICD10 codes enriched in *DICER1* heterozygotes in DiscovEHR and UK Biobank, respectively. However, none of the phenotypes identified in DiscovEHR and UK Biobank overlapped. Of 24 ICD10 codes that were identified, after Bonferroni correction, only Z11 (encounter for screening for infectious and parasitic diseases) and R50 (Fever of other and unknown origin) were statistically significant.Here we provide more refined prevalence, thyroid phenotype penetrance, and additional phenotypes for pathogenic *DICER1* variants in two large cohorts using ClinGen *DICER1* VCEP variant classification guidelines.

Session Title: Cancer Poster Session II

PB4955 A genome-first approach to quantify cancer risk in individuals harboring *CHEK2* pathogenic truncating variants versus pathogenic missense variants in two population-scale cohorts

Authors:

M. Ramos[1], S. Kim[2], J. Kim[1], J. Haley[3], R. Shanker[3], U. Mirshahi[4], B. Graubard[1], H. Katki[1], D. Carey[4], D. Stewart[1], DiscoverEHR; [1]Natl. Cancer Inst., Rockville, MD, [2]Johns Hopkins Med. Inst./NCI, Gaithersburg, MD, [3]Geisinger, Danville, PA, [4]Geisinger Clinic, Danville, PA

Introduction: Pathogenic truncating variants (PTV) in CHEK2 are associated with increased cancer risk; in contrast, pathogenic missense variants (PMV) have more variable effects dependent on whether a critical protein domain is affected. Yet several missense variants in CHEK2 have been associated with increased risk of certain cancers in clinical reports and phenotype-first studies. Using the genome-first approach, we investigated the association of different cancer phenotypes to pathogenic/likely pathogenic (P/LP) variants in individuals who carry PTV and PMV in CHEK2. **Methods:** From the Geisinger (GHS) and UK Biobank (UKBB) cohorts, we investigated cancer phenotypes in individuals over 18 years of age (GHS: n=160,448, UKBB: n=469,802) with germline P/LP variants in CHEK2. Odds ratios measuring the effect of harboring a CHEK2 variant on cancer occurrence for both any cancer and for specific cancers were computed while controlling for current age, sex, smoking, alcohol, and BMI. Bonferroni correction was applied for multiple comparisons ($\alpha<0.05$). Kaplan-Meier survival curves were constructed, and hazard ratios were computed based on the Cox Proportional-Hazards. **Results:** There were 152,662 (GHS) and 305,330 (UKBB) controls and 3,147 (GHS) and 3,226 (UKBB) P/LP heterozygotes. In GHS, 919 (29%) of the P/LP heterozygotes were PTV and 2,227 (70.6%) were PMV. In UKBB, 1,860 (57.5%) were PTV and 1,291 (39.9%) were PMV. The proportion of individuals with P/LP variants who have any type of cancer (GHS: n=864, 27.5%, UKBB: 935, 28.9%) was significantly higher compared with controls (GHS: n=35,053, 22.3%, P-value: <0.0001, UKBB: n=69,840, 22.9%, P-value: <0.0001). From UKBB, the risks of breast, prostate and thyroid cancers were significantly increased in both PTV and PMV. Kidney and bladder cancer, Hodgkin lymphoma, non-Hodgkin lymphoma, myeloid leukemia and acute myeloid leukemia were significantly associated with PTV carriers whereas lymphoid leukemia, chronic lymphocytic leukemia and essential thrombocythemia were significantly associated with PMV. In GHS, risk of breast cancer was significantly increased in both PTV and PMV, while bladder cancer risk significantly increased for PTV and prostate cancer, thyroid cancer, and leukemia risk increased for PMV. There were no differences in overall survival rates and time-to-cancer in between PTV and PMV. Results from the two cohorts largely overlapped. **Conclusions:** The genome-first approach revealed that some of the cancer phenotypes can be different between PTV and PMV carriers which may help inform cancer surveillance and genetic counseling.

Session Title: Cancer Poster Session III

PB4956 A Linear Relationship between the Number of Cancers in the Family History of First-degree Relatives and the Risk of Multiple Primary Cancers

Authors:

S. He[1,2], K. H. Barry[2], B. D. Mitchell[2], S. Chen[2], L. B. Freeman[1], Y. Zhang[2], S. I. Berndt[1]; [1]Natl. Cancer Inst., Bethsda, MD, [2]Univ. of Maryland Sch. of Med., Baltimore, MD

With advances in the early detection and treatment of cancer, the number of cancer survivors has risen over time as well as the incidence of second cancers or multiple primary cancers (MPC) in the same individual. We tested the hypothesis that the number of cancer cases among first-degree relatives is associated with MPC risk in 139,540 participants from the Prostate, Lung, Colorectal, and Ovarian (PLCO) Cancer Screening Trial, a prospective cohort study in the U.S. Cox proportional hazard models were used to calculate hazard ratios (HR) and 95% confidence intervals (95% CI), adjusting for age, sex, self-reported race, smoking, body mass index, and other potential confounders. Overall, reporting a family history of cancer in at least one first-degree relative increased the risk of MPC by 19% (HR=1.19, 95%CI: 1.13-1.25). A positive linear trend was observed between the reported number of cancers among first-degree relatives and MPC risk with HRs (95%CI) of 1.14 (1.07, 1.21), 1.23 (1.14, 1.33), 1.33 (1.18, 1.50), to 1.44 (1.21, 1.70) for 1, 2, 3, and 4+ cancers among relatives, respectively ($P_{trend}=1.31x10^{-13}$). Although the global test for heterogeneity was not significant (P>0.05), small differences in risk were observed based on histology with hazard ratios ranging from 1.12 to 1.97 depending on the cancer subtype. No significant differences were observed by age at first cancer, race, smoking status, or body mass index. Our study demonstrates that family history of cancer among first-degree relatives is associated with an increased risk of developing multiple primary cancers and that the risk increases linearly with the number of cancers reported among first-degree relatives. Although shared environmental factors could contribute to the observed increased risk, our analysis suggests that genetic factors are likely to contribute to the risk of MPC and that the number of cancers among first-degree relatives should be taken into account when assessing the risk of a second cancer.

Session Title: Cancer Poster Session I

PB4957 A novel germline deep intronic BAP1 variant leading to aberrant gene splicing in a patient with familial uveal melanoma

Authors:

M. Abdel-Rahman[1], M. Hussein[2], L. Byrne[2], P. Johansson[3], F. Davidorf[2], N. Hayward[4], C. Cebulla[2]; [1]Ohio State Univ., Columbus, OH, [2]The Ohio State Univ., Columbus, OH, [3]QIMR Berghofer Med. Res. Inst., Brisbane, Australia, [4]QIMR Berghofer Med. Res. Inst., Herston, Australia

Objective: The goal of this study was to assess the contribution of germline non-coding genomic variants including 3' UTR, 5' UTR and deep intronic variants to the pathogenesis of BAP1-Tumor Predisposition Syndrome. Methods: A next-generation sequencing gene panel, which included the whole *BAP1* gene (9.7KB) and coding regions of 90 other known cancer genes, was used. Germline DNA from 38 subjects with personal or family histories suggestive of BAP1-TPDS was tested. The effect of the variant on splicing was assessed by RT-PCR. Results: A novel non-coding variant was identified in 1/38 patients. A deep intronic deletion, c.932-57_932-56del, in *BAP1* was detected in a male who presented with uveal melanoma at age 36 yrs. The patient reported a family history of uveal melanoma. Prior exome sequencing identified no pathogenic variants in other known cancer genes. Splicing prediction software did not predict the variant to affect splicing, however, RT-PCR showed whole exon skipping (exon 11). No other family member was available to assess segregation of the variant and the patient was treated by brachytherapy so no tumor tissue was available for assessment of biallelic inactivation. Conclusions: The study provides evidence of the importance of testing for non-coding variants in patients with strong personal or family history of BAP1-TPDS phenotype with no detectable coding variants. A reflex RNA NGS assay to rule out gene rearrangement and deep non-coding variant is highly recommended.

Session Title: Cancer Poster Session II

PB4958 A novel statistical method to identify germline genetic modifiers of driver mutations in cancer.

Authors:

S. Zhao[1], J. Zhou[1], X. He[2]; [1]Dartmouth Coll., Hanover, NH, [2]Univ Chicago, Chicago, IL

Positive selection on somatic mutations is the driving force for cancer progression. Accumulation of multiple driver mutations, which are mutations under positive selection, lead to malignant transformation of a cell. Growing evidence shows that the emergence of a driver mutation in a sample depends on individual-specific factors, in particular, the individual's germline genetic background. A hypothesis for this phenomenon is that germline background provides the context for somatic evolution; for example, germline background may affect an individual's immune system, influencing which driver genes are positively selected. Recently, a number of germline variants have been identified to affect selection on certain driver genes. However, due to the sparsity of somatic mutations and heterogeneous background mutation process across positions and individuals, current statistical tools have limited power in detecting such interactions and are prone to false positives. To address this, we developed a powerful statistical method that models positive selection on driver genes at the individual level, allowing individual-specific factors, such as germline variants, to modify the strength of selection . Our method, DiffDriver, models background mutation rates, accounting for individual-specific mutational processes and tests whether the strength of selection associates with given individual level factors. It further improves the power of detecting these associations by incorporating functional features of mutations, such as effects on proteins, in our model of selection. Through simulations, we show that compared to current methods, it reduces false positives and boosts the power of detecting interactions. We applied our method to identify interactions between germline background and selection of somatic mutations in TCGA. We revealed a number of germline variants and polygenic risk to certain traits that affect the selection of some well-known driver genes. We believe our method will be of critical value to understand selection of somatic mutations on an individual-by-individual basis.

Session Title: Cancer Poster Session III

PB4959 Accounting for heterogeneity in genetic susceptibility by disease subtype identifies eight novel breast cancer risk loci

Authors:

S. Fu[1], D. Dutta[1], X. Li[2], T. Ahearn[1], Z. Li[3], K. Wang[1], J. Jin[4], X. Li[1], K. Michailidou[5], R. Milne[6], P. Kraft[1], J. Simard[7], P. Pharoah[8], M. Schmidt[9], D. Easton[10], N. Chatterjee[11], M. Garcia-Closas[12], K. Yu[1], H. Zhang[1]; [1]Natl. Cancer Inst., Rockville, MD, [2]Harvard T.H. Chan Sch. of Publ. Hlth., Cambridge, MA, [3]Indiana Univ. Sch. of Med., Indianapolis, IN, [4]Univ. of Pennsylvania, Philadelphia, PA, [5]The Cyprus Inst. of Neurology and Genetics, Nicosia, Cyprus, [6]Cancer Council Victoria, Melbourne, Australia, [7]CHU de Québec - Université Laval Res. Ctr., Quebec, QC, Canada, [8]Univ Cambridge, Cambridge, United Kingdom, [9]Netherlands Cancer Inst., Amsterdam, Netherlands, [10]Univ. of Cambridge, Cambridge, United Kingdom, [11]Johns Hopkins Univ., Gaithersburg, MD, [12]Inst. of Cancer Res., London, United Kingdom

Background: Breast cancer is a heterogeneous disease with diverse molecular subtypes, each characterized by distinct clinical presentations, outcomes, and with variation in etiology. Recent studies suggest that many common germline variants are heterogeneously associated with breast cancer subtypes. Although existing methods, such as the two-stage polytomous model, have proven valuable, they encounter limitations in power and computational efficiency, especially when considering interaction effects across tumor markers. Therefore, there is a need to develop more advanced methods that can effectively incorporate tumor subtype information to identify novel breast cancer loci.

Methods: We developed TOPO, a novel, and computationally efficient method to detect variants exhibiting subtype heterogeneity by subtype by leveraging a comprehensive testing procedure with three sub-model structures. Using 106,278 invasive breast cancer cases and 91,477 controls of European ancestry from the Breast Cancer Association Consortium, we conducted a genome-wide association study and defined breast cancer subtypes by estrogen receptor (ER), progesterone receptor (PR), human epidermal growth receptor 2 (HER2), and tumor grade. Covariate adjustments were made for age and top ten genetic principal components. We filtered out variants within ±500kb or in linkage disequilibrium (LD>0.1) with any previously established risk loci, ensuring the novelty of identified variants.

Results: After filtering out 210 known risk loci, we identified eight novel variants (P-value <5*10^(-8)) using the proposed model. The scaled genomic inflation factor was 1.002, indicating negligible inflation in the association analysis. Conditional analyses on previously known variants further confirmed the independence of identified signals. These variants exhibited strong evidence of breast cancer subtype-specific associations across different breast cancer subtypes. For example, rs6677545 had a case-control odds ratio (OR) of 0.97 (95% confidence interval (CI): 0.95-0.99) for luminal-A like subtype (defined as ER positive or PR positive, but HER2 negative) versus an OR of 1.04 (95% CI: 1.01-1.07) for the triple-negative subtype (defined as ER, PR, HER2 all negative).

Conclusion: Utilizing our innovative method that incorporates correlated tumor characteristics, we uncovered eight new breast cancer loci, enhancing our understanding of breast cancer heterogeneity and its varied etiologies. These findings hold promise for developing subtype-specific genetic scores, enabling more precise personalized prevention strategies.

Session Title: Cancer Poster Session I

PB4960 † Additional impact of genetic ancestry over self-reported race/ethnicity to prevalence of *KRAS* mutations and allele-specific subtypes in non-small cell lung cancer

Authors:

X. Wang[1], K. Hou[2], B. Ricciuti[3], J. V. Alessi[3], X. Li[1], F. Pecci[3], R. Dey[1], J. Luo[3], M. M. Awad[3], A. Gusev[3], X. Lin[1], B. E. Johnson[3], D. C. Christiani[1]; [1]Harvard T.H. Chan Sch. of Publ. Hlth., Boston, MA, [2]Univ. of California, Los Angeles, Los Angeles, CA, [3]Dana-Farber Cancer Inst., Boston, MA

Background: *KRAS* mutations are the most common oncogenic driver in patients with non-small cell lung cancer (NSCLC). However, the interplay between self-reported race and/or ethnicity (SIRE) and genetically inferred ancestry (GIA) on *KRAS* mutations and allele-specific subtypes is largely unknown. **Methods:** A total of 3918 NSCLC patients from the Chinese OrigiMed (OM) cohort (n=2039) and Boston Lung Cancer Survival (BLCS) cohort (n=1879) were included, all of whom had well-annotated baseline covariates including age at diagnosis, sex, clinical stage, SIRE, cancer histology and smoking status. *KRAS* mutation and its subtypes were determined by targeted NGS panels in each cohort. GIA was inferred from paired germline data for the BLCS. Multivariable and multinomial logistic regressions were employed to assess the associations between SIRE, GIA and *KRAS* mutation outcomes. **Results:** In the entire cohort, 804 (20.5%) patients harbored *KRAS* mutation, with a considerably higher prevalence of 31.5% observed in the BLCS cohort compared to 10.4% in OM cohort. SIRE-Asian patients were associated with a significantly lower rate of *KRAS* mutation (OR = 0.44, 95% CI:0.22 - 0.81, P = 0.01), transversion substitutions (OR = 0.24, 95% CI:0.10 - 0.62, P = 0.003) and $KRAS^{G12C}$ (OR = 0.17, 95% CI:0.04 - 0.72, P = 0.02) compared to SIRE-White patients after adjusting for potential confounders. The association of SIRE and *KRAS* mutation was further modified by sex, with Asian women showing lower risk compared to Asian men, while the opposite trend was observed in SIRE-White patients. Increased European ancestry was associated with a higher risk of *KRAS* mutations (OR per 10% increase = 1.05, 95% CI:1.00 - 1.09, P = 0.03), particularly more transition substitutions (OR per 10% increase = 1.10, 95% CI:1.01 - 1.21, P = 0.03) and $KRAS^{G12D}$ (OR per 10% increase = 1.18, 95% CI:1.04 - 1.34, P = 0.01), and this was more pronounced in patients of increased Northern European ancestry. Furthermore, even among SIRE-White patients, a 10% increase in European ancestry was associated with an additional 6% increase in the likelihood of *KRAS* mutations (OR = 1.06, 95% CI: 1.01 - 1.12, P = 0.03), while a 10% increase in Admixed American ancestry was associated with a 6% decreased likelihood of *KRAS* mutations (OR = 0.94, 95% CI: 0.88 - 0.99, P = 0.04). **Conclusion:** SIRE and GIA both demonstrate associations with *KRAS* driver mutations and allele-specific subtypes of NSCLC. By complementing each other in the understanding of molecular heterogeneity in lung cancer, these factors can contribute to more accurate interventions and clinical decision-making, ultimately benefiting diverse patient populations of NSCLC.

Session Title: Cancer Poster Session II

PB4961 † Advancing Precision Oncology Using Integrated Interpretation of the Germline and Tumor Genomes

Authors:

A. Ghazani[1,2], A. Chittenden[3], R. Isidro[2,1], D. Koeller[3], A. Schwartz[3], M. walker[2], B. Unal[2], C. Hayes[2], M. Manam[2], R. Buehler[3], D. Manning[2], M. Redston[2,1], M. Yurgelun[1,3], H. Rana[3,1], J. Garber[3,1], A. A. Ghazani[1,2]; [1]Harvard Med. Sch., Boston, MA, [2]Brigham and Women's Hosp., Boston, MA, [3]Dana Farber Cancer Inst., Boston, MA

The standard interpretation of germline variants in cancer yields a high number of VUS due to the absence of key supporting data in favor of pathogenicity or benign function of the variants. As the largest category of reported variants, VUS bears well-documented economical, clinical, and psychosocial burden. While many in silico tools assess potential pathogenicity of variants, none exist that systematically incorporates information from tumor in the assessment of germline VUS in cancer. Tumor profiles have long been demonstrated to reveal valuable information about cancer biology, therapeutic efficacy, and tumor progression. Here we present the value of tumor genome profile in revealing key features about the constitutional genome. We have previously demonstrated the utility of tumor data in the reassessment of pathogenic germline variants in selected patients with non-syndromic phenotypes. We have now developed an evidence-based platform called INT2GRATE that allows systematic integration of key tumor parameters to aid interpretation of germline VUS in selected cancer susceptibility genes. INT2GRATE further enables collation and sharing of integrated somatic knowledge associated with germline VUS. INT2GRATE logic utilizes qualifying clinical criteria, tumor-derived genetic information and germline findings separately, in Lynch syndrome, hereditary paraganglioma-pheochromocytoma (PGL/PCC), and von Hippel-Lindau syndrome. INT2GRATE was assessed by retrospective analysis of data of different disease cohorts including patients with somatic variants in DNA mismatch repair genes (n=5018), VHL gene (n=2552); and PGL/PCC genes (n=8049). Somatic tumor profiling was performed by OncoPanel at Brigham and Women's Hospital to evaluate 447 genes. Corresponding germline variants and patient clinical data, when available, were collected from the Dana-Farber Cancer Institute Cancer Genetics database. INT2GRATE effectively identified and excluded approximately 99% of cases due to the absence of key parameters, including cases with non-syndromic tumors or those not consistent with the presentation of constitutional disease. To date, we have identified germline variants in the INT2GRATE POSITIVE category consistent with disease presentation in patients with Lynch syndrome (n=23) and PGL/PCC (n=1). We have also identified VUS in the INT2GRATE NEGATIVE category where the absence of key tumor evidence was consistent with the absence of cancer syndrome or diagnosis related to VHL (n=11) and PGL/PCC (n=18). INT2GRATE provides a platform to comprehensively assess the role of germline VUS using well-established tumor and clinical evidence.

Session Title: Cancer Poster Session III

PB4962 Aggregation of over 600,000 predictors for cancer risk stratification

Authors:

K. Liang[1], Y. Farjoun[2], V. Forgetta[3], Y. Chen[4], S. Yoshiji[1], Y. Ilboudo[5], T. Lu[6], G. Butler-Laporte[7], B. Richards[1]; [1]McGill Univ., Montreal, QC, Canada, [2]Lady Davis Inst., JGH, Montreal, QC, Canada, [3]5 Prime Sci., Montreal, QC, Canada, [4]McGill Univ., Montréal, QC, Canada, [5]Jewish Gen. Hosp., Montreal, QC, Canada, [6]Univ. of Toronto, Toronto, ON, Canada, [7]Univ. of Oxford, Oxford, United Kingdom

Background: Early cancer screening for high-risk individuals is a crucial step in cancer prevention. Genetic predisposition to cancer is currently estimated using a combination of family history and known, rare pathogenic variants. However, it is well known that common variants can have an equally significant contribution to cancer risk. Polygenic risk scores are an easy way to capture common genetic predisposition to cancer; however, it relies on the availability of well-powered GWAS. Therefore, an informative PRS for cancer requires not only high genetic heritability but also sufficiently large sample sizes for a well-powered GWAS. Despite the efforts from many population-level biobanks (e.g., UK Biobank), except for a few common cancers, many cancers still lack sufficient sample sizes for a well-powered GWAS limiting the use of PRS in these cancers.
Aims and Methods: Unlike many cancers, we do have an abundance of different -omic measurements and common traits. We, therefore, set out to determine if it is possible to incorporate these data in the form of polygenic risk scores to improve cancer risk prediction. In total, we collected PRS for 1,091 metabolites, 4,719 proteins, and 616,001 transcript measurements across 49 tissues, and 1,471 non-cancer exposures, and combined them into a single multi-PRS (mPRS) to evaluate their performance across 14 cancers.
Results and conclusion: We found that relative to the cancer PRS, the cancer mPRS can correctly reclassify up to 26% (95% CI: 25 -26%) of controls and 14% (95% CI: 0.03-0.26) of cases relative to the standard cancer PRS based on the net reclassification index. The cancer mPRS provided the most improvement in classification accuracy for chronic lymphoid leukemia (5x increase in AUPRC). For other cancers, minor improvements in classification accuracy were seen, mostly for rarer cancers. No improvement was seen for cancers with large sample sizes (i.e., breast cancer, prostate cancer). We also noted that there can be descent performance for mPRS using only biomarker data and non-cancer traits (AUROC ranging from 0.62 for breast cancer to 0.79 for urinary tract cancer). Overall, it suggests that the most effective way to obtain the best PRS for cancer is through strongly ascertained cohorts with large sample sizes. However, when this is not feasible, it is still possible to construct reasonable mPRS using only readily available biomarker data. Whether this multi-PRS approach can be refined for biomarker discovery or used in developing non-linear PRS remains to be determined.

Session Title: Cancer Poster Session I

PB4963 Alterations of super-enhancers in T-cell Prolymphocytic Leukemia.

Authors:

H. Yan, S. Tian, H. Zhang, E. W. Klee, W. Ding; Mayo Clinic, Rochester, MN

T cell prolymphocytic leukemia (T-PLL) is a rare disease, accounting for ~2% of mature lymphocytic leukemia cases in adults. With a rapid clinical course, T-PLL patients show poor response to conventional chemotherapy and immunotherapy. Genetically, T-PLL is characterized by numerous structural variants, including inv(14)(q11q32), t(14;14)(q11;q32) and t(X;14)(q28;q11), as well as recurrent mutations in chromatin regulators. On the other hand, alterations of regulatory regions, in particular enhancers, have been implicated in disease susceptibility and progression. However, the extent of epigenetic changes has not been well defined in T-PLL. We performed ChIP-seqfor histone H3 lysine 27 acetylation (H3K27ac) to identify active enhancers in patients (n=6) and age-matched healthy individuals (n=3). Samples were collected with written consent and approval from the institutional review board at Mayo Clinic. Paired-end reads were mapped to the human reference sequence hg19 using BWA, and peaks were identified by MACS2 and filtered by blacklisted regions. Between 233 and 458 super-enhancers (SEs), i.e., clusters of enhancers, were identified from individual samples using the ROSE algorithm at default parameter setting. These SEs were merged into a list of 819 SE regions. We extracted 109 T-PLL unique SEs, i.e., those present in >=2 T-PLL cases but in none of the normal samples, as well as 40 normal unique SEs, and assigned these SEs to the nearest genes. Pathway analysis with Enrichr package indicated that T-PLL unique SEs are associated with genes enriched in key pathways, such as MAPK signaling, IL-2/STAT5 signaling and T-cell receptor and co-stimulatory signaling, while normal unique SEs are associated with genes enriched in hematopoietic stem cell differentiation, Th17 cell differentiation and regulation of actin cytoskeleton pathways. We further identified 26 and 20 differential SEs with significantly increased and decreased H3K27ac in T-PLL, respectively. The former led to the activation of oncogenes TCL1A and TCL1B in T-PLL. The latter targeted cell surface receptor genes CD84 and CD352 that play key roles in immune response. Our analyses provide insights into the roles of SE alterations in T-PLL.

Session Title: Cancer Poster Session II

PB4964 An *MLH1* Gene Mutation in a 27-Year-Old Male: A Case of Somatic Symbiosis

Authors:

T. Demarco, A. Winer, K. Matthijssen, D. Alaeddin, M. Turner, K. Schwarting; Inova Schar Cancer Inst., Fairfax, VA

We present the case of a healthy 27-year-old male with an identified *MLH1* gene mutation. Mutations in the *MLH1* gene are associated with Lynch syndrome and an increased risk for the development of colon cancer and other gastrointestinal malignancies, including pancreatic cancer. This patient had recently lost his father to metastatic pancreatic cancer. A tumor specimen from his father had been sent to an established somatic laboratory for tumor testing utilizing a comprehensive multi-gene panel. This testing was reported in March of 2022 and did not indicate any germline pathogenic or likely pathogenic mutations. However, in January of 2023 following his father's death, an amended report was provided by the somatic testing laboratory to the patient's father's medical oncologist indicating an update to a variant classification. Specifically, a germline variant in the *MLH1* gene known as p.S556T was reclassified as pathogenic, with a variant allele fraction highly suggestive of germline inheritance. The medical oncologist immediately contacted the patient's son, the executor of his father's estate, and recommended genetic counseling and testing based on his father's amended somatic genetic testing report. The patient underwent genetic counseling and testing utilizing a comprehensive multi-gene inherited cancer panel. The results indicated that the patient also carried the *MLH1* mutation that was identified upon his father's somatic testing, thus confirming germline inheritance. The patient was counseled about his cancer risk and referred to an internal multi-disciplinary program to review recommended cancer screening and prevention for Lynch syndrome. For *MLH1* mutation carriers, colorectal cancer screening utilizing high-quality colonoscopy is recommended to commence between the ages of 20-25. While the patient is not currently a candidate for pancreatic cancer screening, this recommendation can be further reviewed in the future. This case illustrates the importance of integrated patient care with ongoing communication amongst various disciplines as part of a comprehensive cancer center. The patient's father's medical oncologist was proactive about contacting his son and counseling him regarding implications of the somatic testing for this patient's personal risk. If this had not been discussed, the patient may have never been made aware of his potential elevated cancer risk. In addition, it further expands the definition of "patient" to include the individual being treated and their family members, who may further benefit from ongoing clinical care and counseling even after the passing of a loved one.

Session Title: Cancer Poster Session III

PB4965 Analysis of cellular heterogeneity in bladder cancer.

Authors:

O. Doumbia[1], O. Samassekou[2,1], H. Berthé[1,3], M. Goita[4], S. Bamba[5], M. Keita[1], G. Landoure[1,6], I. Sissoko[7], A. Kassogué[1,7], A. Diarra[1,8], M. Traoré[1]; [1]Univ. of Sci., Techniques and Technologies of Bamako (USTTB), Bamako, Mali, [2]Sherbrooke Univ (CHUS), Sherbrooke, QC, Canada, [3]Urology Dept. of CHU of Point G, Bamako, Mali, [4]Univ. of Sci. Techniques and Technologies of Bamako (USTTB), Bamako, Mali, [5]USTTB, Bamako, Mali, [6]Neurology Departement of CHU point G, Bamako, Mali, [7]Ctr. Hosp.ier Univ.ire de Kati, Koulikoro, Mali, [8]Urology Dept. of Hosp. " Mère-Enfant le Luxembourg", Bamako, Mali

Background: Bladder cancer is one of the most common cancers and the first cause of death by cancer in urology worldwide and the same epidemiological trend is observed in Mali. The poor prognosis of patients could be due to the presence of aggressive forms associated with genomic instability which can be determined by cell heterogeneity. Cell Heterogeneity defined by genomic changes on cellular basis can better determine clonality and tumor aggressiveness. Cellular telogenomics which explores pan telomeric changes on cell basis might be an approach to determine phenotype differences between cancer cells based on the genomic profile.
Objectif: Determine cell heterogeneity by measuring cell dispersion in normal and cancer cells from urine of bladder cancer patients.
Methodology: We carried out cytological preparation from urine of bladder cancer patients. We did 3D FISH technique of telomeres, 3D microscopy and measured telomere length and number in 100 normal and cancer cells for each patient. To define the heterogeneity index of a cell type from a given sample, we used the similarity measure by calculating the Euclidean distance between cells of the same cell type in the sample. We developed an index score through statistical analysis by defining each cell according to their telomere number and size.
Results: This method allowed us to characterize both genetically and genomically the phenotypes of normal and cancer cells in bladder cancer. We compared the two categories of cells (cancer and normal) in terms of cellular homogeneity. We found that in more than 90% of patients, cancer cells were more heterogeneous, based on telomere length and number, than normal ones. The phenotypic characteristics of cancer cells were mostly their polyploidy state defined by the number of telomeres and the shortness of their telomeres. Moreover, we found that samples presenting a higher proportion of calls having high number and short telomeres were among the most aggressive. Finally, we defined and found that the index of heterogeneity in cancer cells (3.51) were higher than those of normal cells (2.70) with p <0. 0001.
Conclusion: This new method for analyzing telogenomic data opens prospects for more accurately determining characteristics of cancer cells such as clonal evolution, genomic instability, tumor expansion and the process of invasion of the stroma and tissues neighborhood.

Session Title: Cancer Poster Session I

PB4966 Analysis of *IDH1*positive enchondroma RNA sequencing shows disruption of the HIF1 pathway in patients with enchondromatoses

Authors:

P. Varela-Calais[1], E. Partan[2], R. Martin[1], D. Drehmer[1], C. Morris[3], G. Gottesman[4], G. Semenza[1], N. Sobreira[1]; [1]McKusick-Nathans Dept. of Genetic Med., Johns Hopkins Med., Baltimore, MD, [2]McKusick-Nathans Dept. of Genetic Med., Baltimore, MD, [3]Mem. Sloan Kettering Cancer Ctr., New York, NY, [4]Ctr. for Metabolic Bone Disease and Molecular Res., Shriners Hosp. for Children, St. Louis, MO

Ollier disease (OD), Maffucci syndrome (MS), and metaphyseal enchondromatosis with D-2-hydroxyglutaric aciduria (MCHGA) are diseases characterized by the presence of multiple enchondromas. Gain-of-function somatic variants at the active sites of *IDH1* and *IDH2* have been identified in individuals with OD, MS and MCHGA. Poll et al. (2022) identified rare candidate causative variants in 6 genes (*HIF1A*, *VHL*, *IDH1*, *IDH2*, *KDM4C*, and *CDKN2A*) related to the HIF-1 pathway. Burden analysis found that germline variants in *HIF1A*, *VHL*, and *IDH1* were all significantly enriched in cases compared to controls. Additionally, RNAseq analysis of patient and control fibroblasts showed a significantly reduced number of differentially expressed HIF-1 related genes in patient samples compared to control samples in response to hypoxia. Here, we compared RNAseq data on *IDH1*-p.Arg132Cys positive enchondroma cell lines from two patients with OD and one with MEHGA to the RNAseq data on three control cartilage cell lines. The cell lines were cultured under normoxia (20% O2) and hypoxia (1% O2 for 24 h). Similarly to what was identified in the fibroblasts, we found that the number of differentially expressed HIF-1 related genes in patient samples compared to control samples in response to hypoxia was significantly reduced. In addition, 26 HIF-1-related genes had more than three-fold increased expression in response to hypoxia in controls; and, of these, 19 had less than three-fold increased expression under hypoxia in patients. *HIF1A* regulation in response to hypoxia occurs primarily on the level of protein stability due to post-translational hydroxylation and proteasomal degradation mediated by *EGLN*s and *VHL*. *HIF1A* is an a-ketoglutarate (a-KG)-dependent oxygenase, and *IDH1* and *IDH2* catalyze conversion of isocitrate to a-KG. Mutant *IDH1* and *IDH2* gain the ability to produce the oncometabolite D-2HG, which competitively inhibits a large family of a-KG-dependent enzymes. D-2HG also stimulates activity of EGLN prolyl 4-hydroxylases, leading to diminished levels of hypoxia-inducible transcription factor. Our RNAseq data on three enchondroma cell lines harboring the *IDH1*-p.Arg132Cys variant suggests that the HIF-1 pathway is downregulated in patients' enchondromas under hypoxia compared to control cartilage. To understand better the role of the HIF-1 pathway in the enchondromas of patients with enchondromatoses, we will measure the HIF-1 transcriptional activity at normoxia and hypoxia in the enchondromas using a cell-based dual-luciferase assay.

Session Title: Cancer Poster Session II

PB4967 Analysis of imprinted genes in cancer cell lines using multi-omics data.

Authors:

J. Krushkal[1], T. Jensen[2], G. Wright[1], Y. Zhao[1]; [1]Natl. Cancer Inst., NIH, Rockville, MD, [2]The Emmes Company, LLC, Rockville, MD

Imprinted genes play important roles during embryonic and postnatal development, and many of them promote cell proliferation and body growth. In normal tissues, the imprinted genes are characterized by parent of origin-specific monoallelic expression, which is tightly controlled via epigenetic mechanisms. Some imprinting patterns are dysregulated in tumors, among a broader array of epigenetic changes in malignant cells. Our earlier study found that response of cancer cell lines to chemotherapy was associated with variation in the copy number, DNA methylation, and overall expression levels of selected imprinted genes. In the current project, we investigated which of the imprinted genes maintained monoallelic expression in cancer cell lines. We developed a computational pipeline to evaluate monoallelic and biallelic expression patterns in cancer cell lines that had available matched RNA-seq expression data, whole exome sequencing information, and copy number data. We analyzed expression patterns at the whole gene and isoform-specific levels for 59,283 autosomal protein-coding genes and ncRNA genes, with a special focus on 94 imprinted genes and 8 non-imprinted control genes. Analysis was performed in 108 cell lines from 9 pediatric and adult tumor histologies using multi-omic data obtained from the DepMap Portal for the Cancer Cell Line Encyclopedia. We will discuss the variation of monoallelic expression patterns of imprinted genes in tumor cells and the potential biological impact of imprinting changes in tumor cells. We will also discuss specific cases of monoallelic expression of non-imprinted genes in cancer cell lines.

Session Title: Cancer Poster Session III

PB4968 Analysis of the frequency and association of the rs2758346 variant of the SOD2 gene in breast cancer

Authors:

M. Gallegos-Arreola[1], B. C. Gómez-Meda[2], G. M. Zúñiga-González[3], L. Figuera[1], A. M. Puebla-Pérez[4], A. Salas-Aragon[5], A. F. Garibaldi-Ríos[6], I. A. Dávila-Carrillo[7], E. J. Maciel-Cruz[8], A. Rivera-Cameras[8], P. M. García-Verdín[8]; [1]División de Genética, CIBO, IMSS, Guadalajara, Mexico, [2]Departamento de Biología Molecular y Genómica, Inst. de Genética Humana "Dr. Enrique Corona Rivera", CUCS,, Univ. de Guadalajara, Guadalajara, Mexico, [3]Laboratorio de Mutagénesis. División de Med. Molecular, CIBO, IMSS, Guadalajara, Mexico, [4]Laboratorio de Inmunofarmacología, Centro Univ.rio de Ciencias Exactas e Ingenierías UdeG, Guadalajara, Jalisco, México., Guadalajara, Mexico, [5]Maestría en Ciencias en Química, Departamento de Química, CUCEI, UdeG., Guadalajara, Mexico, [6]Doctorado en Genética Humana, Centro Univ.rio de Ciencias de la Salud (CUCS), Univ. de Guadalajara . División de Genética, Centro de Investigación Biomédica de Occidente (CIBO), Centro Médico Natl. de Occidente (CMNO), Inst. Mexicano d, Guadalajara, Mexico, [7]Doctorado en Genética Humana, Centro Univ.rio de Ciencias de la Salud (CUCS), Univ. de Guadalajara (UdeG).. División de Genética, Centro de Investigación Biomédica de Occidente (CIBO), Centro Médico Natl. de Occidente (CMNO), IMSS, Guadalajara, Mexico, [8]Doctorado en Genética Humana, CUCS, UdeG. División de Genética, CIBO, IMSS, Guadalajara, Mexico

Introduction: rs2758346 is a single nucleotide variant (SNV) found on chromosome 6, in the SOD2 gene that codes for the enzyme superoxide dismutase 2. This enzyme protects cells from oxidative stress caused by free radicals. The T allele of this SNP seems to be more associated with the risk of developing different diseases. However, there are no association studies with breast cancer. **Objective:** Determine the association of the rs2758346 variant of the SOD2 gene with breast cancer. **Material and methods:** 109 DNA samples from breast cancer patients and 128 from the control group were included in the study. The genotypes of the rs2758346 variant were determined by TaqMan®probes, by real-time PCR assays (qPCR), and were analyzed on a CFX96 thermal cycler, C100 touch, BioRad. **Results:** The wild-type genotype (CC) was observed in 32% (36/105) of the patient group and 34% (46/128) of the control group (OR 0.84, 95%CI 0.49-1.44, p=0.62), the heterozygous genotype (CT) was present in 51% (55/109) of the patient group and 48% (62/128) of the controls, (OR 1.08, 95%CI 0.65-1.8, p=0.85) and TT genotype was present in 17% (19/109) of the patient group and 16% (20/128) of the controls (OR 0.86, 95%CI 0.43-1.17, p=0.81), without showing significant differences between the study groups **Conclusions:** The genotype TT of the rs2758346 variant of the SOD2 gene did not show a risk association for susceptibility to the development of breast cancer in the analyzed sample.

Session Title: Cancer Poster Session I

PB4969 Assessing the expression of Long Interspersed Nuclear Elements (LINEs) via long-read sequencing in diverse human tissues and cell lines

Authors:

K. Rybacki[1], M. Xia[1], J. Xing[2], K. Wang[3]; [1]Univ. of Pennsylvania, Philadelphia, PA, [2]Rutgers, the State Univ. of New Jersey, Piscataway, NJ, [3]CHOP, Philadelphia, PA

Transposable Elements (TEs), such as Long INterspersed Elements (LINEs), are DNA sequences that can replicate themselves within genomes independently of the host cell DNA. LINE activity creates genomic structural variants in human populations and alterations in cancer genomes. Long-read RNA sequencing technologies, including both Oxford Nanopore and PacBio, have generated completely sequenced transcripts and provide the opportunity to examine full-length expressed LINEs. This study focuses on the development of a new bioinformatics pipeline for the identification and quantification of active, full-length LINEs in both healthy tissues, and cancerous and normal cell lines. The pipeline utilized RepeatMasker to identify LINE-1 elements in humans, based on specific criteria such as divergence and length, to focus on young L1 elements with intact retrotransposition machinery. Comparisons of cancerous cell lines with normal cell lines, and healthy human tissue samples revealed higher expression levels of younger, active LINEs in cancer, at intact and fully transcribed L1 loci. Active and inactive L1 elements were analyzed to illustrate the presence and impact of highly mutated and less mutated LINEs across various tissue and cell line types. By employing advanced bioinformatics methodologies on full-length transcriptome data, this study demonstrates the landscape of LINE-1 elements and their contributions in normal and oncogenic cellular processes.

Session Title: Cancer Poster Session II

PB4970 Association of inferred global ancestry with cytogenetic subtypes in pediatric acute lymphoblastic leukemia

Authors:

J. Geris[1], P. Kore[1], M. Leon-Camarena[2], M. Castellanos[3], O. A. Taylor[1], V. Garcia-Morales[1], A. L. Brown[1], P. Sok[1], V. Huynh[4], K. Ludwig[5], L. Klesse[5], S. Pruitt[5], A. Hughs[5], K. Tien[1], K. Heym[6], T. Griffin[1], E. Rodrigo[1], J. C. Bernini[1], M. Gramatges[1], M. E. Scheurer[7], P. J. Lupo[1]; [1]Baylor Coll. of Med., Houston, TX, [2]Univ. of Texas - Austin, Austin, TX, [3]Univ. of California, San Francisco, San Francisco, CA, [4]Children's Hosp. of Orange County, Orange, CA, [5]Univ. of Texas Southwestern, Dallas, TX, [6]Cook Children's Hlth.Care System, Fort Worth, TX, [7]Baylor Coll. Med., Houston, TX

Pediatric acute lymphoblastic leukemia (ALL) is a heterogeneous disease with multiple cytogenetic subtypes, each associated with distinct clinical outcomes. Despite advances in therapies, racial and ethnic disparities in ALL incidence and survival exist. Recent studies have suggested ALL subtypes differ by genetic ancestry, but few efforts have replicated these differences, especially for Latinos. This study leveraged the Reducing Ethnic Disparities in Acute Leukemia (REDIAL) Consortium to examine the association between inferred global genetic ancestry and ALL subtypes. REDIAL includes patients diagnosed with ALL at seven major pediatric cancer centers in the Southwestern U.S. Demographic and clinical characteristics include self-reported race/ethnicity (SRRE), sex, and age at diagnosis. Cytogenetic aberrations were dichotomized as present or absent. Inferred global genetic ancestry proportions, including African (AFR), European (EUR), Amerindian (AMR), and Asian (AS) were estimated with the 1000 Genomes Project Phase 3 reference panels using ADMIXTURE. A K-means clustering algorithm was used to classify children into one of four groups of AFR, EUR, AMR, or AS ancestry. Logistic regression models were used to estimate adjusted odds ratios (aORs) and 95% confidence intervals (CI) accounting for sex and age at diagnosis. Of the 593 ALL patients, 58.9% self-reported as Latino, 31.0% non-Latino white (NLW), 5.7% non-Latino Black (NLB), 3.7% Asian/Pacific Islander (API), and 0.7% Other/Unknown. The mean proportion of genetic ancestry varied by SRRE: Latino, AMR: mean (standard deviation) = 59.2% (11.7); NLW, EUR = 83.1% (3.8); NLB, AFR = 82.9% (12.9); API, AS = 60.7% (17.6). Overall, cytogenetic features associated with favorable outcomes were less frequent among those of AMR ancestry compared with those of EUR ancestry (aOR 0.64, 95% CI 0.42-0.96). Differences were also observed for specific subtypes. For example, *ETV6::RUNX1* (t12;21) was less frequent in patients of AMR ancestry compared with those of EUR ancestry (aOR 0.55, 95% CI 0.33-0.90) or AFR ancestry (0.31, 0.12-0.79). *BCR2::ABL1* was less frequent in patients of AMR ancestry (1.6%) compared with patients of EUR ancestry (3.6%), but this difference was not statistically significant (aOR 0.35, 95% CI 0.09-1.31). Additionally, the prevalence of *KMT2A* gene rearrangements was higher in patients of AS ancestry (4.0%) compared with patients of EUR ancestry (0.7%) but was not significant in our adjusted model (aOR 6.24, 95% CI 0.68-56.85). Our results confirm differences in ALL subtypes by genetic ancestry, which could partially account for outcome disparities among children with ALL.

Session Title: Cancer Poster Session III

PB4971 Association of Microsatellite Instability and gene expression in pathogenesis of Colorectal Carcinoma.

Authors:

M. Kibriya[1], F. Jasmine[1], M. Raza[2], M. Kamal[3], M. Bissonnette[1], H. Ahsan[1]; [1]Univ. of Chicago, Chicago, IL, [2]Jahurul Islam Med. Coll., Kishoregonj, Bangladesh, [3]The Lab., Dhaka, Bangladesh

Background: In sporadic colorectal carcinoma (CRC), several genetic factors play important roles such as genetic mutations, genetic instability, chromosomal instability (CIN) or microsatellite instability (MSI), and DNA methylation. The MSI pathway involves the primary loss of function of DNA repair genes. In a previous study, we demonstrated a difference in DNA methylation pattern in microsatellite stable (MSS) and MSI CRC. In the current study, we explore the similarity and differences in gene expression profile in MSS and MSI at gene level and at pathway level to better understand the CRC pathogenesis and /or potential therapeutic opportunity. **Material & Methods:** Seventy one CRC patients (male = 43, Female=28) with different histological stages (stage 1=12 , stage 2=22 and stage 3=37) were studied. Eighteen of the tumors had MSI and 53 were MSS. From surgically resected colon specimens, paired tissue samples (tumor and adjacent normal) preserved in RNA later were used for genome-wide gene expression studies. **Result:** At gene level, we compared the list of differentially expressed genes (fold change (FC) >=3 and FDR 0.05) in tumor tissue compared to corresponding adjacent normal appearing tissue in patients with MSI (190 genes) and in patients with MSS tumors (129 genes). Of these, 107 genes overlapped and these genes were enriched in many of the pathways known to impact cancer. The list of genes that were differentially expressed in MSI only, showed enrichment in mainly two broad categories of pathways - (a) Inflammation related pathways like *IL 17* signaling pathway, *TNF* signaling pathway, chemokine signaling, *NFkB* signaling, cytokine-cytokine interactions, and (b) metabolism-related pathways including pentose and glucoronate interconversions, ascorbate and aldarate metabolism, steroid hormone biosynthesis, retinol metabolism and drug metabolism. Among them, the genes in Inflammation related pathways were up-regulated and the genes in metabolism related pathways were down-regulated in MSI tumor tissue. For pathway level analysis (taking all the genes in a given pathway into account), Gene set ANOVA also revealed similar results confirming the enrichment findings. For example, the 150 genes involved in *IL 17* signaling pathway were on average up-regulated by 1.19 fold (CI 1.16 - 1.21) in MSI compared to 1.14 fold (CI 1.13 - 1.16) in MSS patients (interaction p= 0.0009). **Conclusion:** We document the association between MSI status and differential gene expression which helps to broaden our understanding of CRC pathogenesis. Furthermore, targeting of several these dysregulated pathways could provide the basis for improved therapies for MSI and MSS CRC.

Session Title: Cancer Poster Session I

PB4972 Association of prostate cancer candidate genes with overall and aggressive prostate cancer in men of African ancestry.

Authors:

F. Chen[1], B. Darst[2], X. Sheng[1], A. Wang[1], Y. Xu[1], R. Hughley[1], B. Adusei[3], M. Jalloh[4], S. M. Gueye[4], A. A. Adjei[5], J. Mensah[6], P. W. Fernandez[7], A. Adebiyi[8], O. Aisuodionoe-Shadrach[9], L. Petersen[10], M. Joffe[11], J. McBride[12], J. Bensen[13], J. L. Mohler[14], J. Taylor[15], E. N. Butler[16], S. Ingles[1], B. Rybicki[17], J. Stanford[2], W. Zheng[18], S. Berndt[19], C. Huff[20], J. Lachance[21], L. Multigner[22], C. Andrews[23], T. Rebbeck[23], L. Brureau[24], S. Chanock[25], D. Conti[1], C. Haiman[1]; [1]Univ. of Southern California, Los Angeles, CA, [2]Fred Hutchinson Cancer Ctr., Seattle, WA, [3]37 Military Hosp., Accra, Ghana, [4]Hôpital Général Idrissa Pouye, Dakar, Senegal, [5]Univ. of Ghana Med. Sch., Accra, Ghana, [6]Korle Bu Teaching Hosp., Accra, Ghana, [7]Stellenbosch Univ., Cape Town, South Africa, [8]Univ. Coll. Hosp., Ibadan, Nigeria, [9]Univ. of Abuja & Univ. of Abuja Teaching Hosp., Abuja, Nigeria, [10]Mediclinic Precise, Cape Town, South Africa, [11]Wits Hlth.Consortium, Johannesburg, South Africa, [12]Ctr. for Proteomic & Genomic Res., Cape Town, South Africa, [13]Univ North Carolina at Chaper Hill, Chapel Hill, NC, [14]Roswell Park Comprehensive Cancer Ctr., Buffalo, NY, [15]Natl. Inst. of Environmental Hlth.Sci., Research Triangle Park, NC, [16]Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, [17]Henry Ford Hlth System, Sterling Heights, MI, [18]Vanderbilt Univ Med Ctr., Nashville, TN, [19]Natl. Cancer Inst., Rockville, MD, [20]Univ. of Texas M.D. Anderson Cancer Ctr., Houston, TX, [21]Georgia Inst. of Technology, Atlanta, GA, [22]Univ Rennes, Inserm, EHESP, Irset (Inst. de recherche en santé, environnement et travail), Rennes, France, [23]Dana Farber Cancer Inst., Boston, MA, [24]CHU de Pointe-à-Pitre, Univ Antilles, Univ Rennes, Inserm, EHESP, Irset (Inst. de recherche en santé, environnement et travail), Pointe-à-Pitre, France, [25]Natl. Cancer Inst., Bethesda, MD

Background: There is a growing body of evidence supporting the contributions of germline rare variants to the susceptibility of prostate cancer (PCa), especially aggressive PCa. Here we investigated whether rare germline pathogenic/likely pathogenic/deleterious (P/LD/D) variants and variants of uncertain significance (VUS) in 38 aggressive PCa genes identified mostly from European populations were associated with total and aggressive PCa risk in men of African ancestry. **Methods**: This exome sequencing analysis consists of 7,176 cases and 4,873 controls of African ancestry from North America and Africa. There are 3,283 aggressive cases (tumor stage T3/T4, regional lymph node involvement, metastatic disease, Gleason score ≥ 8.0, PSA ≥ 20 ng/mL or PCa as the underlying cause of death) including 1,074 metastatic cases, and 1,752 non-aggressive cases (Gleason score ≤ 7.0, PSA < 20 ng/mL, and stage ≤ T2). P/LP/D variants were rare (< 1% in controls) and had either a VEP impact score of "high" or a pathogenic or likely pathogenic ClinVar classification. Rare missense variants with a REVEL score > 0.6 were considered as VUS. P/LP/D and VUS variants were assessed separately in gene-based association analysis for total PCa, aggressive PCa, and metastatic PCa. **Results**: Significant associations with total PCa were observed for P/LP/D variants in *BRCA2* (OR=1.86, 95% CI=1.20-2.90, P=0.006), *CHEK2* (OR = 3.22, 95% CI=1.10-9.49, P=0.034), *HOXB13* (OR=2.28, 95% CI=1.34-3.87, P=0.002), and *PALB2* (OR=3.44, 95% CI=1.18-10.1, P=0.024). In the analysis of aggressive vs. non-aggressive cases, the association was significant for *ATM* (OR=2.60, 95% CI=1.13-5.97, P=0.025), *BRCA2* (OR=2.39, 95% CI=1.16-4.95, P=0.018), and *HOXB13* (OR=2.10, 95% CI=1.04-4.27, P=0.039). Similar suggestive associations with metastatic PCa were also observed for *ATM* and *BRCA2* (P < 0.07). In African Americans *PALB2* was strongly associated with aggressive (OR=11.7, 95% CI=1.51-91.2, P=0.019) and metastatic (OR=17.7, 95% CI=1.57-200.3, P=0.02) PCa. There was evidence that VUS of *BRCA2*,*CHEK2*, *MLH1* and *MUTYH* were associated with aggressive and/or metastatic PCa (OR of 1.45-9.01, P of 0.016-0.047).**Conclusion:** The associations of *ATM*, *BRCA2*, and *PALB2* with total PCa and/or aggressive PCa in men of African ancestry are consistent with findings from men of European ancestry. We further confirmed the association of *HOXB13* with risk of total and aggressive PCa which was driven by the known West African ancestry-specific risk variant X285K (rs77179853). These findings further support the importance of these genes in the consideration of screening and active surveillance for high-risk and advanced disease.

Session Title: Cancer Poster Session II

PB4973 Association of stromal transcriptomes with prostate cancer defined by *TMPRSS2:ERG* or *PTEN* molecular subtypes

Authors:

C. Ma[1], Y. Zhou[2], K. H. Stopsack[3,4], T. L. Lotan[5,6], L. A. Mucci[3], **K. L. Penney**[3,7], S. Tyekucheva[2]; [1]Univ. of Massachusetts Amherst, Amherst, MA, [2]Dana-Farber Cancer Inst., Boston, MA, [3]Harvard T.H. Chan Sch. of Publ. Hlth., Boston, MA, [4]Massachusetts Gen. Hosp. and Harvard Med. Sch., Boston, MA, [5]Dept. of Pathology, Johns Hopkins Univ., Baltimore, MD, [6]Dept. of Oncology, Johns Hopkins Univ., Baltimore, MD, [7]Brigham and Women's Hosp. and Harvard Med. Sch., Boston, MA

Background: The *TMPRSS2:ERG* gene fusion and *PTEN* loss are two of the most common somatic molecular alterations in primary prostate cancer. *TMPRSS2:ERG* fusion status has been implicated in the etiology of prostate cancer, while *PTEN* loss has been associated with disease aggressiveness and progression. Prior studies have examined gene expression of tumor with these molecular alterations in prostate cancer. Given the crosstalk between epithelium and stroma, the transcriptome in prostate stroma linked with these subtypes have not been described. In this study, we highlight the whole transcriptome expression associated with ERG gene fusion and PTEN loss in prostate cancer. **Methods:** We performed RNA sequencing after Illumina TruSeq Exome Capture of tumor-adjacent and benign-adjacent macrodissected prostate stromal samples from the Health Professionals Follow-up Study and the Physicians' Health Study. ERG (N=268 prostate cancer cases) and PTEN protein (N=184) were assessed in adjacent tumor tissue using validated immunohistochemistry assays. We performed differential expression analysis comparing stromal gene expression and pathways by tumor ERG gene fusion and PTEN loss status. **Results:** We found 11 genes to be differentially expressed (nominal *P*<0.001) comparing ERG positive vs. negative prostate cancer cases in tumor-adjacent stroma, including *SLCO2B1*, *NDE1*, and *KLHDC1*. Another 12 genes were differentially expressed in benign-adjacent stroma, including *ACAN*, *MIEF2*, and *ZBTB40*. At a pathway level, oxidative phosphorylation, mitochondrial electron transport, and ATP synthesis were differentially enriched in tumor-adjacent stroma comparing ERG positive and negative tumors, and pathways associated with PI3K/Akt/mTOR signaling and E2F targets were differentially enriched in benign-adjacent stroma. Comparing cases with PTEN loss with intact PTEN cases, we identified four genes to be differentially expressed in tumor-adjacent stroma: *FBXL3*, *GABRP*, and *LASP1*, and 34 genes in benign-adjacent stroma, including *DYSF*, *DCLK1*, and *CHD5*. Pathways differentially enriched between PTEN loss and intact PTEN cases included those related to myogenesis, TNFα signaling, angiogenesis, and apoptosis in benign-adjacent stroma. Pathways were not statistically significantly differentially enriched in tumor-adjacent stroma (adjusted *P*>0.2 for all). **Conclusions:** Our study uncovers stroma-specific genes and pathways differentially enriched with ERG gene fusion and PTEN loss, demonstrating that the molecular investigation of the tumor microenvironment can highlight potential new players in prostate cancer etiology and progression.

Session Title: Cancer Poster Session III

PB4974 Association of the rs17880135 variant of the SOD1 gene with breast cancer

Authors:

A. Morales-Hernández[1,2], M. G. Márquez-Rosales[1], A. F. Garibaldi-Ríos[1,3], B. C. Gómez-Meda[4], G. M. Zúñiga-González[5], L. E. Figuera-Villanueva[1,3], I. A. Carrillo-Dávila[1,3], P. M. García-Verdín[1,3], E. J. Maciel-Cruz[1,3], A. M. Puebla-Pérez[6], M. P. Gallegos-Arreola[1]; [1]División de Genética, Centro de Investigación Biomédica de Occidente (CIBO), Centro Médico Natl. de Occidente (CMNO), Inst. Mexicano del Seguro Social (IMSS), Guadalajara, Jalisco, Mexico, [2]Licenciatura en Médico Cirujano y Partero, Centro Univ.rio de Tonalá (CUT), Univ. de Guadalajara (UdeG), Guadalajara, Jalisco, Mexico, [3]Doctorado en Genética Humana, Centro Univ.rio de Ciencias de la Salud (CUCS), Univ. de Guadalajara (UdeG), Guadalajara, Jalisco, Mexico, [4]Departamento de Biología Molecular y Genómica, Inst. de Genética Humana "Dr. Enrique Corona Rivera", Centro Univ.rio de Ciencias de la Salud (CUCS), Univ. de Guadalajara (UdeG), Guadalajara, Jalisco, Mexico, [5]División de Med. Molecular, Centro de Investigación Biomédica de Occidente (CIBO), Centro Médico Natl. de Occidente (CMNO), Inst. Mexicano del Seguro Social (IMSS), Guadalajara, Jalisco, Mexico, [6]Laboratorio de Inmunofarmacología, Centro Univ.rio de Ciencias Exactas e Ingenierías (CUCEI), Univ. de Guadalajara (UdeG), Guadalajara, Jalisco, Mexico

Background: Breast cancer in Mexico and worldwide represents a health problem. This disease is characterized by a multifactorial origin, so different genetic and environmental factors are attributed to its cause. Some studies in cancer have demonstrated the association of different variants in genes that participate in the process of elimination of free radicals. Among them is the enzyme Cu-Zn cytosolic superoxide dismutase or also called *SOD1* (Cu/Zn-*SOD1*). However, there are no association studies between the rs17880135 variant of the *SOD1* gene and breast cancer. Objective: Determine the association of the rs17880135 (TG) variant of the *SOD1* gene with breast cancer. **Methodology:** 304 DNA samples from breast cancer patients and 118 from the control group were analyzed. Using IDT probes, the genotypes of the rs1788880135 variant were discriminated by qPCR. Amplification and analysis were performed on a CFX96 thermal cycler, C100 touch, BioRad. **Results:** The wild-type genotype (TT) was the most frequent genotype, observed in 85% (258/304) of the patient group and in 93% (110/118) of the control group (OR 0.4, 95%CI 0.18-0.83, p=0.032), which was shown to be a protective factor. However, heterozygous genotype (TG) was present in 15% (46/304) of the patient group and 7% (8/118) of the controls, which when comparing the study groups was observed as a risk factor (OR 2.4, 95%CI 1.12-5.36, p=0.032) for susceptibility to breast cancer development. The GG genotype was not observed in the study groups. **Conclusions:** The heterozygous genotype (TG) of the rs1788880135 variant of the *SOD1* gene is a susceptibility risk factor for the development of breast cancer in the analyzed sample.

Session Title: Cancer Poster Session I

PB4975 Association of the rs2758331 variant of the *SOD2* gene with breast cancer in a Mexican population

Authors:

A. Salas-Aragon[1,2,3], A. M. Puebla-Pérez[1], A. F. Garibaldi-Ríos[3,4], B. C. Gómez-Meda[5], G. M. Zúñiga-González[6], L. E. Figuera-Villanueva[3,4], J. I. Delgado-Saucedo[1], P. B. Castro-García[1], M. P. Gallegos-Arreola[3]; [1]Inmunofarmacología, Centro Univ.rio de Ciencias Exactas e Ingenierías UdeG, Guadalajara, Mexico, [2]Maestría en Ciencias en Química, Centro Univ.rio de Ciencias Exactas e Ingenierías (CUCEI), Univ. de Guadalajara (UdeG), Guadalajara, Mexico, [3]División de Genética, Centro de Investigación Biomédica de Occidente (CIBO), Centro Médico Natl. de Occidente (CMNO), Inst. Mexicano del Seguro Social (IMSS), Guadalajara, Mexico, [4]Doctorado en Genética Humana, Centro Univ.rio de Ciencias de la Salud (CUCS), Univ. de Guadalajara (UdeG), Guadalajara, Mexico, [5]Departamento de Biología Molecular y Genómica, Inst. de Genética Humana "Dr. Enrique Corona Rivera", CUCS, UdeG, Guadalajara, Mexico, [6]División de Med. Molecular, CIBO, CMNO, IMSS, Guadalajara, Mexico

Background: The SOD2 (Manganese superoxide dismutase, MnSOD) enzyme is involved in the detoxification of the superoxide anion, and is encoded by the *SOD2* gene, and its function is to protect the cell from free radicals generated by oxidative stress. It has an important participation in cell signaling pathways of apoptosis and oxidative stress. The variant rs27758331 C>A is located on chromosome 6 on the *SOD2* gene at position 159684038 of the sense chain, it is characterized by the change of a nucleotide from C to A and has been associated as a risk factor for susceptibility to multiple diseases. However, these studies do not include breast cancer. **Objective**: Determine the association of the rs27758331variant of the *SOD2* gene with breast cancer in a Mexican population. **Material and methods:** The genotype discrimination analysis of the rs27758331 variant of the *SOD2* gene was performed on 180 DNA samples from breast cancer patients and 172 from the control group. The genotype discrimination was determined by TaqMan® probes, by real-time PCR assays (qPCR), and was analyzed on a CFX96 thermal cycler, C100 touch, BioRad. **Results:** The wild-type genotype (CC) was observed in 29% (52/180) of the patient group and in 50% (86/172) of the control group (OR 0.40, 95%CI 0.26-0.63, p=0.00007), showing significant differences between the study groups as a protective factor for susceptibility to breast cancer. The heterozygous genotype (CA) was present in 34% (61/180) of the patient group and 29% (50/172) of the controls, which when comparing the study groups was not significantly different. Nevertheless, the AA genotype was present in 37% (67/180) of the patient group and 21% (32/172) of the controls, which when comparing the study groups was observed as a risk factor (OR 2.4, 95%CI 1.52-3.9, p=0.0003) for susceptibility to breast cancer development. **Conclusions:** The genotype AA of the rs27758331 variant of the *SOD2* gene is a susceptibility risk factor for the development of breast cancer in the analyzed sample.

Session Title: Cancer Poster Session II

PB4976 Association of the rs8720 variant in *KRAS* gene in patients with breast cancer of the Mexican population

Authors:

A. Garibaldi-Ríos[1,2], L. E. Figuera-Villanueva[1,2], E. J. Maciel-Cruz[1,2], I. A. Carrillo-Dávila[1,2], B. C. Gómez-Meda[3], G. M. Zúñiga-González[4], A. M. Puebla-Pérez[5], H. Montoya-Fuentes[4], M. P. Gallegos-Arreola[2]; [1]Doctorado en Genética Humana, Centro Univ.rio de Ciencias de la Salud (CUCS), Univ. de Guadalajara (UdeG), Guadalajara, Mexico, [2]División de Genética, Centro de Investigación Biomédica de Occidente (CIBO), Centro Médico Natl. de Occidente (CMNO), Inst. Mexicano del Seguro Social (IMSS), Guadalajara, Mexico, [3]Departamento de Biología Molecular y Genómica, Inst. de Genética Humana "Dr. Enrique Corona Rivera", CUCS, UdeG, Guadalajara, Mexico, [4]División de Med. Molecular, CIBO, CMNO, IMSS, Guadalajara, Mexico, [5]Laboratorio de Inmunofarmacología, Centro Univ.rio de Ciencias Exactas e Ingenierías UdeG, Guadalajara, Mexico

INTRODUCTION: Breast cancer (BC) is a disease caused by the abnormal and uncontrolled growth of cancerous cells in the tissues of the mammary gland. It represents the type of cancer with the highest prevalence among women worldwide, including Mexico, positioning itself as one of the leading causes of death in women in this and all countries of the world. The development of this pathology is influenced by various factors, among which hormonal and genetic factors stand out, and it has been observed that these have a greater weight in the development of the disease. The process of malignization of a cell begins with a series of alterations in the DNA sequence of certain genes, such as the *KRAS* gene. In this gene, high frequencies of mutations (85% in the RAS family) and multiple variants have been reported, including single nucleotide variants (SNV). Worldwide, association studies of variants in the *KRAS* gene with BC are limited. In the Mexican population, the association analysis of the rs8720 variant is nonexistent.

Worldwide, studies of the association of variants in the *KRAS* gene with the pathology are limited. In the Mexican population, the association analysis of the rs8720 variant is nonexistent.**OBJECTIVE:** Determine the association of the rs8720 *KRAS* gene variant with breast cancer (BC). **METHODS:** Genomic DNA samples from 492 healthy women and 377 BC patients were analyzed for the rs8720 T>C variant of the *KRAS* gene using qPCR. **RESULTS:** The genotype CC (38%) of the rs8720 variant was most frequently found in the BC group compared to the control group (23%) showed as risk factors for susceptibility to the development of BC ($p<0.05$). Discrepancies were noted in patients with BC carrying the CC and CT genotypes of the rs8720 variant and the presence of toxicity gastric or hematologic to chemotherapy in BC with stages III-IV furthermore genotype were identified as BC susceptibility risk factors. **CONCLUSIONS:** Variant rs8720 of the *KRAS* gene was associated with risk susceptibility factors for BC in the sample analyzed.

Session Title: Cancer Poster Session III

PB4977 Association of the*FGFR4*gene variant (rs351855) with colorectal cancer in a Mexican population

Authors:

I. CARRILLO DÁVILA[1,2], A. GARIBAILDI-RÍOS[3,4], B. C. GÓMEZ-MEDA[5], G. M. ZÚÑIGA-GONZÁLEZ[6], L. E. FIGUERA[3,4], M. A. Rosales-Reynoso[6], P. M. GARCÍA-VERDIN[3,4], M. P. GALLEGOS-ARREOLA[3]; [1]UNIVERSIDAD DE GUADALAJARA, GUADALAJARA, Mexico, [2]Doctorado en Genética Humana, Centro Univ.rio de Ciencias de la Salud (CUCS), Univ. de Guadalajara (UdeG), Guadalara, Mexico, [3]División de Genética, Centro de Investigación Biomédica de Occidente (CIBO), Centro Médico Natl. de Occidente (CMNO), Inst. Mexicano del Seguro Social (IMSS), GUADALAJARA, Mexico, [4]Doctorado en Genética Humana, Centro Univ.rio de Ciencias de la Salud (CUCS), Univ. de Guadalajara (UdeG), Guadalajara, Mexico, [5]Departamento de Biología Molecular y Genómica, Inst. de Genética Humana "Dr. Enrique Corona Rivera", CUCS, UdeG, GUADALAJARA, Mexico, [6]División de Med. Molecular, CIBO, CMNO, IMSS, GUADALAJARA, Mexico

Background: A rs351855 is a single nucleotide variant (SNV) missense which causes a change of substitution of glycine-to-arginine (G/A) at codon 388 (p.Gly388Arg) located in the transmembrane domain of the receptor in the exon 9 (c.1162G/A) in *FGFR4* gene. This variant was reported to increase cancer risk and has been associates with genetic predisposition to multiple types of cancer, included colorectal cancer (CRC). Has been linked to poor prognosis and very poor overall survival for population. **Objective:** Determining the association of *FGFR4* (rs351855) gene variant with colorectal cancer in Mexican population. **Materials and methods:** Genomic DNA samples from 420 healthy controls and 448 CRC patients were analyzed for the rs351855 variant of the *FGFR4* gene using qPCR. **Results:** The wild genotype GG was the most frequent, observed in 46% (205/448) of the CRC patient group and in 59% (246/420) of the control group (OR 0.59, 95%CI [0.45-0.78] p=0.000), which was shown to be a protective factor. However, AA genotype variant was present in 23% (101/448) of the patient group and 6% (24/420) of the controls, which when comparing the study groups was observed as a risk factor (OR 4.8, 95%CI [3 - 7.6] p=0.000) for susceptibility to CRC development. The GA heterozygous genotype didn't find differences between the studied groups, frequencies of 32% (142/448) and 36% (150/420) were detected for the group of patients and the group of controls, respectively (OR 0.83, 95%CI [0.63 - 1.10] p =0.237). **Conclusions:** The AA genotype of the rs351855 variant of the *FGFR4* gene was a susceptibility risk factor for the development of CRC the analyzed sample.

Session Title: Cancer Poster Session I

PB4978 B-allele frequency-based approach to detecting absence of heterozygosity enables detection of low-level mosaic events using optical genome mapping.

Authors:

A. Raksi, W. A. Sherman, M. D. Roytman, A. Hastie, N. Miller; Bionano, San Diego, CA

Introduction: Copy-neutral loss of heterozygosity has been found to be important in several types of cancer and has traditionally been identified using SNP microarray analysis. Previously, a method to detect the absence or loss of heterozygosity (AOH/LOH) using high-resolution optical genome mapping (OGM) results from the Bionano Saphyr® platform was described, based on the B-allele frequency (BAF) observed at label sites overlapping with known SNP positions, which enables the detection of AOH events at low allelic fractions, and performance was evaluated in several samples with known AOH events. Here the analytical evaluation of the method and performance of detecting mosaic AOH events at low aberrant cell fraction (ACF) based on simulated data is described.
Methods: The B-allele frequency of a label is calculated as the ratio of OGM molecules observed with label fluorescence to the total number of molecules aligned to the position. Heterozygous labels are defined as label motif sites that overlap heterozygous SNPs, where a fraction of molecules is observed with a lack of fluorescence at that location. A cohort of samples was simulated to assess performance; 230 AOH/LOH events of sizes ranging from 1 Mbp to 100 Mbp were simulated at various aberrant cell fractions as low as 5% ACF, for a total of 1,350 AOH/LOH events evaluated. Sensitivity and precision were calculated for each size range and cell fraction separately. Simulated samples were analyzed using the Bionano Solve pipeline. BAF values were normalized for the query sample at each SNP position, and AOH regions were called using the NxClinical™ software's SNP-FASST3 algorithm for segmenting and calling events. Additional validation was performed using a cohort of constitutional and cancer samples for which orthogonal testing had been performed and for which cell counts and LOH events were available.
Results: After optimizing parameters used for AOH detection, we demonstrate 92.3% sensitivity and >95% precision for calling AOH events greater than 20 Mbp at 25% aberrant cell fraction. In 15 samples with a total of 37 known AOH events, all events greater than 15 Mbp were called, including a 29 Mbp low level mosaic CN-AOH at 20% aberrant cell fraction.
Conclusion: These results show the analytical performance of an approach for detecting AOH/LOH regions at low allele frequency using optical genome mapping alone and demonstrates the utility of the method to analyzing heterogeneous cancer samples. The software will be made available as part of Bionano Solve 3.8 and VIA™v7.0.

Session Title: Cancer Poster Session II

PB4979 Base resolution deep learning models decipher the regulatory basis of thyroid cancer.

Authors:

A. Singhal, M. Kasowski, A. Kundaje; Stanford Univ., Palo Alto, CA

The initiation and progression of cancer, which has traditionally been seen as a genetic disease, is now increasingly being associated with regulatory abnormalities. Combinatorial binding of transcription factors to complex sequence syntax encoded in cis-regulatory elements mediates the development of normal cell-type specific expression of genes. However, large scale regulatory alterations including transcription factor binding to reprogram the chromatin-accessible landscape of regulatory elements can lead to altered gene function and undesirable cellular transformations that result in the formation of primary and metastatic tumors. The thyroid has been impacted by such regulatory alterations, resulting in cancer growth. Thyroid cancer is the most common cancer of hormone-producing tissue with high metastases frequency, low mutational burden, and a high degree of heritability, making it important to study genetically. To investigate regulatory activity changes in the thyroid, we have developed state-of-the-art base resolution chromatin profiling models to decipher predictive cis-regulatory sequence syntax and to predict the impact of regulatory genetic variants on the strength and shape of chromatin profiles. These models have been developed on a unique thyroid dataset consisting of matched normal, primary tumor, and metastatic tumor for each patient. Many fine-mapped SNPs from genome-wide association studies (GWAS) are in non-coding regions where they are thought to influence gene regulation. We have used our deep learning models to predict regulatory activity and compare transcription factor binding across distinct alleles. Using a prioritized list of SNPs from GWAS loci associated with Thyroid-Stimulating Hormone (TSH), our models predict a set of functional SNPs with a significant difference in the predicted counts and profiles. Our model predicted the upregulation of a transcription factor at *rs16857609*, one such functional SNP previously associated with thyroid cancer and subsequently fine-mapped in a multi-ancestry cohort. This is an expression quantitative trait locus (eQTL) for *DIRC3* and *IGFB5* in thyroid cancer (TCGA). Our deep learning models are the first of their kind to provide base resolution fine-mapping of GWAS hits to quantitative trait loci for a unique thyroid dataset. These models will help address a major gap in the thyroid cancer field by elucidating the role of specific transcription factors and genes in thyroid cancer initiation and progression across heterogeneous clinical cases.

Session Title: Cancer Poster Session III

PB4980 Analysis of the frequency and association of the rs5746094 (T/C) variant of the SOD2 gene in breast cancer

Authors:

P. GARCÍA VERDÍN[1], B. C. Gómez Meda[2], A. F. Garibaldi Rios[3], L. E. Figuera[3], I. A. Carrillo Davila[3], G. M. Zúñiga-González[4], M. P. Gallegos Arreola[5]; [1]Doctorado en Genética Humana, GUADALAJARA, Mexico, [2]2Departamento de Biología Molecular y Genómica, Inst. de Genética Humana "Dr. Enrique Corona Rivera", CUCS, UdeG, Guadalajara, Mexico, [3]Doctorado en Genética Humana, CUCS, UdeG. División de Genética, CIBO, IMSS, Guadalajara, Mexico, [4]Laboratorio de Mutagénesis, División de Med. Molecular, CIBO, CMNO, IMSS, Guadalajara, Mexico, [5]División de Genética, CIBO, IMSS, Guadalajara, Mexico

Introduction: rs5746094 is a single nucleotide variant (SNV) found on chromosome 6, in the SOD2 gene that codes for the enzyme superoxide dismutase 2. This enzyme protects cells from oxidative stress caused by free radicals. The C allele of this SNP seems to be more associated with the risk of developing different diseases. However, there are no association studies with breast cancer. **Objective:** Determine the association of the rs5746094 variant of the SOD2 gene with breast cancer. **Material and methods:** 213 DNA samples from breast cancer patients and 72 from the control group were included in the study. The genotypes of the rs5746094 variant were determined by TaqMan®probes, by real-time PCR assays (qPCR), and were analyzed on a CFX96 thermal cycler, C100 touch, BioRad. **Results:** The wild-type genotype (TT) was the most frequent genotype, observed in 62% (133/213) of the patient group and in 50% (36/72) of the control group (OR 166, 95%CI 0.97-2.84, p=0.085), without showing significant differences between the study groups. However, heterozygous genotype (TC) was present in 16% (34/213) of the patient group and 42% (30/72) of the controls, which when comparing the study groups was observed as a protector factor (OR 0.26, 95%CI 0.14-0.48, p=0.00001). Nevertheless, the CC genotype was present in 22% (46/213) of the patient group and 8% (6/72) of the controls, which when comparing the study groups was observed as a risk factor (OR 3.03, 95%CI 1.25-7.43, p=0.019) for susceptibility to breast cancer development. **Conclusions:** The genotype CC of the rs5746094 variant of the SOD2 gene is a susceptibility risk factor for the development of breast cancer in the analyzed sample.

Session Title: Cancer Poster Session I

PB4981 Breast cancer prevalence and therapeutic complications in carriers of *BRCA1/2* pathogenic variants after *Salpingo-Oophorectomy* risk in *BRCA* gene carriers: A Narrative Review Study

Authors:

A. Ahmadi; Isalmic Azad Univ. Tehran Med. Sci., Tehran, Iran, Islamic Republic of

Introduction: Cancer begins when healthy cells in the breast change and grow out of control and form a lump or sheet of cells called tumors. A tumor can be cancerous or non-cancerous which is also called benign. A tumor is a malignant cancer that can grow and spread to other parts of the body. A benign tumor can grow but has not spread. *Risk Reduction Salpengo Oophorectomy (RRSO)* is the gold standard method of reducing the risk of ovarian cancer, but data on the impact on breast cancer (BC) outcomes are inconsistent. An estimated 43,700 deaths (43,170 women and 530 men) from breast cancer occur in the United States in 2023. Worldwide, breast cancer is the fifth leading cause of death. Women with *BRCA1* or *BRCA2* gene mutations are at risk of breast and ovarian cancer and often undergo surgery to remove both of their ovaries to prevent ovarian cancer. The effect of this operation on breast cancer risk is unclear. The aim of this study was to determine BC mortality rate in *BRCA1/BRCA2* carriers after *RRSO*

Methods: This descriptive study was conducted with narrative review approach in 2023 by searching keywords such as *BRCA*, Breast Cancer, risk - reducing *Salpingo-Oophorectomy, Pathogenic Variants and Prevalence* in valid databases such as Scopus, Science Direct, and Web of Science.

Results: Worldwide, female breast cancer is now outnumbered by lung cancer as the most common cancer. An estimated 2,261,415 new cases of breast cancer were diagnosed in women worldwide in 2020. More women in the United States are diagnosed with breast cancer than any other type of cancer, in addition to skin cancer. It accounts for 1 in 3 new cancers in women annually. The practice is not associated with a reduced risk of breast cancer when considering both *BRCA1* and *BRCA2* carriers together, but it has been linked to a reduced risk of breast cancer when considering the *BRCA2* carrier alone. These findings may have important implications for counselling for women in the clinic.

Conclusion: Carriers of the *BRCA1/2* pathogenic variant increase the risk of breast and ovarian cancer in their lifetime. Women with cancer who have recently been diagnosed with this pathogen may experience psychological distress due to an imminent threat to health.

Session Title: Cancer Poster Session II

PB4982 Cancer predisposition in an Asian population of diverse ancestries

Authors:

S. Chan[1,2], W. Lim[3,4], **J. Ngeow**[1,2]; [1]Natl. Cancer Ctr. Singapore, Singapore, Singapore, [2]Lee Kong Chian Sch. of Med., Nanyang Technological Univ., Singapore, Singapore, [3]SingHlth.Duke-NUS Inst. of Precision Med., Singapore, Singapore, [4]Genome Inst. of Singapore, Agency for Sci., Technology and Res., Singapore, Singapore

Hereditary cancer risk assessment guidelines have primarily been developed based on data from predominantly European-descent populations due to underrepresentation of non-European genomes in human genomics studies. This limits the utility of guidelines in non-European populations, such as Asians, in which genomic variation is less characterized. In this study, we evaluated germline pathogenic variants in cancer predisposition genes from 9,051 genomes of SG10K_Health cohort representing populations of East Asian, South Asian and Austronesian-speaking Southeast Asian ancestries. We found 2.5% (227/9,051) of SG10K_Health at-risk of dominantly inherited cancer predisposition disorders, primarily harboring pathogenic variants (PVs) in breast and ovarian cancer susceptibility genes including *ATM*, *BRCA1*, *BRCA2*, *BRIP1* and *RAD51D*. Hereditary breast and ovarian cancer and Lynch syndrome were the most prevalent cancer syndromes, with an overall prevalence of 1:148 and 1:532 respectively, which are higher than current global estimates. We observed a high carrier burden for recessively inherited bone marrow failure syndromes, particularly Shwachman-Diamond Syndrome (SDS), with 1.2% (109/9,051) of SG10K_Health carrying a *SBDS* PV associated with SDS. We evaluated and found minimal overlap of PVs identified in SG10K_Health with our cancer patient cohort (n=3,537) that underwent clinical multi-gene panel testing. About 12% (25/197) of PVs identified in breast/ovarian cancer patients (n=2,306) were found in SG10K_Health whereas only 4% (2/46) of PVs in colorectal cancer patients (n=251) overlapped and none of the 23 PVs identified in endometrial cancer patients (n=149) were detected in SG10K_Health. Our study highlighted that majority of germline PVs in the cancer cohort are rare and unlikely observed in the general population; cautioning the reliance on population screening data, from European and non-European populations alike, for developing genomic medicine guidelines and making clinical decisions for hereditary cancer risk assessment and management.

Session Title: Cancer Poster Session III

PB4983 Cancer predisposition signaling pathways drive Beckwith-Wiedemann Syndrome Wilms Tumor development

Authors:

S. Nirgude[1], N. Naveh[1], S. Kavari[1], J. Kalish[2]; [1]Children's Hosp. of Philadelphia, Philadelphia, PA, [2]Children s Hosp. of Philadelphia, Philadelphia, PA

Background: Wilms tumor (WT) exhibits structural and epigenetic changes at chromosome 11p15, which also cause Beckwith-Wiedemann Syndrome (BWS). Children diagnosed with BWS have increased risk for WT. The aim of this study is to identify the molecular signaling signatures in BWS driving these tumors.**Methods**: We performed whole exome sequencing, methylation array analysis, and gene expression analysis on BWS-WT samples. Our data were compared to publicly available nonBWS data. We categorized WT from BWS and nonBWS patients by assessment of 11p15 methylation status and defined 5 groups- control kidney, BWS-nontumor kidney, BWS-WT, normal-11p15 nonBWS-WT, altered-11p15 nonBWS-WT.**Results**: BWS-WT samples showed single nucleotide variants in *BCORL1*, *ASXL1*, *ATM* and *AXL* but absence of recurrent gene mutations associated with sporadic WT. We defined a narrow methylation range stratifying nonBWS-WT samples. BWS-WT and altered-11p15 nonBWS-WT showed enrichment of common and unique molecular signatures based on global differential methylation and gene expression analysis. CTNNB1 overexpression and broad range of interactions were seen in the BWS-WT interactome study.**Conclusion**: While WT predisposition in BWS is well-established, as are 11p15 alterations in nonBWS-WT, this study focused on stratifying tumor genomics by BWS status. Further investigation of our findings may identify novel therapeutic targets in WT oncogenesis.

Session Title: Cancer Poster Session I

PB4984 Cancer risk in Simpson-Golabi-Behmel Syndrome: Case Series and Literature Review.

Authors:

A. Nisbet[1,2], S. D. Klein[1,3], J. B. Kalish[1,3,4,5]; [1]Children's Hosp. of Philadelphia, Div. of Human Genetics, Philadelphia, PA, [2]Perelman Sch. of Med., Philadelphia, PA, [3]Dept. of Pediatrics, Perelman Sch. of Med., Philadelphia, PA, [4]Ctr. for Childhood Cancer Res., Children's Hosp. of Philadelphia, Philadelphia, PA, [5]Dept. of Genetics, Perelman Sch. of Med., Philadelphia, PA

*Introduction:*Simpson-Golabi-Behmel Syndrome (SGBS) is an overgrowth syndrome characterized by macrosomia, craniofacial and gastrointestinal anomalies, intellectual disability, and congenital heart defects. SGBS is caused by variants in *GPC3* or *GPC4* on the X chromosome. Less than 250 cases are reported in the literature. Like other overgrowth syndromes, it has been proposed that patients with SGBS have an increased risk of cancer; but the actual risk is unknown. Current cancer screening guidelines are based on those developed Beckwith-Wiedemann Syndrome. Here, we describe the largest cohort of SGBS to date to define the genotype/phenotype correlation and the spectrum of clinical findings.
*Methods:*We conducted a PubMed literature review for all known cases of SGBS and included nine unreported patients. We calculated the cancer risk and developed a SGBS severity score. 90 out of 323 articles were classified as case reports and phenotyped. Each case was reviewed for data on demographics, genetic information, and comorbidities. The nine unreported cases were also phenotyped. Secondary analyses to further characterize the patients who developed tumors were performed.
*Results:*Based on reviewed data from the literature, organomegaly was reported in 32% of cases with tumors developing in 10.4% of cases. Wilms tumor (WT) and hepatoblastoma (HB) were each reported in 3.9% of cases. Metastatic medulloblastoma and ameloblastoma were also identified. Of the patients that developed cancer, 75% were male, 25% were female, 62% had familial germline mutations, and all cases were due to variants in *GPC3*. Specifically, nonsense and intragenic indels appear to have a greater cancer risk than complete gene duplications/deletions, missense, and frameshift variants. In our unreported cohort, organomegaly and renal anomalies were each reported in 75% of cases and the cancer risk was 12.5%. It appears that patients with more severe clinical presentations were more likely to develop cancer.
*Conclusion:*Using the largest known dataset of patients with SGBS, we define an overall cancer risk of 10-12%, with the risk of WT and HB each approaching 4%. There is an additional risk of non-abdominal tumors around 2%. Based on cancer screening guidelines for syndromic causes of WT and HB in the United States, we recommend screening in SGBS in a similar manner to Beckwith-Wiedemann Syndrome. The higher-than-expected risk of non-abdominal malignancies suggests that additional imaging and surveillance could be considered as well. Additionally, the risk of cancer being highest in nonsense and intragenic indels merits further investigation.

Session Title: Cancer Poster Session II

PB4985 Cancer somatic mutations in miRNA genes affect miRNA biogenesis, miRNA level, isomiRs distribution, and effectiveness of targets regulation.

Authors:

M. Machowska, N. Szostak, P. Kozlowski; Inst. of Bioorganic Chemistry, Polish Academy of Sci., Poznan, Poland

MicroRNAs (miRNAs) play an emerging role in cancer development. miRNAs are small non-coding RNAs that usually downregulate gene expression by binding to 3'UTRs of mRNAs, leading to translation repression or mRNA degradation. Changes in the level of many mature miRNAs and their involvement in carcinogenesis are commonly observed in cancer cells. Moreover, many somatic mutations in miRNA-coding genes have been detected in cancers, however, their effects on miRNA biogenesis, miRNA level, miRNA isoforms (isomiRs) generation, and loss- or gain-of-function effect are mostly unknown.

In the study, we analyzed the effects of somatic mutations identified in miRNA genes in cancer samples from The Cancer Genome Atlas (TCGA). Based on the miRNA-seq data from TCGA, we analyzed the influence of mutations located in all regions of miRNA precursor, including miRNA duplex (seed or non-seed), loop, and flanking regions. First, we focused on mutations located in miRNA duplexes, which effect could be directly observed in miRNA-seq data. It allowed to compare miRNAs derived from the wild-type or mutant allele of the same sample. In most analyzed cases, we observed a significant decrease (or complete loss) of miRNAs derived from the mutated allele in comparison to the wild-type allele. Moreover, for most of the mutated miRNAs expressed on the level sufficient for detailed analysis of isomiRs, we found significant changes in isomiRs distribution between wild-type and mutant allele. Such alterations may result in differences in target recognition (independent from the mutations in seed region that directly affect the target recognition) or may influence the miRNA stability. We observed similar effects of mutations located in the loop or flanking regions. Among the mutated miRNA genes with the mutations having a substantial effect on generated isomiRs were genes playing an important role in cancer, e.g., *MIR142*, *MIR205*, and *MIR342*. Furthermore, we conducted a luciferase assay and showed that mutation-specific isomiRs recognize and regulate different targets. Our results showed that most mutations in miRNA genes have an influence on miRNA gene functioning, either affecting miRNA level, miRNA processing, isomiRs distribution, and/or targets recognition and downregulation.

Session Title: Cancer Poster Session III

PB4986 † Cascade testing with comprehensive multigene panels for hereditary cancer identifies unexpected findings in relatives

Authors:

D. Pineda-Alvarez[1], H. Hampel[2], S. M. Nielsen[1], S. Bristow[1], E. D. Esplin[1], R. L. Nussbaum[1], B. Heald[1]; [1]Invitae, San Francisco, CA, [2]City of Hope Comprehensive Cancer Ctr., Duarte, CA

Background: Studies of relatives undergoing multigene panel testing (MGPT) for a familial pathogenic germline variant (PGV) are limited; one small study found that first-degree relatives (FDRs) undergoing cascade testing via a 30-gene panel had a 4.9% rate of unexpected findings (different PGV(s) than the proband; Caswell-Jin J et al 2019 JNCI). The aim of this study was to see if the above findings could be replicated in a larger cohort of relatives undergoing MGPT for a familial PGV.
Methods: Relatives undergoing clinician-ordered cascade testing for a familial PGV in a cancer predisposition gene at a single commercial laboratory were identified. Analysis was limited to those relatives undergoing a MGPT of ≥47 genes. The PGV(s) identified in the relative were compared to the proband's PGV(s). Demographics, relationship to proband, and cancer history data were pulled from test requisition forms; descriptive statistics, t test, and Fisher's exact test were utilized.
Results: 15,362 relatives of 7,433 probands with ≥1 PGVs underwent cascade testing. Among these relatives, 3,706 (24.1%) had clinician-ordered MGPT. Relatives were primarily female (75.9%), White (74.4%), and FDRs of the proband (73.6%), and 713 (19.2%) reported a personal history of cancer. Unexpected PGVs were identified in 230 (6.2%) relatives. This included 11 (0.3%) who were negative for the familial PGV but positive for a different PGV in the same gene as the proband and 144 (3.9%) who were negative for the familial PGV but positive for a PGV in a different gene than the proband. An additional 75 (2.0%) relatives tested positive for the familial PGV and were positive for an additional PGV in a different gene than the proband. Gender, age at testing, personal cancer history were not associated with an unexpected finding, but being White ($p=0.0253$) or at least a second-degree relative were ($p<0.0001$). Among the 219 relatives with a PGV in a different gene, 45 (20.5%) were in a high-risk gene, 54 (25.7%) in a moderate-risk gene, 18 (8.2%) in a gene associated with rare cancer types, and 102 (46.6%) with a low-risk or carrier finding. Based on the unexpected PGV(s), 79 (36.3%) relatives would potentially have qualified for different or additional cancer screening recommendations than the recommendations based on proband's PGV.
Conclusion: Among relatives undergoing cascade testing via a MGPT, unexpected results were found in 6%. 46% of these results were in high- or moderate-risk genes. Limiting cascade testing to only the familial PGV might result in missed, actionable findings for a subset of relatives.
Reference: Caswell-Jin, JL., et al. J Natl Cancer Inst. 2019 Jan 1;111(1):95-98.

Session Title: Cancer Poster Session I

PB4987 Cell type-specific effects of common germline variation on gene expression and pancreatic cancer risk.

Authors:

G. Prasad[1,2], T. Truongvo[1], J. Zhong[2], J. Hoskins[2], L. T. Amundadottir[2], E. H. Arda[1]; [1]Lab. of Receptor Biology and Gene Expression, Ctr. for Cancer Res., Natl. Cancer Inst., NIH, Bethesda, MD, [2]Lab. of Translational Genomics, Div. of Cancer Epidemiology and Genetics, Natl. Cancer Inst., NIH, Rockville, MD

Pancreatic ductal adenocarcinoma (PDAC) is a highly aggressive and fatal cancer of the exocrine pancreas, which is composed of acinar and duct cells. Uncovering genes that influence PDAC risk is crucial for early detection, prevention, and development of effective therapies. While a few high-risk germline mutations have been identified in protein coding regions of the genome, low-penetrance, common variants identified through Genome Wide Association Studies (GWAS) primarily reside in non-coding regions. This suggests that disease-associated variants contribute to risk by modifying non-coding gene regulatory elements. To identify potential gene targets influenced by such non-coding risk variants, we are employing cell type-resolved expression, chromatin accessibility and methylation quantitative trait loci (eQTL, caQTL and meQTL, respectively) analyses in purified pancreatic acinar and ductal cells. Previous QTL studies in the pancreas have used bulk tissue or tumor samples, potentially masking cell type-specific genetic effects underlying disease risk. In this work, we used a flow cytometry-based method to purify human acinar and duct cells obtained from organ donors. To obtain cell type-specific eQTL, caQTL, and meQTL data, we performed high-throughput genotyping, DNA methylation analysis, RNA sequencing, and ATAC sequencing on purified cells from 100 individuals (100 acinar and 100 duct samples). Initial genetic analyses revealed a diverse ancestry, with the majority of samples being from individuals of European descent. All samples exhibited >10 million mapped (usable) expression reads. Acinar samples were enriched for the expression of pancreatic secretory genes including *CPA1*, *PNLIP*, *PRSS1*, and *CEL*, demonstrating that our cell purification approach successfully separated acinar and duct cells. We performed principal component and hierarchical clustering analyses to identify potentially confounding variables and detect outliers in the expression data. Clustering analysis highlighted distinct cell populations and detected two outlier samples. Factors such as age, gender, BMI, diabetic status, sample source, ethnicity, and RNA quality did not correlate with any sample clusters, though differences in sequencing library preparation methods did. Our study has the potential to identify novel genes and putative regulatory mechanisms underlying PDAC risk with enhanced precision compared to bulk tissue QTL studies. Furthermore, these datasets will serve as valuable resources for annotating non-coding risk loci associated with other pancreatic diseases, including pancreatitis and diabetes.

Session Title: Cancer Poster Session II

PB4988 Characterization of cell free DNA and circulating tumor DNA in a pediatric oncology cohort and implementation into the SickKids Cancer Sequencing (KiCS) precision oncology program.

Authors:

R. Venier, R. Ripsman, R. Khan, A. Huynh, A. Villani, D. Malkin, S. Cohen-Gogo, A. Shlien; The Hosp. for Sick Children, Toronto, ON, Canada

Current survival rates for children and adolescents with newly diagnosed cancer are greater than 80% while survival rates for patients with relapsed, refractory, or metastatic disease are significantly lower. Sensitive methods to monitor disease and identify novel therapeutic targets are key to improving survival for these patients. Studies have shown that cell free (cfDNA) and circulating tumor DNA (ctDNA) analysis from blood, is a promising, non-invasive technique that may improve outcomes for this population. Nonetheless, pediatric cohorts exhibit unique challenges compared to their adult counterparts.

SickKids Cancer Sequencing (KiCS) is a precision oncology program that performs next generation sequencing on samples from pediatric patients with rare or hard-to-cure tumors. We recently described the results of the first 300 KiCS participants (PMID: 36585449), of whom the majority had paired tumor/germline cancer panel testing, whole genome sequencing (WGS) and RNA-sequencing. Thus far, the program has enrolled 594 participants and collected tumor and blood specimens with matched clinical data. Here we describe the implementation of a plasma biobank initiative and cfDNA analysis of a small cohort.

Serial blood samples were collected from patients with cancer at scheduled timepoints, at time of treatment change, or if recurrence was suspected. Blood was separated and banked as plasma and matched cell pellet. cfDNA characteristics were measured using a high sensitivity Bioanalyzer chip. Libraries were generated using a low input kit and WGS (60-80X) was performed. cfDNA data were compared to tumor and germline data from the same patient.

KiCS has banked over 300 plasma samples from more than 100 patients at various timepoints of disease. This pilot study extracted cfDNA from 10 plasma samples. cfDNA fragment size was within the expected ranges for all samples. The concentration of cfDNA in these pediatric samples was comparable to published adult cohorts (average 28.4 ng/ml, median 7.53 ng/ml) though total yield was decreased due to lower collection volumes. Optimization of a low input library preparation method and deep WGS identified expected (tumor-informed) and unique ctDNA variants.

The KiCS program implemented a robust blood collection and cfDNA sequencing workflow within an existing precision oncology program at our institution. We successfully isolated cfDNA from minimal plasma volumes and identified ctDNA variants in patient samples. Further work will focus on optimizing sequencing pipelines. The KiCS program and team are well poised to build upon these data to determine the clinical utility of cfDNA in a large pediatric cancer cohort.

Session Title: Cancer Poster Session III

PB4989 Characterization of cytogenetically cryptic abnormalities in pediatric acute myeloid leukemia by optical genome mapping

Authors:

G. Raca, A. E. Kovach, K. Miller, J. Han, K. Ma, C. Fong, A. Doan, J. Ji, R. Schmidt, D. Bhojwani; Children's Hosp. Los Angeles, Los Angeles, CA

Abnormal gene fusions are predominant oncogenic drivers in pediatric acute myeloid leukemia (AML), and are generated by heterogeneous chromosome rearrangements the majority of which cannot be detected by conventional cytogenetic analysis. RNA-Seq- and panel-based fusion testing are not routinely performed for pediatric AML, and the content of commonly used targeted fusion panels is often biased toward adult cases. To evaluate whether this diagnostic gap can be addressed by optical genome mapping (OGM), a powerful new method for genome-wide high-resolution detection of copy number abnormalities and balanced chromosomal rearrangements, we tested the ability of OGM to identify cytogenetically cryptic fusions in pediatric AML. Retrospective review of 91 pediatric AML cases tested in our laboratory between June 2017 and May 2023 identified 36 cases (39.5%) resolved by karyotype analysis, 12 cases (12.5%) in which the primary driver was a DNA sequence variant, and 29 cases (31.8%) in which the primary genetic subtype remained undetermined after karyotype analysis and clinical molecular testing. Finally, we identified 14 cases (15.3%) in which the CHLA custom sequencing panel (OncoKids®) and/or RNA-Seq fusion assay revealed abnormal fusions which were undetected by karyotype analysis. These included fusions known to be cytogenetically cryptic (*NUP98::NSD1*, n=4; *CBFA2T3::GLIS2*, n=2) as well as fusions typically created by visible translocations that nevertheless remained undetected due to association with either complex or variant chromosome rearrangements (*KMT2A::MLLT10* (n=4), *KMT2A::MLLT3* (n=2), *NUP98-KDM5A* and *KMT2A::AFDN*). To date, OGM is completed in 10/14 fusion positive cases with non-informative karyotype, and could not be performed in 4 cases due to lack of samples (n=2) or poor DNA quality (n=2). OGM analysis successfully revealed all fusions previously detected by OncoKids® and/or RNA-Seq, including *NUP98::NSD1* (n=3), *KMT2A::MLLT10* (n=3), *KMT2A::MLLT3* (n=2) and one case each of *CBFA2T3::GLIS2* and *NUP98::KDM5A.* In 2/29 cases processed to date, which remained unknown after multi-modal clinical testing, OGM revealed *MYB::GATA1* and *ETV6::MNX1* fusions; the analysis is in progress for the remaining 27 unknown cases. OGM also matched the performance of chromosomal microarray (CMA) analysis in detection of copy number variants and revealed the nature and detailed structure of the chromosome rearrangements creating the observed gene fusions. Our preliminary results show potential of OGM to identify clinically significant cryptic fusions in pediatric AML and to thus markedly increase diagnostic yield relative to karyotype analysis.

Session Title: Cancer Poster Session I

PB4990 Characterization of the mutational landscape of metastatic thyroid cancer via targeted next-generation sequencing

Authors:

A. Chung[1], C. Pei-Lung[2], S. Shyang-Rong[1]; [1]Natl. Taiwan Univ. Coll. of Med., Taipei, Taiwan, [2]Natl. Taiwan Univ. Hosp., Taipei, Taiwan

Thyroid cancer patients with distant metastasis contribute to a poor prognosis. Profiling somatic mutations in metastatic thyroid cancer may lay the foundation for cancer treatment and prognosis prediction. We collected samples from 22 thyroid cancer patients with distant metastasis, including 11 cases of papillary thyroid cancer (PTC), 5 cases of follicular thyroid cancer (FTC), 1 case of anaplastic thyroid cancer (ATC), 1 case of medullary thyroid cancer (MTC), 1 case of poorly differentiated thyroid carcinoma (PDTC), and 3 cases of co-existence of two distinct types. We sequenced the tumors from both the primary site and distant metastasis using a 50-gene panel specific to thyroid cancer, which included TERT promoter and noncoding genes GAS8-AS1. Additionally, we used the KAPA hyperPETE lung cancer fusion panel, designed for the detection of 17 fusion genes. We identified hotspot mutations in BRAF and TERT in 12 out of 22 (60%) patients. The frequency of hotspot mutations in BRAF was consistent with previous reports, while the frequency of hotspot mutations in the TERT promoter was higher than previously reported in our cohort, suggesting TERT hotspot mutations may associate with distant metastasis. We observed that hotspot mutations in BRAF and the RAS family were mutually exclusive, and hotspot mutations in BRAF were predominantly found in PTC, whereas hotspot mutations in the RAS family were more prominent in FTC. We compared somatic mutations between the tumor tissue from the primary site and paired distant metastatic sites. We consistently found mutations in BRAF and RAS family genes existing in both the primary and distant metastatic sites, while other mutations were not consistent. Regarding fusion genes, we detected one RNA fusion, PAX8-PPARG fusion, in one case of FTC, which was not detected with any pathogenic variants. In summary, we identified at least one pathogenic somatic event in 20 out of 22 patients (91%). The frequency of hotspot mutations in the TERT promoter was higher than in other reports, suggesting its potential association with distant metastasis. This study provides a comprehensive characterization of metastatic thyroid cancer and its primary and metastatic sites.

Session Title: Cancer Poster Session II

PB4991 Characterization of tumor immune cell infiltration in breast cancer predisposition variant carriers identifies enrichment of tumor-associated macrophages.

Authors:

R. Kelly[1,2], Y. Liu[2], A. R. Harris[1,3], C. Peng[4], Y. J. Heng[5], G. M. Baker[5], D. G. Stover[6,7], R. Tamimi[8], P. Kraft[2,9]; [1]Cancer Prevention Fellowship Program, Natl. Cancer Inst., Rockville, MD, [2]Harvard T.H. Chan Sch. of Publ. Hlth., Boston, MA, [3]Integrative Tumor Epidemiology Branch, Dept. of Cancer Epidemiology and Genetics, Natl. Cancer Inst., Rockville, MD, [4]Channing Div. of Network Med., Dept. of Med., Brigham & Women's Hosp., Boston, MA, [5]Beth Israel Deaconess Med. Ctr., Harvard Med. Sch., Boston, MA, [6]Div. of Med. Oncology, Stefanie Spielman Comprehensive Breast Ctr., Ohio State Univ. Comprehensive Cancer Ctr., Columbus, OH, [7]Dept. of BioMed. Informatics, Ohio State Univ., Columbus, OH, [8]Dept. of Population Hlth.Sci., Weill Cornell Med., New York, NY, [9]Trans-Div.al Res. Program, Natl. Cancer Inst., Rockville, MD

Despite improvements in overall survival, breast cancer remains a major cause of cancer mortality and incidence is increasing in the United States. Roughly 5% of breast cancer patients have a pathogenic germline genetic variant that is associated with increased risk of breast cancer. Mutations in > 25 genes have been associated with hereditary breast cancer, many of which are involved in genome stability pathways including DNA double-strand break (DSB) repair. In addition to roles in both carcinogenesis and treatment response, the DSB DNA damage response in tumor cells promotes their crosstalk with immune cells in the tumor microenvironment. We hypothesized that DSB repair-related variant carriers may share common characteristics in the immune microenvironment of their tumors that are distinct from non-carriers. Here, we performed transcriptomic analyses of 362 incident breast cancer cases from the Nurses' Health Study (NHS) I and II, including 26 with germline DSB repair-related pathogenic variants in *ATM, BLM, BRCA1, BRCA2, BRIP1, CHEK2, FANCC, FANCM, NBN, RAD50, RAD51C,* and *RECQL*. Gene expression-based cell mixture deconvolution algorithm CIBERSORTx was applied to tumor microarray data to infer the abundance of 22 immune cell types. Immunohistochemical (IHC) staining for four immune cell markers (CD4, CD8, CD20, and CD163) was used to assess immune cell infiltration in a subset of 55 participants. We observed a significant, positive association between carrier status and M1 macrophage infiltration using multivariable linear regression adjusted for age at diagnosis, tumor stage, estrogen receptor, and the top four genetic principal components (change in mean abundance: 0.0195 [0.0091, 0.0299], p = 0.0003). Carriers also showed significantly higher infiltration of CD163+ cells in our subset of participants for whom IHC data was available (Wilcoxon p = 0.0024), suggestive of M2 macrophages. Abundance of these M2 macrophages was highly correlated with that of inferred M1 macrophages (Spearman's rho = 0.5079, p = 7.56E-05), suggesting that carrier status may be associated with increased abundance of both classically and alternatively activated macrophages in their breast tumors. Our results suggest that breast tumors of DSB repair-related pathogenic variant carriers may possess a distinct immune landscape, which may have broader therapeutic implications in this high-risk population.

Session Title: Cancer Poster Session III

PB4992 Characterizing novel fusion proteins in Acute Megakaryoblastic Leukemia (AMKL)

Authors:

A. Perreault, M. Josling; Elon Univ., Elon, NC

Phase separation is the biophysical process of macromolecule condensate formation through interactions between proteins and nucleic acids. Recent literature has suggested that fusion proteins differentially phase separate, causing dysfunctional chromatin organization that can lead to diseases, such as cancer. Importantly, these phase separation-induced chromatin contacts are independent of the canonical proteins necessary for forming chromatin loops, such as CTCF and cohesin. Therefore, gaining a deeper understanding of the structure and function of fusion proteins and how they contribute to impaired gene expression that leads to disease is critical. Using data and tools available on the St. Jude Cloud Genomics Platform, we have identified and characterized fusion proteins found in Acute Megakaryoblastic Leukemia patients. Since intrinsically disordered regions (IDRs) are thought to be the driving force of phase separation capabilities of proteins, IUPRED3 was leveraged to further investigate features of the candidate fusion proteins.

Session Title: Cancer Poster Session I

PB4993 Characterizing selection signatures in coding and noncoding regions of 14,886 cancer genomes

Authors:

M. Zheng[1], S. Xiwei[1,2], Y. Jian[1,2]; [1]Westlake Univ., Hangzhou, China, [2]Westlake Lab. of Life Sci. and Biomedicine, Hangzhou, China

Natural selection plays a crucial role in cancer development, yet the extent of selection within the noncoding cancer genome remains largely unknown. Here, we developed a novel approach that integrates functional impact scores into the dN/dS framework, broadening the identification of selection signals to include not only coding regions but also the entirety of the genome. We applied our method to whole-genome sequencing (WGS) data from 14,886 cancer patients, revealing pervasive signatures of negative selection throughout the genome and co-occurrence of negative and positive selection signals in specific genomic regions. We identified 585 positively selected genes with an FDR < 0.05, of which 81% (474/585) have not been previously reported. Among the 111 known driver genes, 84.7% (94/111) are supported by additional selection signals in noncoding elements of the corresponding genes, suggesting a convergence of functionality between coding and regulatory variants in cancer pathogenesis. Moreover, we discovered a cluster of negatively selected genes enriched in essential and cancer-dependent genes, reflecting the vital roles these genes play in survival and proliferation of cancer cells. Taken together, our findings provide new insights into the role of selection during cancer cell evolution and present an extensive compendium of putative driver genes and noncoding elements with potential therapeutic target applications.

Session Title: Cancer Poster Session II

PB4994 Circulating LINE-1 and MicroRNA Profiles in Plasma Exosomes of Non-Small Cell Lung Cancer Patients Correlate with Histological Subtype, Early Diagnosis, and Clinical Prognosis: Implications for Precision Oncology

Authors:

A. Hassanin[1,2], K. Ramos[1]; [1]Texas A&M Inst. of BioSci.s and Technology, Houston, TX, [2]2Dept. of Animal Wealth Dev., Faculty of Vet. Med., Suez Canal Univ., Ismailia, Egypt

Liquid biopsy can provide insights into a patient's cancer status at successive time points, an essential domain for monitoring temporal tumor burden and early detection of tumor recurrence and treatment resistance. Extracellular vehicles, a new liquid biopsy analyte, are gaining growing interest in thoracic oncology as increased exosome levels have been detected in the body fluids of individuals with lung cancer. In the present study, we evaluated LINE-1 retrotransposon and microRNA profiles in plasma-derived exosomes from patients with non-small cell lung cancer (NSCLC). Both LINE-1 and a panel of eight miRNAs, namely, miR-21-5p, miR-126-3p, miR-210-3p, miR-221-3p, Let-7b-5p, miR-146a-5p, miR-222-3p, and miR-9-5p, were expressed early in NSCLC development in male and female patients with squamous cell lung cancers (SQCLC) and lung adenocarcinomas (LUAD), including metastatic tumors. Patterns of expression correlated closely with clinicopathological characteristics, which combined with ROC curve analyses, established the utility of LINE-1 levels in distinguishing tumor types, as well as the superior diagnostic performance of four miRNAs compared to single miRNA analytes. Ten genes within a discretized LINE-1 oncogenic regulatory network were validated as downstream targets of the eight plasma-derived exosomal miRNAs, confirming postulated mechanisms by which LINE-1 retrotransposons influence the occurrence of NSCLC. We conclude that measurements of the LINE-1 and miRNA load in plasma-derived exosomes from NSCLC patients may serve as precision-based, non-invasive, liquid biopsies for early diagnosis and clinical evaluation of NSCLC.

Session Title: Cancer Poster Session III

PB4995 *Cis*- and *trans*-eQTL TWAS of breast and ovarian cancer identify more than 100 risk genes in the BCAC and OCAC consortia.

Authors:

T. Head[1], F. Dezem[2], J. Plummer[2], A. Todor[1], J. Yang[1], S. Gayther[3], S. Kar[4], J. Schildkraut[1], M. Epstein[1]; [1]Emory Univ., Atlanta, GA, [2]St. Jude Children's Res. Hosp., Memphis, TN, [3]Cedars Sinai Med. Ctr., Los Angeles, CA, [4]Univ. of Cambridge, Cambridge, United Kingdom

Genome-wide association studies (GWAS) have identified numerous common risk variants for breast and ovarian cancer, with most GWAS-derived risk variants lying in non-coding regions of the genome. This has spurred considerable interest in transcriptome-wide association studies (TWAS) that aim to investigate how genetically regulated transcriptional activity plays a role in both the shared and distinct etiologies of these two complex diseases. However, the TWAS methods utilized to date have only considered the regulatory effects of variants located close to the target gene (*cis*-SNPs). With growing evidence for non-trivial distal regulatory effects of common variants (*trans*-SNPs) on gene expression, we performed TWAS of breast and ovarian cancer using a recently proposed Bayesian genome-wide method (BGW-TWAS) that incorporates both *cis*- and *trans*-expression quantitative trait loci (eQTL). We used whole genome sequencing and RNA sequencing data in breast and ovarian tissue from the Genotype-Tissue Expression Project V8 to train genome-wide gene expression imputation models with BGW-TWAS. We then used these imputation models to perform a TWAS on large-scale GWAS summary statistic data from the Breast Cancer and Ovarian Cancer Association Consortia (BCAC and OCAC) to identify genes whose genetically predicted expression levels are associated with risk of breast cancer overall, five breast cancer subtypes (luminal A-like, luminal-B like, luminal B/HER2-negative-like, HER2-enriched-liked, triple-negative), non-mucinous epithelial ovarian cancer, and five ovarian cancer histotypes (high grade serous, low grade serous, mucinous, endometrioid, clear cell). We identified 101 significant genes across breast cancer phenotypes and 8 across ovarian cancer phenotypes. These loci include several novel loci such as *ACAP3* (overall breast cancer risk) whose associations are predominantly driven by *trans*-eQTL effects. These *trans*-driven genes were not identified by a competing TWAS approach that considered *cis*-eQTL only. We replicated several gene associations for breast phenotypes using summary statistics from an independent GWAS of these cancers. We further used genotype and gene expression data in both normal breast and tumor tissue from the Cancer Genome Atlas to explore the robustness of our genome-wide imputation models. Results provide further insight into the complex genetic architecture underlying risk of breast and ovarian cancer subtypes. Efforts to validate top *trans*-eQTL using genome-wide chromatin interaction data are ongoing.

Session Title: Cancer Poster Session I

PB4996 Clinical consequences of a genetic predisposition toward higher benign prostate-specific antigen levels

Authors:

M. Shi[1], J. Shelley[2], K. Schaffer[1], J. Tosoian[1], M. Bagheri[1], J. Witte[3], L. Kachuri[4], J. Mosley[1]; [1]Vanderbilt Univ. Med. Ctr., Nashville, TN, [2]Vanderbilt Univ. Sch. of Med., Nashville, TN, [3]Stanford Univ., Stanford, CA, [4]Stanford Univ., San Francisco, CA

Background: Prostate-specific antigen (PSA) is a screening biomarker for prostate cancer, and a PSA>4 ng/mL may prompt a diagnostic evaluation to look for cancer. PSA levels are influenced by genetic variation unrelated to prostate cancer risk. We leveraged the VUMC biobank (BioVU) data to investigate whether a genetic predisposition to a higher PSA influences diagnostic assessments and outcomes. **Methods:** We developed a polygenic score (PGS) comprising 111 SNPs associated with PSA levels, but not with prostate cancer risk after accounting for selection bias due to screening. We curated two cohorts of male participants ages 45-70 years without a history of prostate cancer. The "biopsy cohort", comprised 468 men (407 European [EA] and 61 African [AA] ancestry) who underwent a first prostate biopsy for an elevated PSA, was used to test PGS associations with prostate cancer detection on biopsy. The "longitudinal cohort" was used to evaluate incident clinical outcomes associated with the PGS, and comprised 3,110 (2,118 EA and 992 AA) men with a baseline PSA<4 ng/mL undergoing routine PSA screening. Multivariable Cox proportional hazards models were used to test whether the PGS was associated with an incident: 1) PSA value>4 ng/mL, 2) International Classification of Diseases (ICD) code for an elevated PSA, 3) encounter with a urologist, or 4) prostate biopsy. Analyses were stratified by age (45-59 years and 60-70 years). Association estimates were per standard deviation increase in the PGS. **Results:** In the biopsy cohort, prostate cancer was identified on biopsy in 236 men (50%), of which 108 (23%) had a high grade (GG≥2) cancer. The PGS was significantly inversely associated with the outcome of detecting any prostate cancer (OR=0.77 [95% CI, 0.62 - 0.95], p=0.02) and high-grade cancer (OR=0.70 [0.55 - 0.89], p=0.004) on biopsy. In the longitudinal cohort, 2,118 (68%) participants were aged 45-59 years and the median (IQR) baseline PSA level was 1.0 (0.6 - 1.7) ng/mL. Among men 45-59 years, the PGS was associated with a PSA>4 (hazard ratio [HR]=1.35 [95% CI, 1.17-1.57], p=4.5x10-5), a clinical diagnosis of an elevated PSA (HR=1.30 [1.12-1.52], p=8.0x10-4), an evaluation by a urologist (HR=1.34 [1.14-1.57], p=4.8x10-4), and undergoing a prostate biopsy (HR=1.35 [1.11-1.64], p=0.002). Among men 60-70 years, association effect sizes were attenuated and not statistically significant. **Conclusions:** A polygenic predisposition toward higher benign PSA levels was associated with a decreased likelihood of detecting prostate cancer on biopsy, but an increased likelihood of undergoing a diagnostic evaluation including invasive procedures for an elevated PSA among 45-59 years.

Session Title: Cancer Poster Session II

PB4997 Clinical usefulness of multi-gene panel testing in hereditary cancer analysis in Japan

Authors:

M. Yoshida, E. Takamine, M. Egawa, H. Kohbata, M. Komine, S. Takahashi, Y. Ebana; Tokyo Med. and Dental Univ., Tokyo, Japan

Introduction: Multi-gene panel testing (MGPT) is now major genetic testing used in hereditary cancer in the US. However, single gene testing (BRCA1/2, RB1, MEN1, TSC1/2, and RET) is still predominate in Japan partly due to the national universal health insurance that covers around 70% of the medical costs. We thus decided to demonstrate the clinical usefulness of MGPT in hereditary cancer analysis in Japan to see if it would be feasible and acceptable in the Japanese genetic medicine environment. **Method:** Participants were recruited at the genetics department from January 2019 to September 2022 at the TMDU hospital, Kameda General Hospital and Kameda Kyobashi clinic. Those who were suspected of high risk of hereditary cancer by the physicians at the genetics department were eligible regardless of their personal cancer history. Informed consent was obtained prior to the testing. Two different sizes of MGPT were included (47 or 84 genes) in the study and either size was chosen based on the participants' personal and family history. The test fee was paid by the participants since the Japanese universal health insurance does not cover MGPT for hereditary cancer. **Results:** A total of 94 people aged between 0 and 89 participated in the study. Of those, 56 and 38 underwent the 47-gene panel and the 84-gene panel respectively. For the 47-gene panel, 16 participants (28.5%) had positive results. At least one variant of uncertain significance (VUS) was detected in 36 participants (64.2%) and, among them 26 participants (46.4%) had only VUSs for the initial test reports. For the 84-gene panel, seven participants (18.4%) had positive results and 27 (71.1%) participants had at least one VUS, with 25 (65.5%) of which had only VUSs for the initial test reports. In total, seven variants were updated later with four being likely benign or benign from VUS, and two being pathogenic from likely pathogenic. The genes which came back with positive results were ATM, BRCA1, BRCA2, CDH1, MLH1, MSH2, MSH6, PALB2, PTEN, RAD50, RAD51D, RECQL4, TP53, VHL. The turnaround time which included shipping from Japan to San Francisco ranged from 6 to 40 days with a median of 10.5 days. **Discussion:** Our findings demonstrate that MGPT in Japan is as acceptable and useful as in other countries, despite its unique national health insurance constraints. However, the VUS rate was over 64% even with the smaller panel. Therefore, it is important for genetic professionals to explain carefully to clients what VUSs mean and how they are treated.

Session Title: Cancer Poster Session III

PB4998 Clonal dynamics of *TP53* mutations in male germline cells and implications for inheritance

Authors:

R. Sanghvi[1], M. Neville[1], A. Lawson[1], P. Nicola[1], F. Abascal[1], T. Bayzetinova[1], M. Jorgensen[2], R. Rahbari[1]; [1]Wellcome Sanger Inst., Hinxton, United Kingdom, [2]Great Ormond Street Hosp., London, United Kingdom

The origin of heritable genetic variation and disorders lies in mutations that occur in the germline. To understand evolution and disease, it is crucial to comprehend the mechanisms behind germline mutations and the selection pressures that shape them. Inherited cancers play a major role in about 5 to 10 percent of all cancers and exert a devastating toll on the probands, the wider families and future generations. Li-Fraumeni Syndrome (LFS) is an inherited, autosomal-dominant cancer predisposition syndrome caused by a mutated *TP53* gene, leading to a high risk of wide-spectrum of tumours presenting throughout different stages of life.

Approximately 7-20% of germline mutations in *TP53* in individuals with LFS occur *de novo* in the germ cells of the parents, with the remainder following typical patterns of maternal or paternal inheritance. In this study, we analysed whole genome sequences from the nuclear families (parents and offsprings) diagnosed with Li-Fraumeni syndromes. We found 25% of cases are through *de novo* mutation in *TP53*, while the rest were inherited. We identified a median of 83 mutations per LFS child (ranging from 51 to 116) correlating with parental age at birth and in line with the previously published trio studies. Mutational signatures were consistent with previously reported germline signatures i.e. SBS1 and SBS5 (with median contributions of 15% and 85% respectively). Additionally, we investigated the mutational landscape and clonal dynamics of mutant *TP53* in sperm of the fathers (age ranging from 34 to 62 years-old) in the *de novo* LFS subset, using a deep targeted sequencing. In the majority of sperm samples we did not observe any expansion of pathogenic *TP53* except in one father in whom we observed 2.5% of sperm had pathogenic *TP53*, p.V272M.

These findings have important implications for genetic counselling on recurrence risks for families with LFS caused by *de novo* mutations. Understanding the degree of penetrance and cause of *de novo* mutations in *TP53* in germline cells will help develop prevention strategies for cancer predisposition syndromes in offspring.

Session Title: Cancer Poster Session I

PB4999 Clonal hematopoiesis, the aging genome and mantle cell lymphoma.

Authors:

H. Hill[1,2], P. Jain[2], C. Y. Ok[2], S. Vinay[1], M. Wang[2], M. Kimmel[1], K. Chen[2]; [1]Rice Univ., Houston, TX, [2]Univ. of Texas MD Anderson Cancer Ctr., Houston, TX

Clonal hematopoiesis of indeterminate potential (CHIP) describes clonal expansion of hematopoietic stem cells which contribute to distinct mutations and are found in pre-malignant and malignant states. Hemopoietic cells are found in the blood and bone marrow, but infiltrate all tissues. CHIP is thought to be a process of the aging genome. However, the prevalence of CHIP mutations is linked to cardiovascular disease, myeloproliferative disease and various cancer types independent of age.

Mantle cell lymphoma (MCL) is a rare, incurable type of B-cell lymphoma. MCL occurs frequently in older individuals with a median age of diagnosis of 69 years. Males are disproportionately affected; the male-to-female ratio of MCL diagnosis is approximately 4:1.

We identified seven CHIP related mutations (*TET2*, *DNMT3A*, *GNAS*, *ASXL1*, *JAK2*, *SF3B1*, *PPM1D*) in pre-treatment MCL patient (n=253) samples using targeted NGS sequencing of peripheral blood, bone marrow, lymph node and gastrointestinal biopsies. Clinical, cytogenetic and environmental features were extracted from electronic health records (EHR) for correlative studies.

Of the full cohort, 28.06% (n=71) had one or more CHIP variant: *TET2* (30%, 51 nonsynonymous variants/170 tested), *ASXL1* (11.2%, 19/170), *JAK2* (5.2%, 9/173), *DNMT3A* (3.5%, 6/173), *SF3B1* (1.2%, 3/259), and *GNAS* (1.2%, 2/173). *GNAS* and *JAK2* mutations were found only in peripheral blood and bone marrow, however enrichment of mutations in any sample type was not significant. Three patients had identical CHIP variants observed in more than one sample type. It's unknown if these were germline variants or clonal somatic mutations. CHIP variants were represented in population databases with population allele frequency (PAF) ranging from <1-10%. However, several mutations were identified that weren't in any databases or associated with pathogenicity. Two genes had congruent mutations occurring in more than one MCL patient (*GNAS:* c.602G>A and *JAK2*: c.1849G>T).

We filtered CHIP mutations by PAF in public genome databases and designated them "somatic or of unknown origin". These mutations were associated with age ($p < 0.01$) and progression-free survival (Log-Rank, $p = 0.045$). Patients with these mutations were also more likely to have had prior cancer diagnoses (OR = 3.33, p=0.014). In a Cox regression model, *TP53* mutations and chromosome 13 deletions were also significant in predicting PFS. CHIP mutations were reflected in the AACR Project GENIE database in MCL patients (n=453): *TET2* (4.5%), *DNMT3A* (4.3%), *SF3B1* (2.0%) *and ASXL1*(0.4%). This and other databases will be used for meta-analyses and comparative studies with our patient cohort.

Session Title: Cancer Poster Session II

PB5000 Combatting cisplatin resistance through induced gene expression changes: Trichostatin A as a candidate.

Authors:

B. Waldmann, A. Stark; Univ. of Notre Dame, Notre Dame, IN

One of the main challenges in treating cancer is how cancer cells have an ability to resist treatment. Some cancer cells do this through intrinsic resistance while others develop resistance over time. One particular chemotherapeutic agent, cisplatin, is vulnerable to both acquired and intrinsic resistance. To counter the induced expression changes by the cancer cells, other drugs can be combined with cisplatin to diminish chemoresistance. One such drug: Trichostatin A, has been considered in combination treatment with cisplatin for increased effectiveness. Trichostatin A is a HDAC inhibitor which helps reduce cancer proliferation by causing an increase in accumulation of acetylated histones. One important protein to the proliferation of cancer cells in HIF-1α. HIF-1α is crucial to chemoresistance of several cancer lines (including breast cancer) as it plays a key role in cancer cell survival. Trichostatin A, at a concentration of 100nM, was found to effectively reduce HIF-1α expression in MCF-7 cancer cells. Trichostatin A treatment of MCF-7 cells led to a reduction of expression of 8.9% (16 hours), 26.5% (39 hours), and 41.1% (63 hours). Trichostatin A's ability to reduce the expression of HIF-1α suggests a promising avenue to combat chemoresistance to cisplatin treatment. Further evaluation of additional candidates, time course, and synergistic effects with cisplatin treatment are currently ongoing. The goal is to decrease cancer cell's chemoresistance to cisplatin by using induced gene expression, which may have broad applications to other chemotherapeutic agents.

Session Title: Cancer Poster Session III

PB5001 Common and distinct patterns of acquired uniparental disomy and homozygous deletions between lung squamous cell carcinomas and lung adenocarcinoma,

Authors:

M. Tuna[1], G. B. Mills[2], C. I. Amos[1]; [1]Baylor Coll. of Med., Dan L Duncan Inst. for Clinical and Translational Res., Houston, TX, [2]Oregon Hlth.& Sci. Univ., OHSU Knight Cancer Inst., Portland, OR

Acquired uniparental disomy (aUPD) is a form of allele-based alteration that can lead to homozygosity of existing aberrations. We used data from The Cancer Genome Atlas SNP-based arrays to identify distinct and common aUPD profiles for LUAD and LUSC and their impact on homozygous deletion (HD), overall survival (OS), and recurrence-free survival (RFS). Overall, we found significantly higher aUPD (q = 5.34E-09) in LUSC than in LUAD. A significant portion of HD was due to aUPD in LUSC (24.9%) and LUAD (19.7%). We identified segmental, whole-chromosome arm and whole-chromosome aUPD, in which whole 7p arm aUPD was restricted to LUSC, while whole-chromosome 3 aUPD was observed only in LUAD, and whole-chromosome 21 aUPD was common to both LUSC and LUAD. The most frequent aUPD and HD were observed at *CDKN2A/B* region in both LUAD and LUSC. In LUAD, aUPD and HD at *CDKN2A/B* region were associated with shorter OS ($q < 0.021$ and $q < 0.005$), and RFS ($q < 0.005$ and $q < 0.005$), while heterozygous deletion was not associated with OS and RFS. In contrast, no association was found between aUPD at *CDKN2A/B* region and survival in LUSC. CTL4A expression was significantly lower in samples with aUPD at *CDKN2A/B* regions than in samples without copy number and allele-based changes in LUAD. Immune infiltration correlates with whole-chromosome *q*-arm or whole chromosome aUPD in LUAD. Both LUSC and LUAD have common and distinct patterns of aUPD regions with differing frequencies of occurrence and associations with outcome.

Session Title: Cancer Poster Session I

PB5002 Comparing Breast Cancer Medical History between Participants surveys and Electronic Health Records in All of Us Research Program

Authors:

L. Bastarache[1], E. Venner[2], L. Sulieman[1]; [1]Vanderbilt Univ. Med. Ctr., Nashville, TN, [2]HGSC - Baylor Coll. of Med., Houston, TX

About 55-72% and 45-69% women carrying BRCA1 and BRCA2 variants will be diagnosed with breast cancer (BC) by age of 70-80. Electronic health records (EHR) linked to genetic data can be used to measure the penetrance of genetic variants, including BRCA, by computing the percentage of patients with a pathogenic variant and BC diagnosis. However, EHR is an imperfect medical history source, as they are subject to missingness and fragmentation leading to underestimating the true penetrance of a genetic variant.
Surveys offer alternative ways of collecting patients/participants' medical history (PRMH). Prior work showed disagreement between EHR and surveys; patients with a diagnosis record in EHR do not always report that diagnosis in a survey, and vice versa. More research is needed to know 1) how EHR missingness affects penetrance estimate 2) how to use PRMH with EHRs to ascertain diagnosis more accurately.
To address the knowledge gap, we compared BC history (BCH) between EHR and PRMH in BRCA carriers in the All of Us Research Program, a national initiative to collect longitudinal biomedical data from historically underrepresented populations. We assessed EHR BC missingness in female participants who had BRCA mutations by comparing EHR to PRMH. In latest release (CDR-7), we identified female participants with pathogenic/likely pathogenic BRCA variants, extracted EHR BC diagnosis codes, and identified female participants who reported BCH in medical history survey (MHS). CDR-7 had 248135 female participants, 145563 females had genome data, 52% were white. Among those, 709 and 2447 had BRCA1 and BRCA2 variants, with 48.6 as mean age. 50115 females filled MHS (mean age 49.2) where 10418 (20.8%) reported BCH.
In BRCA1 cohort, we found substantial discordance between BCH in MHS and EHR. Of 189 participants who completed MHS, 94 (49.7%) reported BCH. 46 (48.9%) participants reported BCH but lacked EHR BC codes. In BRCA2 cohort, 527 filled MHS and 180 (34.1%) confirmed BCH where 81 participants reported BCH and had BC codes, while 99 (55.0%) confirmed BCH and lacked EHR BC codes. In BRCA1 and BRCA 2, 13 and 32 participants had EHR but didn't report BC in MHS.
Our results show that BC prevalence in participants with BRCA mutations is slightly lower than nationally reported values possibly due to either having PRMH for a subset of participants or lower age at enrollment (<50 years on average). Almost half of participants who reported BCH lacked EHR BC codes. EHR missingness is a major issue that can bias genetics studies if not controlled properly. We showed possible use of genetic variants to quantify EHR missingness. Moreover, PRMH represent valuable source to estimate accurate penetrance.

Session Title: Cancer Poster Session II

PB5003 Comprehensive germline and somatic testing for precision risk-assessment for children with retinoblastoma and their siblings

Authors:

F. Taylan[1,2], M. Fili[3,4], K. Lagerstedt Robinson[1,2], T. Diaz de Ståhl[5,6], S. Wessman[5,6], B. E. Damato[3], G. Stålhammar[3,4], K. Henning[7,8], T. Frisk[7,8], P. Grillner[7], J. Sandgren[5,6], G. Ljungman[9], D. Gisselsson[10,11], A. Nordgren[1,2,12,13], B. Tesi[1,2,14], E. Tham[1,2]; [1]Dept. of Molecular Med. and Surgery, Karolinska Inst.t, Stockholm, Sweden, [2]Dept. of Clinical Genetics, Karolinska Univ. Hosp., Stockholm, Sweden, [3]Dept. of Ocular Oncology, St. Erik Eye Hosp., Stockholm, Sweden, [4]Dept. of Clinical NeuroSci., Div. of Ophthalmology and Vision, Unit of Ocular Oncology and Pathology, Karolinska Inst.t, Stockholm, Sweden, [5]Dept. of Oncology-Pathology, Karolinska Inst.t, Stockholm, Sweden, [6]Dept. of Pathology and Cancer Diagnostics, Karolinska Univ. Hosp., Stockholm, Sweden, [7]Dept. of Pediatric Oncology, Astrid Lindgren Children's Hosp., Karolinska Univ. Hosp., Stockholm, Sweden, [8]Childhood Cancer Unit, Dept. of Women's and Children´s Hlth., Karolinska Inst.t, Stockholm, Sweden, [9]Dept. of Women's and Children's Hlth., Uppsala Univ. Hosp., Uppsala, Sweden, [10]Div. of Clinical Genetics, Dept. of Lab., Med., Lund Univ., Lund, Sweden, [11]Med. Services, Section for pathology, Region Skåne, Lund, Sweden, [12]Dept. of Lab. Med., Inst. of Biomedicine, Univ. of Gothenburg, Gothenburg, Sweden, [13]Dept. of Clinical Genetics and Genomics, Sahlgrenska Univ. Hosp., Gothenburg, Sweden, [14]Dept. of Med. Huddinge, Ctr. for Hematology and Regenerative Med., Karolinska Inst.t, Stockholm, Sweden

Background: Retinoblastoma (Rb) is a rare form of eye cancer that mainly affects children under the age of 5 years. Hereditary Rb is caused by germline disease-causing variants (DCVs) in the tumor suppressor RB1, which is also inactivated in non-hereditary cases. Not all germline DCVs are detected by current genetic tests, so siblings of a proband with Rb without detected germline *RB1* DCVs require intensive ocular surveillance due to residual risk. As part of the prospective nationwide Genomics Medicine Sweden childhood cancer initiative, a universal testing strategy, including germline and tumor DNA whole genome sequencing (WGS), has been offered to all children with solid malignancies at diagnosis and relapse since 2021 to refine diagnosis, individualize treatment and surveillance as well as enable identification of at-risk relatives. This study presents our genetic results for children with Rb.
Methods: Thirteen children between 3 months and 4 years old, diagnosed with Rb (unilateral n=12, bilateral n=1) at St. Erik Eye Hospital in Stockholm, underwent germline WGS (gWGS) between May 1, 2021 and April 30, 2023. The gWGS data were analyzed with an *in silico* childhood cancer predisposition gene panel (n=189), including the *RB1* gene. WGS from fresh frozen tumor tissue (tWGS) was performed on 5 cases, while exome sequencing was performed on 5 other cases for which only formalin-fixed paraffin-embedded tumor samples were available. In the remaining 3 cases, tumor material was not available.
Results: We identified 6 patients with a germline finding in *RB1*. No germline DCV was detected in the other 188 genes. Of the 7 patients without germline *RB1* DCV, 5 had somatic mutations in both alleles of *RB1*. All cases had at least one somatic aberration of a structural type. Cases with two somatic *RB1* mutations without germline DCV were considered sporadic or potentially low-grade mosaic. In total 5 of the 13 cases did not confer risk for siblings, who could therefore avoid intensive surveillance. All findings were discussed at multidisciplinary rounds and reported to the treating physicians. We also offered genetic counseling to the families.
Conclusion: Overall, gWGS, combined with patient-matched tWGS, allows characterization of *RB1* variants, providing precise risk assessment for children with Rb and carrier testing of parents and future offspring. This approach also identifies at-risk siblings within the family and reduces unnecessary surveillance for existing or future siblings in cases where 2 somatic, and no germline, hits in *RB1* are detected. Our results highlight the effectiveness of paired germline-tumor analysis when interpreting germline and somatic variants in Rb.

Session Title: Cancer Poster Session III

PB5004 Computational epigenetic analysis of colorectal cancer identifying Key regulators of genomic and transcriptional dysregulation with chip-seq method.

Authors:

M. HABIBI[1], K. GHDIRA[2]; [1]Faculty of medicine od Tunis (TUNISIA), Ben Arous, Tunisia, [2]Inst. pasteur of tunis, Tunis, Tunisia

Colorectal cancer (CRC) is one of the most prevalent malignant tumors worldwide, characterized by diverse clinical phenotypes and responses to current treatments. The considerable heterogeneity between patients is one of the most important features of CRC. Efforts have been made in recentyears to genetically and phenotypically classify genetically and phenotypically diverse CRC tumorsinto distinct molecular subtypes based on gene expression profiling. Epigenetic dysregulation hasbecome a paradigm of cancer biology that underlies the characteristics of cancer cells. Globalchanges in DNA methylation, chromatin states, and cis-regulatory elements, as well as geneticaberrations in chromatin proteins, characterize more than 50% of human cancers.The identification of precise activators via global mapping of histone modifications contributes tothe cellular reprogramming towards tumor growth and metastasis, making activator dysfunction apromising diagnostic biomarker and potential therapeutic target. A systematic study of chromatinmodifications in cancer and pioneering reports on the molecular mechanisms underlyingtranscriptional addiction have begun to emerge. However, defining the chromatin states thatcharacterize human pathology remains a challenge, partly due to the limited number of cellsavailable from primary tumors. Another critical problem in the study of the epigenetic profiles ofindividual tumors is the use of appropriate research platforms that capture the intrinsic profiles ofcancer cells.In this study, we aim to develop a computational pipeline for epigenetic analysis of genomic andepigenetic events mediating colon cancer and provide clues on the main regulators that orchestrategenomic and transcriptional dysregulation on which cancer cells in colorectal cancer depend. Toachieve this goal, we will identify significant transcription factors relevant to colorectal cancer andexamine the different levels of variation as well as study the variation in copy number and DNAmethylation of these factors. We will also identify potential cis-regulatory non-coding RNAs andnon-coding RNAs that regulate transcription factor activity. Using ATAC-Seq and ChIP-Seqprofiles, we can detect target genes to infer intrinsic associations of target genes based on Bayesiannetworks. We aim to identify potential biomarkers, if possible. The bioinformatics approach ofChIP-Seq will be used to study the interaction between proteins and DNA.

Session Title: Cancer Poster Session I

PB5005 Contribution of Driver Genes in Classical Myeloproliferative Neoplasms (MPNs): Report from developing country

Authors:

M. Rashid[1], R. Ahmed[2], U. Zaidi[1], T. Shamsi[1]; [1]NIBD, Karachi, Pakistan, [2]Univ. of Karachi, Karachi, Pakistan

Background: Classical Myeloproliferative neoplasms (MPNs) are characterized by excessive production of terminally differentiated blood cells that are fully functional. The genetic landscape of MPNs is in large part elucidated. **Materials & Methods:** Total 133 MPNS patients, including 40 polycythaemia vera (PV), 50 essential thrombocythaemia (ET) and 43 primary myelofibrosis (PMF) were investigated for the presence of *CALR* exon 9 by Sanger sequencing and *JAK2*V617F mutations by ARMS-PCR. **Results:** *JAK2*V617F mutation was found significantly (P-value = 0.000) more common in PV (72.5%; 29/40) then in ET (42%; 21/50) and PMF (23.2%; 10/43). Mutations in *CALR* exon 9 were found significantly (P-value = 0.000) more common in ET than PMF with frequencies 38% (19/50) and 20.9% (9/43), respectively whereas absent in PV (0/40). Incidence of *CLAR* mutation in non-*JAK2*V617F mutated ET and PMF combine was 43.5% (27/62), where in ET it was 62% (18/29) and in PMF it was 27.2% (9/33). In PV *JAK2*V617F mutated patients were significantly older than wild type patients and had high platelets count. In ET significantly (P value = 0.009) high platelet count (1272 x 10^9/L mean ±615 SD) and low leukocyte count (8.4 x 10^9/L mean±2.7 SD) (P value = 0.017) were seen in *CALR* mutated ET as compare to wild type. In PMF significantly (P-value = 0.009) high Hb (11 g/dl mean±1.3 SD) and high platelets count (472 x 10^9/L mean ±432 SD) (P value = 0.005) were observed in *CALR* mutated PMF than wild type. **Conclusion:** This study emphasizes on molecular diagnostic approach to adopt in classical MPNs.

Session Title: Cancer Poster Session II

PB5006 Copy Number Variations (CNVs) Contribute to Malignant Tumors Development in Birth Defects Children

Authors:

Y. Liu[1], J. Glessner[2], H. Qu[3], X. Chang[1], H. Qiu[4], T. Wang[1], K. Nguyen[2], F. Mentch[3], X. Wang[2], A. Saeidian[5], D. Watson[3], H. Hakonarson[3]; [1]Children's Hosp. of Philadelphia, Philadelphia, PA, [2]Children s Hosp. of Philadelphia, Philadelphia, PA, [3]Children's Hosp. of Philadelphia, Philadelphia, PA, [4]Children's Hosp. of Philadelphia CHOP, Philadelphia, PA, [5]The children's Hosp. of philadelphia CHOP, Philadelphia, PA

There are two key signatures of pediatric cancers when compared to adults: higher prevalence of germline alterations and heterogeneity in the alteration types. Recent population-based assessments have demonstrated that children with birth defects (BD) are more likely to develop cancer even without chromosomal anomalies; therefore, explorations of genetic alterations in BD children with cancers could provide insights into the underlying mechanisms for pediatric tumor development. In this study, we performed whole genome sequencing (WGS) on blood-derived DNA samples from 1566 individuals without chromosomal anomalies, including 454 BD probands with at least one type of malignant tumor, 767 cancer-free BD children, and 345 healthy individuals, focusing on copy number variations (CNV), one of the most important structural alternations observed in cancer. Roughly half of the BD-cancer children have CNVs that are not identified in BD-only/healthy individuals, and those CNVs are not evenly distributed among the patients. Strong heterogeneity was observed, given that only a few cancer predisposition genes contain CNVs in more than three patients, including deletions in KCND2, SDK1, SP4, and duplications in NEURL1B. Moreover, functional enrichment of genes harboring CNVs show that a large number of patients cluster on the same biological pathways, such as deletions of genes with neurological functions, and duplications of immune response genes. Phenotype clustering uncovered recurrences of patients with sarcoma, such as tumors in bones or soft/connective tissues. A notable observation is that non-coding RNAs, including microRNAs and long non-coding RNAs (lncRNAs), which we identified as gene regulators, show strong signals related to growth and cancer regulations in functional enrichment analysis. In conclusion, we conducted one of the first genomic study to explore the impact of CNVs in cancer developments in children with BD, providing new insights into the underlying biological processes, especially the impact CNVs play in regulating non-coding RNAs.

Session Title: Cancer Poster Session I

PB5008 Deciphering anti-ovarian cancer associated mechanistic regulatory pathways of indigenous Saudi Arabian plants using transcriptomic systems biology and bioinformatics approaches

Authors:

A. Alblihy; KFSC, Riyadh, Saudi Arabia

Ovarian cancer (OC) is the deadliest cancer among women and ranks fourth among all fatal diseases in women. In 2019, the estimated number of new cancer cases was 222,240 and the number of deaths was approximately 14,170. Previous research has suggested that prolonged use of unopposed estrogen and estrogen plus progesterone is associated with an increased risk of OC, particularly in menopausal women. The lack of efficient early detection methods, increased chemotherapy resistance, and the heterogeneity of the disease all contribute to the poor prognosis of most OC cases. Therefore, it is crucial to develop safer and more effective therapeutic strategies to improve the survival rate of women with OC. Sensing these opportunities, the current study integrates bioinformatics and network pharmacology approaches to explore the potential molecular mechanisms of local flora of Saudi Arabia including *Onopordum heteracanthum*, *Acacia ehrenbergiana*, *Osteospermum vaillantii*, *Cyperus rotundus*, *Carissa carandas*, *Carissa spinarum*, and *Camellia sinensis* in ovarian cancer treatment. Initially, the information on active constituents of these local plants and target genes of both liver cancer and plants were retrieved from literature as well as from publicly available databases. Based on the matching results between plant-potential targets and ovarian cancer targets, the protein-protein interaction (PPI) network was constructed using the STRING database and imported into Cytoscape for screening of hub genes based on their degree of connectivity. Later, the interactions network between compounds and overlapping genes was constructed using Cytoscape software to analyze the network pharmacological prospective effects of active compounds on ovarian cancer. Following that, a compound-target protein-pathway network was constructed which uncovered that namely hispidulin, stigmasterol, ascorbic acid, octopamine, cyperene, kaempferol, pungenin, citric acid, d-Tartaric acid, beta-sitosterol, (−)-epicatechin gallate, and (+)-catechin decisively contributed to the cell growth and proliferation by affecting AKT1 and VEGFA proteins. Moreover, the molecular docking and Molecular Dynamic (MD) simulation of 20ns well complemented the binding affinity of the compound and revealed strong stability of predicted compounds at the docked site. This research not only provides a theoretical and experimental basis for more in-depth studies but also offers an efficient method for the rational utilization of active compounds as anti-tumor drugs

Session Title: Cancer Poster Session II

PB5009 Deciphering disease-relevant cell types in prostate, breast, and lung cancers through integration of GWAS and single-nucleus RNA sequencing data.

Authors:

T. Harrison[1], E. Kaplan[2], K. Hou[3], C. M. Vachon[4], R. M. Tamimi[5], B. Pasaniuc[3], S. Lindstroem[1]; [1]Univ. of Washington, Seattle, WA, [2]Wesleyan Univ., Middletown, CT, [3]UCLA, Los Angeles, CA, [4]Mayo Clinic, Rochester, MN, [5]Weill Cornell Med., New York, NY

Hundreds of genetic variants have been identified as associated with prostate, breast, and lung cancers through genome-wide association studies (GWAS). However, the mechanisms by which they confer risk, and the tissues and cell-types involved in disease susceptibility are poorly understood. Single-cell and single-nucleus RNA sequencing (snRNA-seq) data provide potential insights into the tissues, cell types, and cell populations affected by disease-relevant variants. To identify populations of disease-associated nuclei, we utilized existing germline GWAS summary statistics for prostate (up to n=177,526), breast (n=133,384), and lung (n=85,716) cancer, and existing snRNA-seq data from GTEx (a cross-tissue set of 209,126 nuclei in n=8 tissues from a total of n=16 donors across tissues). Specifically, we mapped germline genetic variants to genes and calculated gene-specific z-scores. We then assessed if expression of the 1,000 most strongly associated genes were enriched across individual cell types using the single-cell disease relevance-score (scDRS). We calculated nuclei-level disease associated scores for 13 grouped cell types, and further evaluated 44 individual cell types to better understand which specific cells contributed to cancer associations. Our findings recapitulated known associations of prostate and breast cancer with cell-types, such as the involvement of epithelial cells in cancer risk. Additionally, we identified several genes for which expression was correlated with scDRS scores, but GWAS-based analyses were not statistically significant (p-value > 0.05). We demonstrate that GWAS-derived signatures from prostate, breast, and lung cancer data can be used to identify enrichment of disease-related expression from snRNA-seq data at the individual cell level. Further scDRS may offer unique insights into genes relevant to disease that have not be identified using GWAS statistics alone.

Session Title: Cancer Poster Session III

PB5010 Decoding the Genetic Landscape of Thyroid Cancer: Identifying Critical Single Nucleotide Polymorphisms for Improved Diagnosis and Prevention

Authors:

J. Ke; Dougherty Valley High Sch., San Ramon, CA

In our study, we explored the genetic intricacies of thyroid cancer, a complex endocrine malignancy, with a primary focus on identifying specific single nucleotide polymorphisms (SNPs) associated with the disease. The multifaceted nature of thyroid cancer, marked by a diversity of genetic variations, underscores the importance of research aimed at decoding its genetic architecture. Our investigation began with a focused examination of the NRAS gene region, noted for its documented significance in thyroid cancer pathogenesis. However, recognizing the potential of a broader genetic landscape, our study expanded to encompass the entirety of chromosome 1. We used Genome-Wide Sequencing data from the Sequence Read Archive and constructed robust pipelines to align sequences against chromosome 1 using the Bowtie 2 tool. The subsequent process involved indexing and variant calling. Our stringent protocols resulted in the identification of over 700 statistically significant SNPs within the thyroid cancer cohort. Interestingly, these SNPs were notably absent in the non-cancer cohort. This revelation accentuates possible genetic foundations of thyroid cancer and emphasizes the genetic disparity between cancer-affected and unaffected cohorts. Our findings suggest that the identified SNPs could be useful in developing innovative strategies for early thyroid cancer diagnosis and prevention.

Session Title: Cancer Poster Session I

PB5011 Detection of actionable variants using liquid biopsy samples for therapeutic approach.

Authors:

S. Temel, B. DOGAN, N. Kaya, N. Nalbant, A. Alemdar, L. Aliyeva, S. Ozemri Sag; Bursa Uludag Univ., Bursa, Turkey

Liquid biopsy is used for NGS-based somatic profiling that enables a precision oncology approach and is now showing its advantages over classic tissue biopsy. By bringing NGS into a clinical context, revealing treatable alterations in multiple genes may help clinicians to assess novel options for diagnostic, prognostic and therapeutic strategies. The current study presents analyses of targeted NGS gene panels to detect actionable somatic variants in patients using the GeneReader NGS System. In this study, 203 patients who were admitted to our clinic between 2019 March and 2021 January with different cancer syndromes (mainly lung & colorectal cancer) were screened using liquid biopsy samples with GeneRead QIAact Lung DNA Panel UMI Kit. This panel is designed to enrich specific target regions in selected genes and covers 550 variants across these 19 genes and targets mostly lung cancer-relevant mutations (SNVs, InDels and, CNVs). Targeted sequence analysis was performed via GeneReader NGS System. All detected somatic variants can be further interpreted by QIAGEN Clinical Insight (QCI) analysis and using online tools (My Cancer Genome, cBioPortal, OncoKB, Franklin, HGMD, ClinVAR) and ACMG criteria. This study included 203 patients with a mean age of 64 (149 male and 54 female) 124 were metastatic, 39 were non-metastatic, and 40 were not determined. The most common cancer types in the patient group are patient lung related-cancer (*n*=144), colorectal-related cancer (*n*=26) and other cancer types (*n*=33), respectively. In 203 patients, 65 actionable somatic variants, 3 actionable variants with amplifications, 14 amplifications, and 2 deletions were annotated via QCI. 17/68 heterozygous variants showed over %40 variant allele frequency evaluated. The 3 genes with the most frequent variant changes were *EGFR* (23/68), *ERBB2* (16/68) and *KRAS* (12/68). This study shed light on the importance of liquid biopsy standards for NGS-based testing, mainly the identification of actionable variants impact on the clinical implications. For tumor profiling tissue biopsy is always the gold standard for initial diagnosis but in case of insufficient tissue and follow of treatment response liquid biopsy can be an option.

Session Title: Cancer Poster Session II

PB5012 Detection of large inversions involving the Lynch syndrome gene *PMS2* using Cas9-assisted targeted sequencing.

Authors:

C. Boles[1], T. Huang[2], F. Chen[2]; [1]Sage Sci., Inc., Beverly, MA, [2]Illumina, Inc., Foster City, CA

Deleterious alterations in the *PMS2* mismatch repair gene on chromosome 7 (band p22.1) are associated with two hereditary cancer syndromes, constitutional mismatch repair deficiency (CMMRD), and Lynch syndrome (LS). Genetic tests for *PMS2* alterations are complicated by the presence of 15 *PMS2* pseudogenes scattered across chromosome 7, and a 100kb low copy number repeat (LCR) sequence that covers the 3' portion of the *PMS2* gene (exons 9-15). A second copy of the LCR sequence in inverted orientation is located 700kb toward the centromere, and shares greater than 95% homology with the telomeric copy, including a *PMS2CL* pseudogene. Inversion events between the LCRs have been shown to cause *PMS2* gene activation, but are extremely difficult to detect by next generation sequencing methods due to the great length and sequence homology between the LCRs. We have used Cas9 endonuclease in vitro to selectively excise and electrophoretically purify a 200kb genomic DNA fragment spanning the entire telomeric LCR plus about 50kb of unique flanking sequence on both sides of the LCR - a fragment that encompasses entire unrearranged *PMS2* gene. Purification of the 200kb fragment enables targeted long-read sequencing of the *PMS2* locus without complications from the centromeric LCR or the *PMS2* pseudogenes. To test for inversions involving the LCR regions, a different Cas9 digestion is performed using the same telomere-proximal cleavage site outside of the telomeric LCR, but using a new second cleavage site in unique sequence about 50kb inside of the centromeric LCR. This second Cas9 digestion will also liberate a 200kb product from the inversion allele, but the sequencing reads will map in a very different pattern from the unrearranged allele, with high coverage of the region between the second cleavage site and the border of the centromeric LCR - a region approximately 650kb away from the location of the unrearranged *PMS2* gene. We demonstrate the inversion detection method using Coriell cell lines from the Ashkenazi trio (HG002, HG003, HG004), in which the maternal haplotype inherited by the son carries the *PMS2* inversion.

Session Title: Cancer Poster Session III

PB5013 Developing a Machine Learning model for classifying ctDNA variants.

Authors:

R. Maruzani, A. Fowler, A. Jorgensen, L. Brierley; Univ. of Liverpool, Liverpool, United Kingdom

Circulating tumour DNA (ctDNA) has shown great potential as a minimally invasive biomarker for early diagnosis, personalised treatment, and disease monitoring in cancer patients. All tissues release fragmented cell free DNA (cfDNA) into the bloodstream via a range of mechanisms. Circulating tumour DNA is a subset of cfDNA originating from tumour tissues. Accurate detection of ctDNA variants in next-generation sequencing data is critical in realising the potential of ctDNA analysis as a minimally invasive cancer biomarker. However, ctDNA variants present a challenge for variant calling tools due to the low variant allele frequencies expected in ctDNA NGS data. Additionally, none of the established variant calling algorithms incorporate ctDNA specific features for variant detection, particularly the shorter fragment lengths and our previous work has shown a high level of discordance between variant calling tools. There is a need for computational tools for more accurate detection of true ctDNA variants. Here, we evaluate a range of machine learning algorithms for classifying true positive ctDNA variants from sequencing and PCR artifacts. Models were trained using publicly available ctDNA datasets, with matched tissue samples providing a truth set for low frequency variants. All models were tested using independent datasets. We leverage ctDNA fragment length information and ensemble variant calling as features in training models. Development of ML models for ctDNA variant analysis provide tools to improve detection and prioritisation of cancer mutations.

Session Title: Cancer Poster Session I

PB5014 Diagnostic Challenges: A case of acute promyelocytic leukemia with classical morphology and immunophenotype but cryptic *PML::RARA* rearrangement detected by RT-PCR

Authors:

G. George, M. Elsadawi, B. Zhang, A. G. Evans, **M. A. Iqbal**; Univ. of Rochester Med. Ctr., Rochester, NY

Acute promyelocytic leukemia (APL) is a subtype of acute myeloid leukemia, characterized by the predominance of abnormal promyelocytes and most often the classical t(15;17) (q22;q21) reciprocal translocation, resulting in the fusion of the promyelocytic leukemia (*PML*) and retinoic acid receptor alpha (*RARA*) genes (*PML::RARA*). The diagnosis of APL hinges on the demonstration of the *PML::RARA* fusion, and the classical translocation is present in 98% of cases. Rarely, patients may present with a variant rearrangement or cryptic insertion of *RARA* into *PML*. Combination therapy with all-trans retinoic acid (ATRA) and arsenic trioxide (ATO) is largely curative provided that the diagnosis of APL is established early cytogenetically. Although the incidence of APL in which the classical *PML::RARA* translocation cannot be demonstrated by fluorescence in situ hybridization (FISH) with targeted t(15;17) and RARA breakapart probes is low, it is important to identify these cases by alternate modalities to initiate immediate treatment. Herein, we report a case of APL diagnosed in a 36-year-old male who presented with spontaneous ecchymosis, petechiae, and exertional dyspnea. Laboratory investigations revealed low fibrinogen, prolonged prothrombin time, international normalization ratio, and activated partial thromboplastin time, and elevated D-dimer, indicative of disseminated intravascular coagulation. Peripheral blood examination demonstrated anemia, leukopenia, and thrombocytopenia. Scattered promyelocytes with occasional Auer rods and variable cytoplasmic granules were noted by microscopy. Flow cytometry revealed an abnormal cell population within the CD45-dim gate with high side scatter and classical immunophenotype (expression of CD13, CD33, CD38, CD117, CD123, CD64, and cytoplasmic myeloperoxidase (MPO), while negative for other markers). Although this case appeared characteristic by morphology, it was found to be negative for both the classical translocation by the routine t(15;17) *PML::RARA* probe and APL variants by the breakapart RARA probe. RT-PCR later revealed a cryptic *PML::RARA* fusion. Fortunately, the patient was started on ATRA/ATO early based on the morphological/immunophenotypic findings. This case underscores the importance of performing confirmatory RT-PCR in FISH-negative cases of APL to identify specific *PML:RARA* breakpoints. Moreover, revising the current paradigm to recommend upfront RT-PCR in all APL cases may mitigate the likelihood of missing these cases. A systematic literature review is underway to understand the prevalence, diagnosis, and prognosis of APL with cryptic *PML::RARA* translocations.

Session Title: Cancer Poster Session III

PB5016 Disrupted DNA methylation signals at lung cancer diagnosis differ from smoking-related and pre-diagnostic signals: results based on one hospital-based and four population-based studies in Europe.

Authors:

T. H. Nøst[1,2], I. Urbarova[2], Y-Q. Sun[1,3], X-M. Mai[1], O. Røe[1,4], P. Sætrom[1,3], M. Johansson[5], M. Johansson[6], B-H. Grønberg[1,3], K. Hveem[1,3], M. Chadeau-Hyam[7], T. Sandanger[2]; [1]Norwegian Univ. of Sci. and Technology, Trondheim, Norway, [2]UiT The Arctic Univ. of Norway, Tromsø, Norway, [3]St Olav's Univ. Hosp., Trondheim, Norway, [4]Levanger Hosp., Nord-Trøndelag Hlth.Trust, Levanger, Norway, [5]Umeå Univ., Umeå, Sweden, [6]Intl. Agency for Res. on Cancer, Lyon, France, [7]Imperial Coll. London, London, United Kingdom

Background: Lung cancer (LC) remains the leading cause of cancer deaths worldwide and there is an urgent need to improve LC risk assessment as current screening criteria miss a large proportion of cases and result in a high rate of false positives on low-dose CT screening. Risk markers for LC based on minimally invasive blood-based tests could help refine screening strategies and for clinical nodule management. **Methods:** We evaluated DNA methylation in blood drawn at LC diagnosis or pre-diagnostically in five different cohorts using the Illumina Infinium HumanMethylation450 and EPIC arrays. Diagnostic samples from 127 patients with confirmed LC, and 55 individuals with suspected but confirmed negative LC were obtained in a hospital-based study in the Norwegian Lung Cancer biobank (NLCB). We identified candidate CpG sites differentially methylated in LC cases at diagnosis, and further evaluated candidate CpGs in pre-diagnostic, population-based samples of 820 future LC cases and 829 matched controls (samples drawn up to 21 years before diagnosis). The pre-diagnostic samples were obtained from nested case-control studies within the Norwegian Women and Cancer Study (NOWAC), the Northern Sweden Health and Disease Study (NSHDS), the Trøndelag Health Study (HUNT), and the Italian part of the European Prospective Investigation into Cancer (EPIC) cohort. Logistic regression models formed the basis of the tests and age, sex, chip plate and chip position were covariates in addition to cohort as random effect. Models with adjustment for smoking status was also estimated. **Results:** We identified 15,385 and 49,407 FDR-significant CpG sites for which either hyper- or hypomethylation was observed in all and late-stage LC cases at diagnosis, respectively. We considered 14,981 candidate CpGs identified at the time of diagnosis. Less than 5% were among 18,760 CpGs associated with smoking in past large meta-analyses and the top ranked CpGs were not identified in previous prospective LC studies. Among candidate CpGs, 14 were also FDR-significant in the pre-diagnostic samples. A methylation score is being established for a subset of the candidate CpGs and its discriminative performance for risk prediction of LC is currently being explored both at and prior to diagnosis. **Conclusions:** Specific methylation biomarkers have a strong potential to improve lung cancer risk assessment. Here we present a unique design to identify novel markers with a focus on samples from time of diagnosis and the CpGs identified differ from those identified in previous studies on LC and smoking. At the conference, we will present risk estimates based on a targeted methylation risk prediction model.

Session Title: Cancer Poster Session I

PB5017 Distribution and frequency of BRCA1/BRCA2 variants in Turkish population with BRCA related cancers: A single center experience.

Authors:

A. Ceylan[1,2], G. Güleç Ceylan[1,2], M. Altan[2], B. Cavdarli[2], V. Topcu[2], E. E. Kurt[1,2], M. Zorluer[2], S. F. Yildirim[2], C. Semerci Gündüz[1]; [1]Ankara Yildirim Beyazit Univ., Sch. of Med., Dept. of Med. Genetics, Ankara, Turkey, [2]Ankara City Hosp., Ankara, Turkey

Germline mutations in *BRCA1* and *BRCA2*, play a significant role in cancer pathogenesis and predisposition, particularly in breast, ovarian, and prostate cancers. Identification of recurrent and novel mutations is very substantial to understanding the differences by population and the structure of the genes. The frequency of variation in the Turkish population is not clearly known. Using the Next Generation Sequencing (NGS) and Multiplex Ligation Probe Amplification (MLPA), we analyzed germline variants of *BRCA1* and *BRCA2* genes for the first 2000 individuals who had high pre-test probability or cancer diagnosis from the patients referred to Ankara City Hospital between 2019-2023. We identified 294 variants classified as VUS or LP/P at *BRCA1* and *BRCA2* genes using the NGS method. While 150 out of 294 variants were LP/P; 144 out of 294 variants were VUS. 129 and 165 variations were identified in *BRCA1* and *BRCA2*, respectively. Eleven of them were novel. In addition to single nucleotide variations, copy number changes in *BRCA1* and *BRCA2* genes were also investigated by MLPA method and CNV was detected in 15 patients. Thus, the rate of disease-causing variants *BRCA1* and *BRCA2* genes was calculated as 8.4%. This rate was found to be similar to those reported in the literature. In this study, we aimed to demonstrate the clinical and demographic findings of patients who had *BRCA*-related cancer. It is crucial to identify the known variants in the specific genes and their functional effect for improvement in the targeted therapy protocols.

Session Title: Cancer Poster Session II

PB5018 DNA Methylation Influences the Relationship Between Polysubstance Use and Physical Frailty Among People with Chronic HIV

Authors:

X. Liang[1], A. Justice[2,3], B. Aouizerat[4,5], K. Xu[2,6]; [1]Dept. of Epidemiology and Biostatistics, Coll. of Human Med., Michigan State Univ., East Lansing, MI, [2]VA Connecticut Hlth.care System, West Haven, CT, [3]Yale Univ. Sch. of Med., New Haven VA Connecticut Hlth.care System, New Haven, CT, [4]Bluestone Ctr. for Clinical Res., Coll. of Dentistry, New York Univ., New York, NY, [5]Dept. of Oral and Maxillofacial Surgery, Coll. of Dentistry, New York Univ., New York, NY, [6]Dept. of Psychiatry, Yale Sch. of Med., New Haven, CT

Polysubstance use (PSU) refers to the consumption of more than one substance, either simultaneously or sequentially. Individuals with PSU have an increased risk of physical frailty, especially among people with chronic HIV (PWH), who have systematically elevated inflammatory function. We and others have demonstrated that both substance use and HIV-1 infection are associated with DNA methylation (DNAm), an epigenetic mechanism involved in regulating immunogenomic function. However, the role of DNAm in the relationship between PSU and physical frailty among PWH remains unclear. In this study, we hypothesized that a subset of DNAm CpG sites (CpGs) mediate the relationship between PSU and physical frailty among PWH. We employed the widely recognized Veteran Aging Cohort Study (VACS) 2.0 Index, known as the VACS Index, to evaluate physical frailty. This measurement tool effectively predicts mortality risk and health outcomes in people living with HIV (PWH). By considering various factors such as age, viral load, CD4 count, and coinfections, the VACS Index offers a comprehensive assessment of overall health. Using data from the VACS that included 1,944 DNA samples from PWH, we first performed Epigenome-Wide Association Study (EWAS) on substances of alcohol, tobacco, cannabis, cocaine, stimulants, and opioids separately and jointly. We identified a total of 2,112 CpGs that were associated with at least one substance (False Discovery Rate, FDR<0.05). Among them, 1,708 CpGs were unique for single substances. We then built a model using 2,112 CpGs to predict PSU by applying multi-response Gaussian regression with elastic net regularization method. The model comprising 1,474 CpGs showed the best prediction of PSU. Subsequently, we constructed a methylation risk score (MRS) based on the 1,474 CpGs, which showed excellent prediction for PSU with an Area Under the Curve (AUC) of 83.92%. Finally, we conducted a mediation analysis to quantify the effects of PSU on VACS Index explained by MRS. MRS showed a significant mediation effect of PSU and VACS Index (Indirect Effect IE_{psu}=3.75, p=0.018). Together, our findings provide evidence that DNAm plays a partial mediating role in the linkage of PSU and increased frailty among PWH.

Session Title: Cancer Poster Session III

PB5019 DNA methylation-predicted prenatal smoking exposure and adult lung cancer risk: a nested case-control study

Authors:

M. Ru[1], D. S. Michaud[2], N. Zhao[2], K. T. Kelsey[3], D. C. Koestler[4], L. Jiayun[1], E. A. Platz[1], C. Ladd-Acosta[1]; [1]Johns Hopkins Bloomberg Sch. of Publ. Hlth., Baltimore, MD, [2]Tufts Univ., Boston, MA, [3]Brown Univ., Providence, RI, [4]Univ. of Kansas, Kansas City, KS

The prenatal period is a critical window for adult disease development, including cancer. Studies show that DNA methylation signatures of prenatal smoking exposure persist into adulthood. Because active smoking alone does not explain all lung cancer, and not all smokers develop lung cancer, it is possible that prenatal smoking exposure may contribute to adult lung cancer risk independent of adult smoking, and possibly act synergistically with adult smoking. A prior study reported null or inverse associations between scores for these methylation signatures and adult lung cancer risk adjusting for methylation-predicted adult smoking. That study did not consider notable trends in maternal smoking prevalence by birth cohort before the 1960s. To address this gap, we examined the association between methylation scores for prenatal smoking exposure and adult lung cancer risk, independent of methylation-predicted adult packyears, and by birth cohort. We conducted a nested case-control study in participants in both CLUE I and II. We included 196 incident lung cancer cases and 196 controls who were incidence density sampled and matched 1:1 on age, sex, race, and smoking status. Cases were ascertained by cancer registry linkage through 2018 and histologically confirmed. Leukocyte DNA methylation was measured using the Illumina MethylationEPIC BeadChip in prediagnostic blood collected in CLUE II in 1989. We calculated 2 methylation prenatal smoking scores detected in (i) older children (19-CpG), (ii) younger adults (15-CpG); and a methylation-predicted adult packyears score. Conditional logistic regression was used to estimate odds ratios (OR) and 95% confidence intervals (CI) adjusting for methylation-predicted adult packyears and principal components for batch effects. We also stratified by birth cohort in quartiles. The 15-CpG score was associated with increased lung cancer risk (OR=1.30 per SD, 95% CI 1.02-1.66, p-trend=0.04), which was driven by the 1930-40 birth cohort (49 pairs, OR=2.76, 95% CI 1.34-5.70, p-trend=0.006). Likewise, the 19-CpG score was associated with increased risk only in the 1930-40 birth cohort (OR=1.89, 95% CI 1.03-3.45, p-trend=0.039). Compared to those <=median on both prenatal scores, participants >median on both scores had higher lung cancer risk (OR=1.92, 95% CI 1.12-3.30), including the 1930-40 birth cohort (OR=4.32, 95% CI 1.19-15.7). This prospective study suggests a positive association between prenatal smoking exposure and adult lung cancer risk, especially in the 1930-40 birth cohort. Studies with multiple birth cohorts are needed to further investigate this association. Funding: AACR-Johnson&Johnson, NCI, NPCR

Session Title: Cancer Poster Session I

PB5020 † Do candidate genes increase clinical utility of hereditary cancer panels? When less is more

Authors:

J. Herrera-Mullar[1], M. Towne[2], C. Horton[1], A. Weaver[1], J. Huang[1], D. Thrush[1], K. Radtke[1], B. Wayburn[1]; [1]Ambry Genetics, Aliso Viejo, CA, [2]Ambry Genetcs, Aliso Viejo, CA

Next generation sequencing has spurred new gene discovery for hereditary cancer predisposition (HCP) and allowed the development of multigene panel tests (MGPT). While the inclusion of more genes on HCP-MGPT increases diagnostic yield compared to single gene testing, data is lacking on how including new candidate genes on HCP-MGPT impacts diagnostic yield and variant of uncertain significance (VUS) rates. We assessed the frequency of variant classifications by gene-disease validity (GDV) scoring categories of genes on HCP-MGPT.
Reported variants in genes on HCP-MGPTs offered at a commercial laboratory from 2014-2021 were retrospectively reviewed. The frequency of pathogenic (P), likely pathogenic (LP) variants, and VUS were recorded and categorized according to four different GDV categories: uncharacterized (genes with limited/disputed GDV), moderate, strong, and definitive based on evaluation of published clinical and experimental evidence in accordance with ClinGen GDV guidelines. Frequencies were determined by reviewing cases with a reported relevant variant (>240,000 cases). Genes uncharacterized for HCP but characterized for an autosomal recessive disorder were excluded from this analysis (*FANCC, RAD50, MRE11, XRCC2, BLM, NBN*). Genes with moderate or limited GDV for HCP at the start of the time frame (n=20) were reviewed for changes to GDV category. Variants were classified in accordance with ACMG guidelines.
Of >240,000 cases reported, the frequency of P/LP variants was 31.5%, 19.9%, 11.0%, and 0% among genes in the definitive, strong, moderate, and limited/disputed categories, respectively. VUS rates were inversely proportional to positive rate and increased as GDV scores decreased (68.5%, 80.1%, 89.0%, 100% for the definitive, strong, moderate, and uncharacterized categories, respectively). Among 20 genes with moderate/limited GDV at time of addition to MGPT, only 1 (5%) became characterized to a definitive GDV. Instead, the majority (55%) were downgraded to limited or disputed, and the remainder were unchanged limited (25%) or moderate (15%).
Rates of clinically actionable variants (P/LP) are correlated with higher GDV scores, and the rate of VUS is inversely related to GDV scores. Most genes with limited or moderate GDV at the time of panel addition were downgraded to limited/disputed categories or remained uncharacterized. The inclusion of newer candidate genes on HCP-MGPT did not affect the diagnostic yield, but substantially increased the VUS rate on HCP-MGPT. These results emphasize the importance of standardized GDV scoring to inform variant classification and gene inclusion criteria on MGPT to optimize clinical utility.

Session Title: Cancer Poster Session II

PB5021 Drug-target Mendelian randomization revealed significant association of genetically proxied metformin use with increased prostate cancer risk

Authors:

X. Shu[1], X. Sun[1], J. Ping[2], J. Long[2], X. Guo[2], Q. Cai[2], X-O. Shu[2]; [1]Mem. Sloan Kettering Cancer Ctr., New York, NY, [2]Vanderbilt Univ. Med. Ctr., Nashville, TN

Background: Previous observational studies have assessed the association between metformin use and incidence of prostate cancer in general populations or diabetic patients, but findings remained mixed.**Methods:** This study used drug-target Mendelian randomization (MR) approach to evaluate potential causal relationship between metformin use and the risk of developing prostate cancer. *Cis*-expression quantitative trait loci (*cis*-eQTL) variants in the gene targets of metformin were identified in the Genotype-Tissue Expression (GTEx) project and eQTLGen. A MR analysis was conducted to filter eQTL variants that also affected blood HbA1c levels using UK Biobank data. Summary statistics for prostate cancer were obtained from the Prostate Cancer Association Group to Investigate Cancer Associated Alterations in the Genome (PRACTICAL) consortium (79,148 cases and 61,106 controls). We also obtained male-specific summary-level data for type 2 diabetes, body mass index (BMI), total testosterone, bioavailable testosterone, estradiol, and sex hormone binding globulin (SHBG) from corresponding international consortia or UK Biobank. Inverse-variance weighted (IVW) regression, weighted median, MR-Egger regression, MR-PRESSO, multivariable MR, and colocalization test, were performed.**Results:** Genetically proxied metformin use (1 SD of target-associated HbA1c reduction, equivalent to 6.75 mmol/mol, 1.09%) was associated with higher risk of prostate cancer (odds ratio [OR]: 1.55, 95% CI:1.23-1.96, $P = 3.00\times10^{-3}$). Two metformin targets, mitochondrial complex I (MCI) (OR: 1.48, 95% CI:1.07-2.03, $P = 0.016$) and gamma-secretase complex (GSC) (OR: 2.58, 95% CI:1.47-4.55, $P = 0.001$), showed robust associations with prostate cancer risk. We further observed that the MCI and GSC-specific effects were partially mediated through BMI and total testosterone levels, with a mediated proportion of 16.4% and 34.3%, respectively. MR and colocalization suggested that expressions of *NDUFA13* (MCI-related gene) in testis and BMI, *APH1A* (GSC-related gene) in hippocampus and total testosterone levels may be influenced by shared genetic factors in the corresponding loci, respectively.**Conclusion:** Genetically proxied metformin use was associated with increased prostate cancer risk. Repurposing metformin for chemoprevention of prostate cancer is not supported by our findings.

Session Title: Cancer Poster Session III

PB5022 Educational attainment and variation of blood DNA methylation among survivors of childhood cancer: A bidirectional causal inference using polygenic scores

Authors:

Q. Dong[1], C. Chen[2], N. Song[3], H. Mulder[4], J. Easton[1], E. Walker[4], G. Neale[4], G. Kang[5], G. T. Armstrong[4], K. K. Ness[4], M. M. Hudson[4], I-C. Huang[4], Z. Wang[1]; [1]St. Jude Children s Res. Hosp., Memphis, TN, [2]Shanghai Jiaotong Univ., Shanghai, China, [3]Chungbuk Natl. Univ., Cheongju, Korea, Republic of, [4]St. Jude Children's Res. Hosp., Memphis, TN, [5]St. Jude Children's Res. Hosp., Memphis, TN

BACKGROUND: We previously identified a DNA methylation (DNAm) signature encompassing a total of 88 DNAm sites for educational attainment (Edu) among survivors of childhood cancer (CCS). We now evaluate bidirectional causal inferences with polygenic scores (PGS). **METHODS:** Study participants comprised 1,618 CCS (median age at DNAm=33.9 years) of European ancestry from the St. Jude Lifetime Cohort Study with pre-existing whole-genome sequencing (WGS) and DNAm profiling (Illumina EPIC BeadChip) data. PGS was calculated as the sum of the trait (Edu or DNAm) measurement increasing alleles. Bidirectional causal inference was analyzed using either Edu-associated genetic variants to derive the PGS (forward: Edu as an exposure and DNAm as an outcome) or methylation quantitative trait loci (meQTL) to derive the PGS (backward: DNAm as an exposure and Edu as an outcome). **RESULTS:** For forward analysis, we started with 1,271 independent Edu-associated SNPs established in general population studies and excluded SNPs not associated with Edu in CCS (i.e., P>0.05), resulting in 105 SNPs for an Edu-PGS. A substantial reduction in SNPs was expected, considering previous evidence of heterogeneous effects across different populations. A total of 88 DNAm sites were tested, and we set statistical significance to be 5.7×10^{-4} (i.e., 0.05/88). We found that Edu-PGS was only significantly associated with DNAm of cg03028088 ($P=5.30\times10^{-6}$). Further analysis using Mendelian Randomization (MR) with one-sample (two-stage least-squares) showed similar results ($P=2.1\times10^{-6}$). Sensitivity analyses excluding SNPs from PGS showed consistent results indicating that horizontal pleiotropy is not a major concern. For backward analysis, we searched SNPs within 1Mb-window centered around cg03028088, evaluated the associations between SNP dosages and the DNAm of cg03028088, and identified 214 significant meQTLs ($P_{FDR}<0.05$). After excluding correlated SNPs ($r^2<0.1$), 14 remained and were used to construct a cg03028088-PGS. We found a nominally significant association between cg03028088-PGS and Edu ($P=7.2\times10^{-3}$), which failed multiple testing, and further one-sample MR analysis confirmed the non-significance (P=0.27). cg03028088 was mapped to *ARID5B*, a gene implicated in inflammation, immune response, and carcinogenesis. Further MR analysis with two-sample methods (inverse variance weighted and MR-Egger) is underway. **CONCLUSION:** We found that DNAm of cg03028088 (*ARID5B*) was associated with Edu as an exposure. Future analyses will evaluate whether cg03028088 is predictive of adverse health outcomes among CCS and assess whether other Edu-associated DNAm sites may influence Edu.

Session Title: Cancer Poster Session I

PB5023 Elucidating the clonal evolutionary dynamics of hypermutated tumours using single-cell whole-genome sequencing

Authors:

M. Kalyva[1,2], F. Gibson[3], F. Muyas[1], C. Sauer[1], J. Espejo Valle-Inclan[1], S. Bhosle[3], S. Barthorpe[3], C. Herranz-Ors[3], H. Francies[3], G. Picco[3], M. Garnett[3], I. Cortes-Ciriano[1]; [1]EMBL-EBI, Hinxton, Cambridge, United Kingdom, [2]Univ. of Cambridge, Cambridge, United Kingdom, [3]Wellcome Trust Sanger Inst., Hinxton, Cambridge, United Kingdom

Immune checkpoint blockade (ICB) has revolutionized the cancer treatment landscape. Yet, durable responses are only observed in a subset of patients. The highest response rates to ICB have been observed in mismatch repair-deficient (MMRD) tumors, which are caused by the inactivation of the MMR pathway. Tumors with MMRD are characterized by a high burden of neoantigens. However, only ~50% of MMRD tumors respond to ICB. Previous clinical and preclinical studies of MMRD tumors indicate that the clonality rather than the burden of mutations determines the variable responses to ICB observed in the clinic. Given that the clonality of mutations is determined by the evolutionary dynamics and patterns of intra-tumour heterogeneity, it is paramount to elucidate the evolutionary trajectories of MMRD tumors to understand the variable immunogenicity of MMRD tumors. Prior genomics analysis of MMRD tumor evolution used bulk whole-genome sequencing (WGS), which is limited to infer the clonal structure and dynamics of tumors due to the impossibility to assign mutations to subclones and limited sensitivity for mutation detection.

Here, to study the evolution of MMRD tumors, we performed single-cell WGS on ~200 single cells from eight patient-derived colorectal MMRD organoids. Mutational burden, though variable across models, was comparable across single cells from the same tumor. Phylogenetic analysis uncovered multiple clones per tumor, with most pathogenic mutations mapping to the trunk of the phylogenetic tree, demonstrating that these mutations are shared across all cancer cells. Notably, few copy number aberrations were observed, indicating that tumor heterogeneity is driven by point mutations and indels. Furthermore, we developed a phylodynamic analysis software, called ClonalSim, to analyze the growth dynamics of tumors using the phylogenetic trees constructed for each tumor, and to infer the fitness of cancer cells and the timing of clonal expansions.

In sum, we present a novel experimental and computational framework that has allowed us to study the patterns of genetic heterogeneity and clonal dynamics underpinning the evolution of MMRD colorectal cancer at single-cell resolution. This framework is broadly applicable to study genetic heterogeneity and clonal evolution in other malignancies and cancer types.

Session Title: Cancer Poster Session II

PB5024 Elucidation Of Tumor Clonal Diversity in An AML Drug Resistance Model Using A High Throughput Single Cell Genome Amplification Method.

Authors:

S. Velivela, J. Dahl, J. West, J. Zawistowski; BioSkryb Genomics, Durham, NC

Introduction: Genomic plasticity within tumors contributes to the cellular heterogeneity, which can drive treatment resistance. However, detection of rare, treatment-resistant progenitors is extremely difficult using either conventional bulk population analysis or when analyzing a few individual cells. The goal of our study was to develop a high-throughput automated workflow that can detect intrinsic and acquired mechanisms of resistance to quizartinib, an FLT-3 inhibitor, in an acute myeloid leukemia (AML) cell line, MOLM-13. Methods: A "continual" drug dosage resistance model was created by 2-month duration of 2nM quizartinib, replenished every 3 days with fresh media, while a "dose-escalation" model was created by increasing the quizartinib dose weekly by 100 pM increments up to 2 nM. The single-cell whole genome amplification was performed using ResolveDNA (BioSkryb Genomics) with digital cell dispensing (HP D100), digital liquid dispensing (HP D300), and automated library preparation (Agilent Bravo). 184 resistant and 184 parental cells were sequenced by lowpass sequencing (Illumina NextSeq2000). 1,104 were processed at 272plex for whole exome sequencing. Bioinformatics analysis was performed using BaseJumper (BioSkryb Genomics). Results: The degree of genomic plasticity observed in the parental and two drug dosage models is significant, finding both loss (5p) and gain (19q) of chromosomal regions as a direct result of the drug regime. Copy number variation (CNV) within these groups shows dynamic genomic rearrangements within a single drug treatment model. Specifically, in the continual dosage model CNV analysis clusters into two predominant models with genomic rearrangements in the pathway for drug resistance of FLT-3 and or downstream MAPK activation circumventing the primary drug target. Principal component analysis (PCA) of single cells reveals 6,444 genomic variants that stratify the treatment groups, including a *FLT3* N841K secondary mutation predominant in the resistant population but identified at low frequency in the parental population. Conclusions: Detection of a *FLT3* secondary mutation in the treatment-naïve cell population demonstrates the diversity within treatment naïve cells and highlights the role of natural selection during drug treatment driving resistance. Similarly, the distinct mechanisms of acquired therapy resistance between models shed light on the marked genomic plasticity resulting from varying modes of selection pressure.

Session Title: Cancer Poster Session III

PB5025 Enhanced Breast Cancer Subtyping Integrating Complete Gene Expression Profiles and Genetic Ancestry

Authors:

J. Stepanian-Rozo[1], J. Zabaleta[2], J. Duitama[1]; [1]Univ. de los Andes, Bogota, Colombia, [2]Stanley S. Scott Cancer Ctr., LSU Hlth.Sci. and Louisiana Cancer Res. Consortium Ctr., New Orleans, LA

Breast cancer (BC) is the first leading cause of death in women around the world. Accurate subtyping of BC based on gene expression is crucial for optimizing treatment strategies, and ancestry has shown to be a modulator in gene expression. This study aims to improve BC subtyping by integrating comprehensive gene expression profiles and ancestry information.We acquired and analyzed a dataset of 406 RNA-Seq samples sequenced from patients with diverse ancestries, geographical origins, and ethnicities, combining 318 publicly available samples with 88 new samples.RNA-Seq from breast tumor data were used to genotype 10,397 SNPs and predict ancestries using the software ADMIXTURE, and combining the RNA-seq genotype calls with calls from the 1000 genomes project. We ran the genefu R package for predicting PAM50 subtypes, and recreated the results using the supervised machine learning algorithms random forest (RF) and support vector machine (SVM), achieving accuracy rates of 0.95 and 0.92, respectively.We integrated ancestry prediction to supervised machine learning experiments resulting in less favorable metrics, achieving accuracy values of 0.86 and 0.85 for RF and SVM, respectively. Additionally, we observed discrepancies between the assigned PAM50 subtypes, the gene expression patterns, and ancestry prediction.To overcome possible biases in the genefu predictions, we investigated the clusters obtained following an unsupervised machine learning strategy, based on the K-means algorithm. The results suggest novel gene expression-based subgroups within breast cancer tumors, where a part of the variance is explained by ancestry.In conclusion, this study highlights the importance of integrating gene expression profiles and ancestry information in BC clustering, contributing to understanding BC heterogeneity. The findings presented lay the groundwork for improved treatment decision-making in BC management.

Session Title: Cancer Poster Session I

PB5026 Enhancer Analysis Identifies *PAK6* and *CHST14* as Regulatory Targets of Endometrial Cancer Risk Variation at the 15q15.1 Locus

Authors:

D. Glubb[1], B. Naumann[1], C. Chapman[1], J. Gertz[2], T. O'Mara[1]; [1]QIMR Berghofer Med. Res. Inst., Brisbane, Australia, [2]Univ. of Utah, Salt Lake City, UT

The largest published endometrial cancer GWAS identified a credible set of 47 risk variants at the 15q15.1 risk locus, all of which are located in non-coding regions. Based on previous observations of enrichment in endometrial cancer enhancer/promoter chromatin loops, we hypothesised that the causal risk variant(s) mediates its effect through regulation of gene expression. To identify credible risk variants that regulate gene expression through enhancer activity at the 15q15.1 locus, we focused on eight credible risk variants located in five putative enhancers marked by H3K27Ac in Ishikawa endometrial cancer cells. We used matching H3K27Ac chromatin looping data to identify 17 genes whose promoters physically interact with the putative enhancers and prioritised 13 genes that had robust expression. Using CRISPR interference to inhibit the putative enhancers, we confirmed enhancer activity for four genes (*CHST14, GPR176*, *KNSTRN*, and *PAK6*). We further investigated *CHST14* and *PAK6*, as both genes have function related to sex hormones, such as estrogen and testosterone, which are endometrial cancer risk factors. Reporter gene assays showed that the presence of risk alleles in one of the enhancers targeting *CHST14* and *PAK6* increased their promoter activity by ~2-fold compared to the alternate alleles. Knockdown experiments in Ishikawa cells suggested that *CHST14* plays a critical role in colony formation, while PAK6 knockdown had no consistent cellular effects. Lastly, we used RNA-seq to explore the effects of *CHST14* and *PAK6* knockdown on gene expression and identify other phenotypes or pathways impacted by these genes. In summary, we identified credible risk variants at the 15q15.1 locus that regulate *CHST14* and *PAK6* expression through enhancer activity, providing insight into their potential contribution to endometrial cancer development.

Session Title: Cancer Poster Session II

PB5027 Enhancing risk stratification of Acute Myeloid Leukemia with next generation cytogenomics

Authors:

S. Eacker[1], A. Muratov[1], M. Malig[1], B. Nelson[1], M. Wood[1], M. VanDyke[1], K. Langford[1], L. Beppu[2], O. Sala-Torra[2], M. Fang[2], C. Yeung[2], S. Sullivan[1], I. Liachko[1], J. Radich[2]; [1]Phase Genomics, Inc., Seattle, WA, [2]Fred Hutchinson Cancer Ctr., Seattle, WA

Cytogenetic diagnostics including karyotyping, FISH, and chromosome microarray are regularly applied in cases of suspected leukemia, including Acute Myeloid Leukemia (AML). Each of these three tests can identify non-overlapping sets of chromosome aberrations known to have prognostic or diagnostic value. Because each test addresses different classes of prognostic variants, all three tests are often performed in series or in parallel, adding time and expense to the diagnostic process. Here we describe OncoTerra, a platform based on proximity ligation sequencing (PLS), that can be used to identify variants commonly assayed by karyotyping, FISH, and array in a single assay. OncoTerra uses a specialized library preparation chemistry that requires no specialized equipment. Following Illumina sequencing, data is analyzed on a secure cloud platform. In a pilot study of 48 AML patient samples, OncoTerra identified variants previously observed by cytogenetics and additional variants that previously identified. These variants were composed of translocations, inversions, and copy number variants either too small or too complicated to be accurately identified by cytogenetics. We identified a novel recurrent inv(9)(p13) that likely escapes resolution by karyotyping and involves two paralogous genes making it difficult to detect by WGS. OncoTerra's resolution differentiates canonical and non-canonical inv(16), a variant for essential for patient risk stratification. Importantly, additional variants identified by OncoTerra changed risk categories for a significant fraction of patients based on ELN 2022 standards. These findings support deployment of OncoTerra as a time- and cost-effective, high-resolution cytogenomic solution to improve patient care.

Session Title: Cancer Poster Session III

PB5028 Equitable machine learning counteracts ancestral bias in precision medicine, improving outcomes for all

Authors:

K. Graim, **L. A. Smith**, J. A. Cahill; Univ. of Florida, Gainesville, FL

Gold standard genomic datasets severely underrepresent non-European populations, leading to inequities and a limited understanding of human disease. Therapeutics and outcomes remain hidden because we lack insights that we could gain from analyzing ancestry-unbiased genomic data. To address this significant gap, we present PhyloFrame, the first-ever machine learning method for equitable genomic precision medicine. PhyloFrame creates equitable transcriptomic signatures of disease and corrects for ancestral bias by leveraging two big-data resources (global population variation data and functional interaction networks) with disease-relevant genomic data. Application of PhyloFrame to cancer shows marked improvements in predictive power across all ancestries, less model overfitting, and a higher likelihood of identifying known cancer-related genes. Our ability to provide accurate predictions for underrepresented groups is substantially increased. These results demonstrate how AI can mitigate ancestral bias in training data and contribute to equitable representation in medical research.

Session Title: Cancer Poster Session I

PB5029 Establishing a single-cell eQTL dataset of lung tissues from Asian women never-smokers to elucidate cell-type specific genetic regulation

Authors:

T. Luong[1], E. Long[1], J. Yin[1], J. Shin[2], A. Kane[1], A. Golden[1], Y. Chang[2], J. Shi[1], J. Lee[3], E. Kim[2], J. Choi[1]; [1]Div. of Cancer Epidemiology and Genetics, Natl. Cancer Inst., NIH, Bethesda, MD, [2]Dept. of Internal Med., Yonsei Univ. Coll. of Med., Seoul, Korea, Republic of, [3]Dept. of Thoracic and Cardiovascular Surgery, Yonsei Univ. Coll. of Med., Seoul, Korea, Republic of

Expression quantitative trait loci (eQTL) studies have been powerful in linking GWAS variants to potential target genes, providing genetic mechanisms underlying common diseases such as cancers. However, the current eQTL resources lack ancestral diversity and are primarily based on bulk tissues. Emerging single-cell eQTL (sc-eQTL) approaches can detect context-specific gene regulation but are mainly of blood samples or cultured cells and still representing European populations. This limits our abilities to test GWAS variants in cancer-relevant cell types as well as in diverse populations. One of the most common cancers and the leading cause of cancer-related death worldwide is lung cancer, and GWAS has identified multiple loci in diverse populations including East Asians. To generate a resource to characterize lung cancer GWAS loci we are building a lung sc-eQTL dataset of Asian population while addressing common challenges of tissue sc-eQTL. Namely, processing fresh tissue in a population scale is logistically challenging and costly, and epithelial cells (including cell types of lung cancer origin) are vulnerable to the dissociation and freezing/thawing process. To address these issues, we incorporated sample multiplexing and cell type balancing of cryopreserved tissues. We collected fresh tumor-adjacent normal lung tissues from 128 never-smoking Korean women and dissociated them before cryopreservation. We then performed single-cell RNA sequencing (scRNA-seq) using 10x Chromium Single Cell 3' v3.1 chemistry with multiplexing of 6 samples/batch. To enrich for epithelial cells, we utilized flow cytometry with surface markers of four major lung cell types (epithelial: EpCAM+/CD45-, immune: EpCAM-/CD45+, endothelial and stromal: EpCAM-/CD45-) before 10X library preparation. Concurrently, we performed DNA genotyping and imputation using matched blood samples. Following scRNA-seq (~36,000 reads/cell) we performed a genotype-based sample demultiplexing using Demuxlet. By integrating Demuxlet and Scrublet we identified ~88% of the detected cells as singlets. After applying QC measures to filter empty droplets and low-quality cells we obtained 435,739 cells or 3,404 cells/patient. After clustering, an initial label transfer using Azimuth and Human Lung Cell Atlas reference data identified ~50% of our cells in epithelial groups. We will further perform eQTL analyses for individual cell types and integrate the results with lung cancer GWAS data to identify susceptibility genes in population and cell-type specific manner. Our approach presents a disease-customized and cost-effective method of building sc-eQTL dataset for solid tissues.

Session Title: Cancer Poster Session II

PB5030 Establishing the effect of reproductive factors on breast cancer risk: a multivariable Mendelian randomization analysis.

Authors:

C. Prince[1,2], L. D. Howe[1,2], G. C. Sharp[3,1,2], A. Fraser[1,2], B. Lloyd-Lewis[4], R. C. Richmond[1,2]; [1]MRC Integrative Epidemiology Unit, Univ. of Bristol, Bristol, United Kingdom, [2]Population Hlth.Sci., Bristol Med. Sch., Univ. of Bristol, Bristol, United Kingdom, [3]Sch. of Psychology, Univ. of Exeter, Exeter, United Kingdom, [4]Sch. of Cellular and Molecular Med., Univ. of Bristol, Bristol, United Kingdom

Background: Childbirth is thought to be associated with breast cancer risk due to the changes that occur to the mammary tissue during pregnancy and lactation. Previous studies have identified an early age at first full-term birth is a protective factor against later life breast cancer, whereas nulliparity or a late age at birth and thought to increase risk. However, there may be an increase in risk of breast cancer immediately post pregnancy due to the breast remodelling, which creates a inflammatory and wound healing-like environment. While there has been some investigation of causality in this setting, it is unclear how reproductive traits affect the risk of breast cancer independently from other reproductive and menstrual factors, given that the traits are often highly correlated. **Methods**: We used genetic data from UK Biobank and the Breast Cancer Association Consortium for four reproductive factors; age at first live birth, age at last live birth, number of live births and ever having children, and 3 breast cancer outcomes; overall, oestrogen-receptor positive and oestrogen-receptor negative breast cancer. We applied univariable and multivariable mendelian randomization, to account for genetic correlation and potential confounding, to estimate the causal effects of the reproductive factors on breast cancer. **Findings**: We found limited evidence for an effect of earlier age at first birth on overall (Odds ratio (OR):1.00; 95% confidence interval: 0.85,1.18, per standard deviation (SD) increase) and oestrogen-receptor positive breast cancer risk (OR:1.04; 0.89,1.20, per SD increase), however some evidence that an earlier age increases risk of oestrogen-receptor negative breast cancer (OR:0.76; 0.61,0.95, per SD increase). Furthermore, we identified limited evidence that age at last live birth, number of live births or ever having children impacts breast cancer risk. **Conclusion**: Our findings contrast with previous observational research regarding age at first live birth and highlights the importance of considering a lifecourse approach and the inter-relationships between reproductive factors when assessing risk on breast cancer.

Session Title: Cancer Poster Session III

PB5031 ETCHING: Ultra-fast Prediction of Somatic Structural Variations and Fusion Genes

Authors:

J-W. Nam; Hanyang Univ., Seoul, Korea, Republic of

Variant callers typically produce massive numbers of false positives for structural variations, such as cancer-relevant copy-number alterations and fusion genes resulting from genome rearrangements. Here we describe an ultrafast and accurate detector of somatic structural variations that reduces read-mapping costs by filtering out reads matched to pan-genome k-mer sets. The detector, which we named ETCHING (for efficient detection of chromosomal rearrangements and fusion genes), reduces the number of false positives by leveraging machine-learning classifiers trained with six breakend-related features (clipped-read count, split-reads count, supporting paired-end read count, average mapping quality, depth difference and total length of clipped bases). When benchmarked against six callers on reference cell-free DNA, validated biomarkers of structural variants, matched tumour and normal whole genomes, and tumour-only targeted sequencing datasets, ETCHING was 11-fold faster than the second-fastest structural-variant caller at comparable performance and memory use. The speed and accuracy of ETCHING may aid large-scale genome projects and facilitate practical implementations in precision medicine.

Session Title: Cancer Poster Session I

PB5032 Evaluating analytical validity of diverse cancer mutation calling pipelines of whole exome sequencing data

Authors:

P-M. Chien[1,2], T-H. Yuan[2], Y-B. Wang[2], J-H. Huang[2], J. Hsu[1]; [1]Graduate Inst. of Med. Genomics and Proteomics,Natl. Taiwan Univ., Taipei, Taiwan, [2]AI Labs, Taipei, Taiwan

Identification of somatic mutations from tumor tissue samples holds substantial clinical consequences for making informed medical decisions. Therefore, evaluating the performance of various tool combinations within somatic variant calling pipelines has become an important issue when employing whole exome sequencing (WES) analysis in clinical settings. In the study, we utilized 9 set of WES reads data of tumor samples released by the Sequencing and Quality Control Phase 2 (SEQC2) project to benchmark the analytical validity of somatic mutation calling workflows. The SEQC2 project, conducted in collaboration with the U.S. Food and Drug Administration (FDA), aims to assess the performance and accuracy of various genomic sequencing technologies and analysis pipelines for clinical applications.We explored the benchmarking approach to compare various combinations of aligner and caller tools. Included aligning tools BWA, Bowtie2, Dragen-Aligner, DRAGMAP, and HISAT2. For the caller, we included Mutect2, TNscope, Dragen-Caller, and DeepVariant. Among these tools, Dragen-Aligner and Dragne-Caller combination showed the best performance with mean F1-score of 0.9659 in SNV detection ,while the combination of BWA and Mutect2 showed the highest mean F1-score of 0.9485 among the open-source software combinations. Notably, results showed that the variant callers exerted a significant influence compared with the impact of the aligners. Furthermore, in true positive set, all variant callers except DeepVariant share the same detect limitation of variant allele frequency (VAF) at 0.0063 whereas the benchmark lowest VAF is at 0.0058. Among several cancer-related indicators, we identified varying concordance results in some tools. For example, TNscope exhibited a tendency to underestimate tumor mutation burden and missed drug-resistance variants such as FLT3(c.G1879A:p.A627T) and MAP2K1(c.G199A:p.D67N). To conclude, our investigation provides a valuable and inclusive guide for clinical genomic researchers on tumor mutation identification, accomplished through an in-depth performance comparison among diverse tool combinations.

Session Title: Cancer Poster Session II

PB5033 Evaluating the Two-Hit Hypothesis: Analyzing differences between germline and somatic cancer-associated mutational patterns in tumor suppressor genes

Authors:

S. Lulla[1], D. I. Ritter[1], C. Kesserwan[2], S. E. Plon[1]; [1]Baylor Coll. of Med., Houston, TX, [2]NYU Sch. of Med., New York, NY

Background: The 'two-hit' hypothesis of cancer predisposition requires two inactivating events in *trans* for complete loss of tumor suppressor gene (TSG) function and consequent tumorigenesis. Both events can be somatic but inheriting one germline loss of function (LoF) variant increases cancer risk. However, *DICER1*-associated tumors display one LoF variant (somatic or germline) in *trans* with one missense variant (exclusively somatic). We are comprehensively analyzing TSG germline and somatic mutational data to assess concordance of germline and somatic mutational patterns consistent with the two-hit hypothesis.
Methods: We identified 44 TSGs associated with autosomal dominant cancer predisposition. For these, we extracted Pathogenic/Likely Pathogenic (P/LP) germline variants from ClinVar and developed a weighted score by translating nominal classifications to numeric values to include conflicting variants with a high proportion of P/LP submissions. Somatic variants curated as Oncogenic/Likely oncogenic (O/LO) were extracted from cBioPortal (213 non-redundant studies). We harmonized variant nomenclature between germline and somatic data using the ClinGen Allele Registry. For each TSG, we also evaluated the proportion and germline classification of somatic O/LO variants that overlap with any ClinVar variant.
Results: From 44 TSGs, we identified $3.2x10^4$ P/LP variants out of $1.7x10^5$ germline variants in ClinVar (median=323; IQR=122-733 P/LP variants per TSG). By applying our weighted score, we retained an additional 4 (IQR=1-13; range=0-43) conflicting variants per TSG with primarily P/LP assertions. O/LO somatic variants from cBioPortal totaled $1.7x10^4$ (median=198; IQR=94-435 per TSG), of which $5.1x10^3$ variants (median=69; IQR=20-130 variants per TSG) were in ClinVar as germline variants. Of these, 82.6% (IQR=71.3-91.3%) intersecting variants per TSG were included in our P/LP set with other germline classifications ranging from Benign to variant of uncertain significance.
Conclusion: Although germline and somatic variants are interpreted using different criteria, we find that both classifications are highly concordant for the shared variants. However, approximately 90% of the 45,277 unique cancer variants in these large databases are exclusive to either germline or somatic data. We are further assessing consistency with the two-hit hypothesis by comparing the distribution of annotated variants by mutation type, mutational spectrum, and location within TSGs to inform genetic mechanisms of tumorigenesis and contribute to improved classification of hereditary and somatic cancer variants.

Session Title: Cancer Poster Session III

PB5034 Evaluation of germline variants in retinoblastoma using non-negative matrix factorization

Authors:

L. Chapman Hannah[1], J. Kim[2], D. Stewart[2], Z. Boukouvalas[3]; [1]NIH, Rockville, MD, [2]Natl. Cancer Inst., Rockville, MD, [3]American Univ., Washington, DC

Approximately 8-10% of children with cancer harbor rare, pathogenic or likely pathogenic (P/LP) germline variants in known tumor-predisposition genes. Common variants have smaller effect sizes and collectively may affect risk. Most children with brain tumors, however, do not harbor known risk variants. To identify additional risk alleles, we used non-negative matrix factorization (NMF), a feature-reduction algorithm commonly used in somatic mutation analysis, to evaluate rare germline variants in a set of retinoblastoma samples (Rb) from the Pediatric Cancer Genome Project (PCGP) and the St. Jude LIFE (SJLIFE) Study. NMF is an unsupervised method that reduces high dimensional feature spaces to a latent feature space.

In this pilot study, variants from chromosomes 13 (harboring known tumor-predisposition genes RB1 and BRCA2) were evaluated. Whole genome sequencing (WGS) data from 145 Rb patients and 364 non-cancer controls from PCGP and SJLIFE were downloaded from St. Jude Cloud. Single nucleotide variants (SNVs) and indels from WGS were called using an ensemble-based pipeline (GEMSCAN) and then filtered based on a set of quality metrics. In the following study NMF was used to identify a latent set of features (variants) within chromosomes 13, and as a result we are able to evaluate underlying groups within the data. The NMF model was generated with the R package RcppML (v0.3.7) using 2 features to capture the most variance in the lowest feature space. We hypothesized that distinct clusters of cases and controls can be identified within chromosome 13.

The analysis included 7,631 rare coding germline variants from chromosome 13 for NMF analysis. In this study, NMF is used to extract sets of germline variants from a high dimensional feature space from cases and controls. The highest ranking features within the coefficient matrix (H) were evaluated. The variants within RB1 that had NMF coefficient weights greater than zero had a higher frequency in cases versus controls (P-value:0.02857). NMF clustering identified 4 RB1 variants all found in Rb cases, including 2 pathogenic variants: NC_000013.11:g.48379594C>T and NC_000013.11:g.48362859C>T. The results of these studies could help identify putative germline variants that collectively contribute to cancer risk.

Session Title: Cancer Poster Session I

PB5035 Evaluation of hereditary predispositions to hematologic malignancy in stem cell transplant: A case series

Authors:

N. Deuitch[1], J. Davis[2], K. Craft[3], L. Cunningham[4], P. Liu[1]; [1]Natl. Human Genome Res. Inst., Bethesda, MD, [2]Natl. Inst. of Allergy and Infectious Disease, Bethesda, MD, [3]Natl. Human Genome Inst., Bethesda, MD, [4]Natl. Cancer Inst., Bethesda, MD

Hereditary predispositions to hematologic malignancy (HHMs) are present in ~10-20% of hematologic malignancies. Allogeneic hematopoietic stem cell transplantation (HSCT) is the gold standard of treatment in hematologic malignancies (HMs). However, HHMs can complicate the transplant process. One such HHM, familial platelet disorder with associated myeloid malignancies (FPDMM) caused by germline RUNX1 pathogenic variants, is characterized by thrombocytopenia, platelet functional defects and predisposition to HM. Identifying these conditions is imperative for diagnosis, management, and treatment of affected individuals and family members. Here we present three families with RUNX1-FPDMM, who underwent HSCT each highlighting unique issues with HHMs. Case 1: A 59 yo Hispanic male was referred for HSCT for myelodysplastic syndrome (MDS). He had been transplanted 20 years prior for MDS, using his brother as the donor. In the interim his brother had also developed MDS and had died. Somatic NGS revealed homozygous loss of *RUNX1*, and fibroblast testing confirmed a heterozygous germline deletion. Cascade screening indicated the brother was an obligate carrier of the *RUNX1* deletion and raising suspicion for donor derived malignancy. Unfortunately, no acceptable donor could be identified, and this patient ultimately died. Case 2: A 42 yo Hispanic male was referred for HSCT for acute myeloid leukemia (AML), which had progressed from MDS. Somatic NGS testing revealed pathogenic variants in *RUNX1*, *RAD50* and *BCOR*. Fibroblast sequencing confirmed that the *RUNX1* and *RAD50* variants were germline, and cascade testing was initiated as a part of donor identification. While there was limited evidence for impact of the *RAD50* variant on transplant outcomes, a familial donor was identified who was negative for both variants. The patient is now 2.5 years post-transplant and doing well. Case 3: A 33 yo White male was identified to have a duplication of exons 3-7 of RUNX1 on NGS panel testing for a longstanding history of thrombocytopenia and family history of AML. Baseline bone marrow biopsy revealed increased myeloblasts and anemia, raising concern for evolving HM. Given the family history and diagnosis of RUNX1-FPD, the patient was referred for HSCT. A matched unrelated was used. The patient is now one-year post-transplant and doing well. The cases highlight the importance of identifying HHMs and issues including evaluation of VUSs, detection of CNVs, racial disparities in donor availability, sample type, and timing of transplant and genetic testing.

Session Title: Cancer Poster Session II

PB5036 † Evolutionary routes of oesophageal adenocarcinoma genomes following whole genome doubling.

Authors:

G. Contino[1,2,3], M. Yap[1], F. Swirsky[1]; [1]Inst. of Cancer and Genomic Sci., Univ. of Birmingham, Edgbaston, United Kingdom, [2]Univ. Hosp. of Birmingham Univ., Birmingham, United Kingdom, [3]Von Hugel Inst., Univ. of Cambrdige, Cambrdige, United Kingdom

Background: Oesophageal Adenocarcinoma (OAC) is the 5th most common cancer related cause of death in the UK and is characterized by a dismal 5-year survival of less than 20%. Whole genome sequencing studies including Pan-cancer analysis highlighted that OAC is driven by large scale loses, gains and rearrangements of the DNA that result in Chromosomal Instability (CIN) and aneuploidy (i.e. change in chromosomal number). A critical event in the evolution of OAC genomes is whole genome doubling (WGD), an endoreduplication of the genetic material that occurs in about 30-50% of OAC genomes. WGD occurs at the early stages of Barrett's dysplasia giving rise to molecularly distinct genomic asset that evolves into invasive adenocarcinoma. We developed *in vitro* models to trace the evolution of OAC genomes after WGD and identify mechanisms of dependency and maintenance of WGD. We validated the findings in a large series of whole genome sequences (WGS) from OAC patients and patient-derived organoids(n=651patients). **Methods**: We selected a non-dysplastic cell line of Barrett's oesophagus (CP-A) and a diploid oesophageal cancer cell line (OACM5.1). Isogenic WGD models were isolated and FACS sorted according to DNA content and grown and clonal cultures. WGD clones were verified with Flow cytometry, metaphase count and centromeric FISH. Corresponding CRISPR/Cas9 TP53 knock out (TP53KO) models were generated for CP-A. DNA and RNA from each clone were collected longitudinally and characterised with single cell shallow genome sequencing, long read sequencing and RNAseq. In addition, we characterised a large cohort (496 patients) of whole genome sequences and RNAseq of Oesophageal cancers with a custom bioinformatic pipeline.**Results**: We isolated multiple clones of WGD isogenic models of CP-A and OACM5.1. WGD is a rare event occurring in <1% of live OAC cells. Single Cell Karyoseq analysis confirmed the acquisition of genomic instability after WGD. TP53 mutation dramatically increases chromosomal instability and accelerates the acquisition of arm wide chromosomal alterations. We estimated the absolute copy-number and presence of WGD in WGS from the OCCAMS consortium (UK, 496 patients), the 100k genome project of Genomics England (UK, 120 patients) and tumour derived organoids (35 patients). We identified three subgroups of patients (diploid, WGD-derived aneuploidy, non-WGD derived aneuploidy) with specific transcriptional profiles. **Conclusions**: Our data identify transcriptomic changes and recurrent patterns of genomic evolution associated with WGD that have implication for patient stratification and treatment.

Session Title: Cancer Poster Session III

PB5037 Ewing sarcoma with two fusion transcripts involving *EWSR1*, *AP1B1*, and *ERG* genes: Further evidence of *EWSR1::ERG* fusions arising from complex genomic mechanisms.

Authors:

N. Uddin, A. L. Sengupta, M. Nagel, J. Y. Park; UT Southwestern Med Ctr. at Dallas, Dallas, TX

The molecular hallmark of Ewing Sarcoma (ES) is *EWSR1* gene rearrangement resulting in its fusion with various partner genes. *EWSR1::ERG* fusion is the second most common rearrangement in ES (~ 5% of the cases). A few sequencing studies have demonstrated complex genomic rearrangements result in *EWSR1::ERG* fusion. Here we present a patient with *EWSR1::ERG* fusion positive ES that was found to have two fusion transcripts involving *EWSR1*, *AP1B1*, and *ERG* genes. The patient is a 17-year-old male who presented with a posterior right flank subcutaneous cystic mass, which was subsequently excised (measuring 22.0 x 11.0 x 8.0 cm). No metastatic disease was found on further evaluation. The histopathological examination of the tumor showed a small round cell tumor, diffusely and strongly positive for CD99 (membranous) and NKX2.2 (nuclear) by immunohistochemistry. Fluorescent in situ hybridization (FISH) performed on formalin fixed paraffin embedded (FFPE) tumor was negative for *EWSR1* rearrangement. RNA sequencing of the FFPE tumor tissue using a clinically validated, targeted anchored multiplex next-generation sequencing (NGS) panel showed two fusion transcripts *EWSR1*:NM_013986.3(Exon 8)::*ERG*:NM_001136154.1(Exon 12) and *AP1B1*:NM_001127.3(Exon 1)::*ERG*:NM_001136154.1(Exon 4). Targeted reverse transcription polymerase chain reaction (RT-PCR) confirmed the NGS results.
The *EWSR1::ERG* fusion was consistent with the diagnosis of ES. The *AP1B1::ERG* is a novel fusion transcript. It likely represents a secondary fusion event in context of *EWSR1::ERG* fusion, in this case. Due to opposite orientation of *ESWR1 (22q12.2)* and *ERG (21q22.2)* on their respective loci, fusion between these two genes was thought to occur by complex genomic rearrangements involving inversions and multiple chromosomal breaks, instead of simple reciprocal translocation. Few NGS studies of ES with *EWSR1::ERG* fusion have shown complex rearrangement mechanisms, including chromothripsis or chromoplexy resulting in *EWSR1::ERG* fusion, confirming this hypothesis. Our case also illustrates a complex rearrangement mechanism resulting in *EWSR1::ERG* and *AP1B1::ERG* fusions. *AP1B1* is near to *EWSR1* on *22q12.2* locus (about 27 kb downstream / telomeric), and is oriented opposite to *EWSR1*, in the same orientation as *ERG*. Therefore *AP1B1::ERG* fusion, in addition to *EWSR1::ERG* fusion is indicative of further rearrangements of *22q12.2* locus and *ERG* gene, possibly involving multiple chromosomal breaks and inversions. Some studies have shown higher chances of relapse in ES patients with complex genomic rearrangements. Our patient is currently undergoing treatment for ES.

Session Title: Cancer Poster Session I

PB5038 Examination of modifiable risk factors and risk of mosaic chromosomal alterations in the UK Biobank

Authors:

C. Young[1], A. Hubbard[2], P. Saint-Maurice[1], S. Khan[3], C. Matthews[1], S. Moore[1], E. Loftfield[4], M. Machiela[4]; [1]NCI-DCEG, Rockville, MD, [2]NCI, Montgomery Village, MD, [3]Natl. Cancer Inst., Bethesda, MD, [4]Natl. Cancer Inst., Rockville, MD

Observational studies suggest modifiable exposures such as sleep, physical activity and alcohol use could impact cancer risk, but biologic mechanisms underlying these associations are poorly characterized. Mosaic chromosomal alterations (mCAs), a type of clonal hematopoiesis in which a detectable fraction of circulating leukocytes harbor large structural changes to chromosomes, present an opportunity to examine the association between modifiable risk factors and clonal expansion in the hematopoietic compartment. We utilized raw intensity and phase information from blood-derived DNA genotyping array data of 485,028 participants free of hematologic malignancies in the UK Biobank (UKBB) to detect mCAs. Modifiable risk factors were self-reported at baseline. Multivariable logistic regression models adjusted for age, sex, smoking, and ancestry were used to evaluate the association between modifiable exposures and risk of detectable mCAs.
In total, 11,826 autosomal mCAs, 15,499 loss of the X chromosome (mLOX) and 43,044 loss of the Y chromosome (mLOY) events were detected. For sleep patterns, we observed evidence for a reduced risk of mLOY in individuals getting >9 hours of sleep compared to individuals getting 7-9 hours of sleep, but limited evidence for associations with autosomal mCAs and mLOX. For physical activity, we noted an association between high levels of vigorous activity (e.g., >2 hours daily) and increased mLOY, but no evidence for associations with autosomal mCAs or mLOX. For alcohol use, multiple lines of evidence indicated potential associations between alcohol intake on increased autosomal mosaicism, mLOX and mLOY. These observations provide initial evidence suggesting possible associations of modifiable risk factors with mCAs. As mCAs are intermediate markers of hematologic cancer risk, mCAs could highlight selective mechanisms relevant to clonal expansion as a potential mechanism by which modifiable exposures contribute to cancer risk.

Session Title: Cancer Poster Session II

PB5039 Expanding the spectrum of germline pathogenic variants among Mexican patients with colorectal cancer

Authors:

D. Aguilar[1], R. Ortiz López[2], A. Rojas-Martínez[2]; [1]Hosp. Zambrano Hellion Tec Salud, San Pedro Garza García, Mexico, [2]Escuela de Med. y Ciencias Médicas del Tecnológico de Monterrey, Monterrey, Mexico

Backorpund:Colorectal cancer (CRC) is the third most common cancer in the worl and tha second leading cause of death due to cancer. 5-10% of CRC are explained by hereditary predisposition syndromes. The most frequent being Lynch syndrome (LS) followed by Familial Adenomatous polyposis (FAP). The purpose of this study is to describe the spectrum of germline pathogenic variants identified among Mexican CRC patients using Next-generation sequencing.Methods: 81 patients with early-onsent and/or family history of CRC were enrolled. All the patients signed the informed consent. NGS was performed using cancer sequencing panel (Illumina) wich targeted a set of 94 genes known to play a role in cancer predisposing. Bioinformatics data analysis was carried out using Pediatric Cancer Sequencing Pathogenicity Information exchange (PeCanPIE). Variant were annotated according to nomenclature recommendations from Human Genome Variation Society and further classified according to the 5 tier ACMG class system.Results: Mean age of CRC diagnosis was 44 years old, 39 (48.15%) were woman and 42 (51.85%) were man. 32 (39.5%) had family history of CRC, 15 fulfilled Amsterdam II criteria and 74 (91.4%) the Bethesda criteria. We identified 28 pathogenic/likely pathogenic variants in 27 individuals, in 11 Lynch Syndrome was molecular confirmed. 17 PV variants were identified in genes typically and not-tipically associated with CRC: ATM(2), CHEK2 (3), MUTYH (3), STK11 (1), BRCA (1), FANCA (1), MEN1 (1), PPMD1 (1), PRF1 (2), RUNX1 (1) and TP53(1). We identified 7 novel variants potentially pathogenic according to in silico and clinical data. Conclusions: NGS technology allow the discover of pathogenic in genes not typically associated with CRC as well as novel variants potentially pathogenic. The better understanding of genes involved in hereditary CRC could be used to identify new biomarkers as well as design target therapies. Fanconi Anemia pathway genes emerges as new candidate genes associated with hereditary CRC, previous evidence supports these observations.

Session Title: Cancer Poster Session III

PB5040 Exploring the non-invasive potential of miRNA 145 and miRNA 363 in Prostate Cancer

Authors:

A. Manoj, M. Ahmad; King George's Med. Univ., Lucknow, India

Owing to the misdiagnosis and over-treatment of prostate cancer, due to its indolent growth and false-positive errors produced by diagnostic ways like PSA level, DRE, and biopsy. Consequently, there is an urgent need for reliable diagnostic and detection technologies to diagnose aggressive prostate cancer non-invasively by exploring the potential of circulating microRNAs. Efficient and reliable markers within the body fluids can help in personalized treatment decisions for monitoring disease and survival. MicroRNA (miRNA) are highly conserved small non-coding RNA that modulate genes involved in numerous biological processes and form part of complex networks that play a significant role in prostate cancer initiation and progression. In keeping this view, we have designed our study to investigate biomarker potential miRNA 145 and miRNA 363 in prostate cancer. We determined the comparative expression of miRNA 145 in benign prostatic hyperplasia (BPH) and prostate cancer blood samples by Real Time-PCR and performed the Receiver Operating Characteristics (ROC) curve analysis. Further, we explored target genes of miRNA using In-Silico tools like TargetScan, miRDB, and miRWalk3.0 and performed gene and pathway enrichment analysis using DAVID 6.8. Our observation suggests a significant tumor-suppressive role of miRNA 145 and the oncogenic role of miRNA 363 in prostate cancer compared to BPH. Computational analysis reveals that miRNA 145 and miRNA 363, both target genes like e2f3, nras, pten, foxo1 etc., play a significant role in prostate cancer survival and progression. Finally, ROC curve analysis predicts that miRNA 145 (AUC 0.822***) and 363 (AUC 0.852***) have good biomarker potential for distinguishing BPH from prostate cancer. Collectively these results reveal that miRNA 145 and miRNA 363 expression may act as a non-invasive biomarker for effective diagnosis of Prostate Cancer. For future suggestions, the target genes evaluation may provide therapeutic targets for better treatment and management of prostate cancer.

Session Title: Cancer Poster Session I

PB5041 Exploring the population-specific landscape of germline genetic drivers of neuroendocrine neoplasms in Mexican Mestizos.

Authors:

L. Hernández Ramírez[1], R. G. Rebollar-Vega[1], M. Torres-Morán[1], A. L. Franco-Álvarez[1], A. I. Manzo-Osnaya[1], E. Martell-Jiménez[1], J. M. Zuarth-Vázquez[2], W. E. Hernández-Núñez[2], V. E. Juárez-Rivera[3], D. Morales-Lara[3], D. Sánchez-Villa[2], J. Eseiza-Acevedo[2], D. Cuevas-Ramos[4], Á. F. Hernández-Martínez[5], I. Hernández-García[5], E. Sosa-Eroza[5], E. Espinosa-Cárdenas[5], G. Vargas-Ortega[5], B. González-Virla[5], M. García-Sáenz[5], L. Balcázar-Hernández[5], A. Gamboa-Domínguez[6], M. Mercado[7], C. Ramírez-Rentería[7]; [1]Red de Apoyo a la Investigación, Univ. Natl. Autónoma de México, e Inst. Natl. de Ciencias Médicas y Nutrición Salvador Zubirán, Mexico City, Mexico, [2]Dept. of Endocrinology, Salvador Zubirán Natl. Inst. of Med. Sci. and Nutrition, Mexico City, Mexico, [3]Dept. of Genetics, Salvador Zubirán Natl. Inst. of Med. Sci. and Nutrition, Mexico City, Mexico, [4]Neuroendocrinology Clinic, Dept. of Endocrinology, Inst. Natl. de Ciencias Médicas y Nutrición Salvador Zubirán, Mexico City, Mexico, [5]Dept. of Endocrinology, Specialties Hosp. Centro Médico Natl. Siglo XXI, Mexican Inst. of Social Security, Mexico City, Mexico, [6]Dept. of Pathology, Salvador Zubirán Natl. Inst. of Med. Sci. and Nutrition, Mexico City, Mexico, [7]Unidad de Investigación Médica en Enfermedades Endocrinas, Hosp. de Especialidades, Centro Médico Natl. Siglo XXI, Inst. Mexicano del Seguro Social, Mexico City, Mexico

Introduction. Neuroendocrine neoplasms (NENs) are heterogeneous tumors with variable clinical behaviors. Given that NENs are frequently inherited, genetic tests are key for early diagnosis, tailored clinical management, and counseling. Yet, platforms for genetic diagnosis are not widely available and current data on disease drivers do not accurately represent all human populations. *Aims.* To report preliminary data on the repertoire and frequency of germline defects underlying NENs in a prospective cohort of Mexican Mestizos. *Methods.* Adult individuals with diagnosis of bronchopulmonary, gastrointestinal, pancreatic, pituitary, and thymic NENs, medullary thyroid carcinoma, primary hyperparathyroidism, and paragangliomas or pheochromocytomas, including sporadic, familial, isolated, and syndromic presentations were recruited from two reference hospitals in Mexico City. Under informed consent, DNA samples from blood were obtained. The study started in August 2022 and is designed as a prospective 15-year cohort. A next generation sequencing (NGS) panel was designed and optimized *ad hoc*. Variants with frequency <0.1% in public databases were classified in accordance with the American College of Medical Genetics and Genomics and Association for Molecular Pathology guidelines, using *in silico* analyses, ClinVar reports, and experimental and clinical data from the literature. *Results.* As of April 2023, 79 individuals (69.2% women) with mean age at recruitment of 50.4±17.6 years had been included. Disease presentation was sporadic in 44.6% of cases; among those with familial presentation, 21.6% were from kindreds with cancer predisposition. The presentation was multiple endocrine neoplasia type 1 in 20.3% of participants, the rest were distributed among 18 other phenotypes. Two individuals underwent testing in external laboratories, one with a commercial NGS panel and one via Sanger sequencing. Forty-five DNA samples were sequenced by means of our custom NGS panel, obtaining ~300x depth of sequencing of target regions. Likely germline drivers of NENs were identified in 17 samples (36.2%), including four cases in which the genetic defects were not expected from the phenotypes. The affected genes were: *AIP*, *CDKN1B*, *DNMT3A*, *MEN1*, *NF1*, *PIK3CA*, *RET*, *SDHA*, *SDHC*, *SDHD*, and *VHL*. Likely population-specific variants were identified among both tumor drivers and benign variants. *Conclusions.* Germline drivers of NENs were found in one-third of individuals from a genetically diverse population that is underrepresented in international studies. Our data support the idea that genetic drivers underlying phenotypes of NENs might differ among populations.

Session Title: Cancer Poster Session II

PB5042 Exploring the Potential: Evaluating the Celemics WES panel for NGS-based Comprehensive Cancer Analysis and Precision Medicine

Authors:

H. Jang, H. Seo, J. Jeong, H. Kim; Celemics Inc., Seoul, Korea, Republic of

Whole exome sequencing (WES) has emerged as a powerful tool for comprehensive genomic analysis, enabling researchers to study genetic variations associated with various diseases, including cancer. In this study, our aim was to present the performance of the Celemics Whole Exome Sequencing (WES) panel, which is the most comprehensive WES panel covering all target regions of major WES panels available in the market. With a target size of 37.1 Mb, the panel does not compromise performance in terms of coverage and uniformity, enabling highly efficient and cost-effective sequencing of the human whole exome, along with complete in-house bioinformatics solutions provided by Celemics Analysis Service (CAS). Featuring these distinctive characteristics, we evaluated the WES panel using OncoSpan FFPE (Cat. ID: HD832, Horizon Discovery, USA) reference material. OncoSpan FFPE is a well-characterized, cell line-derived reference standard containing over 380 variants in 152 cancer-associated genes. The results of the Celemics WES panel showed exceptional sequencing performance, successfully detecting all relevant known variants in the reference sample, even in hard-to-capture regions such as GC-rich and homologous regions. The outstanding sensitivity and specificity, as well as the superior on-target ratio, indicate that the Celemics WES panel can not only be used for research and molecular diagnostics in the field of oncology but also for discovering cancer-associated novel SNPs or conducting comprehensive exome analysis, bringing NGS one step closer to robust clinical diagnostics and precision medicine.

Session Title: Cancer Poster Session III

PB5043 Frequency of 50 bp Ins-del and rs4817415 variant of SOD1 gene and their association with vitiligo patients

Authors:

B. Gómez Meda[1], G. M. Zúñiga-González[2], L. D. Bañuelos-Díaz[3], M. P. Oceguera-Angel[3], A. Galindo-Gómez[3], M. G. María G. Sánchez-Parada[4], B. M. G. Torres-Mendoza[2], J. I. Delgado-Saucedo[5], M. G. Márquez-Rosales[2], A. L. Zamora-Perez[1], M. P. Gallegos-Arreola[2]; [1]Univ. de Guadalajara, Centro Univ.rio de Ciencias de la Salud, Guadalajara, Mexico, [2]Inst. Mexicano del Seguro Social, Centro de Investigación Biomédica de Occidente, Guadalajara, Mexico, [3]DermaMed Clinic, Guadalajara, Mexico, [4]Univ. de Guadalajara, Centro Univ.rio de Tonalá, Guadalajara, Mexico, [5]Univ. de Guadalajara, Centro Univ.rio de Ciencias Exactas e Ingenierías, Guadalajara, Mexico

Vitiligo is a common, systemic, idiopathic, autoimmune disease affecting skin, hair and mucous membranes. Among the causal factors of the disease, oxidative stress plays an important role, which also involves the imbalance of cellular homeostasis, causing damage to the cell and to the integrity of the genetic material, as well as to multiple processes of cell survival and risk of carcinogenesis. Therefore, the aim of the present study was to analyze the frequency of SOD1 gene variants (50 bp Ins/Del and rs4817415 in patients with a previous diagnosis of vitiligo and in a control group. Blood samples were collected from patients with a clinical diagnosis of vitiligo, adults of both sexes, who underwent analysis of gene variants by real-time-PCR, and were compared with a reference group without the disease. The population analyzed were young people aged 29 to 37 years with a higher percentage of women. The population was found in HWE. The 50 bp Ins/Del and rs4817415 variants showed no significant difference between groups ($p>0.05$). In conclusion, the 50 bp Ins/Del and rs4817415 variants of the SOD1 gene showed no association in patients with vitiligo in the sample analyzed.

Session Title: Cancer Poster Session I

PB5044 Frequency of pathogenic and likely pathogenic variants in genes associated with an increased risk for cancer in patients undergoing population based carrier screening.

Authors:

E. Cooper, L. Petrarca; LabCorp, Philadelphia, PA

INTRODUCTION: Carrier screening is used to identify individuals at risk for having children with autosomal recessive or X-linked genetic disorders. In some instances, carrier testing may reveal unintended findings, including identifying individuals who have an increased risk to develop cancer. The objective of this study was to investigate the frequency at which individuals test positive on carrier screening for pathogenic and likely pathogenic variants in genes associated with a risk for malignancies. These types of results highlight the need for appropriate pre and posttest counseling. METHODS: Carrier screens (n=1581), collected from July to November 2021, were analyzed to identify carriers of 4 conditions (6 genes) associated with cancer predisposition: Ataxia-telangiectasia (ATM), Bloom syndrome (BLM), Fanconi anemia (FANCL, BRIP1, FANCC), and Nijmegen breakage syndrome (NBN). All patients had the same carrier screening panel drawn which included over 500 genes. RESULTS: Out of the 1581 carrier screens, we identified 20 patients (1.3%) who were carriers for one or more genes associated with cancer predisposition. The number of carriers for each gene is as follows: ATM (8), BLM (3), FANCL (3), BRIP1 (2), FANCC (2), NBN (2). CONCLUSIONS: These data suggest that approximately 1.3% of individuals who undergo a carrier screening panel of this size will be positive for a variant that may increase their risk to develop cancer. These findings are outside the original intention of carrier screening which is to identify reproductive risks. Additionally, they underscore the need for appropriate pre-test counseling that includes a discussion of the possibility of identifying a variant associated with an increased risk for malignancy. Post-test counseling regarding these results is critical, not only to review reproductive risks, but also to review potential implications for one’s own healthcare.

Session Title: Cancer Poster Session II

PB5045 Frequency of the rs8720 variant of the KRAS gene in patients with colorectal cancer from a Mexican population

Authors:

M. Rentería-Ramírez[1], J. I. Cruz-Sanchez[2], C. A. Ronquillo-Carreón[2], A. F. Garibaldi-Ríos[1], G. M. Zúñiga-González[3], L. E. Figuera[1], B. C. Gómez-Meda[1], J. I. Delgado-Saucedo[4], M. P. Gallegos-Arreola[1]; [1]División de Genética, Centro de Investigación Biomédica de Occidente (CIBO), Centro Médico Natl. de Occidente (CMNO), Inst. Mexicano del Seguro Social (IMSS), Guadalajara, Jalisco, México., Guadalajara, Mexico, [2]Inst. Mexicano del Seguro Social, Guadalajara, Mexico, [3]División de Med. Molecular, CIBO, CMNO, IMSS. Guadalajara, Jalisco, México., Guadalajara, Mexico, [4]Laboratorio de Inmunofarmacología, Centro Univ.rio de Ciencias Exactas e Ingenierías UdeG, Guadalajara, Jalisco, México., Guadalajara, Mexico

Background: Colorectal cancer (CRC) is considered a public health problem, as it is one of the most common malignant neoplasms in developed countries and Mexico is no exception. CRC represents the third most common malignant neoplasm in the world and the second cause of death. Different studies have evaluated the association of variants in the KRAS gene with various diseases. In this sense, the rs8720 T>C variant in the KRAS gene (identified in Kirsten rat sarcoma virus) has been associated with risk susceptibility to cancer. However, in patients with CRC from the Mexican population, this association is unknown.**Objective:** To evaluate the association of the rs8720 T>C variant of the KRAS gene in patients with colorectal cancer from Western Mexico.**Methodology:** 386 CRC genomic DNA samples were included and 311 for the control group. Genomic DNA samples from both study groups were analyzed to verify the amount of DNA by spectrophotometry. Using the real-time Polymerase Chain Reaction (PCR) method, the alleles and genotypes of the rs8720 variant of the KRAS gene were identified in the study groups.**Results:** the CC genotype (variant) was observed in 38% (146/346) of the patient group and 26% (82/311) of the control group (OR 1.69, 95%CI 1.22-2.35, p=0.001) showed as a risk factor, TT wild-type genotype was present in 23% (89/386) of the patient group and 24% (74/311) of the controls (OR 0.95, 95%CI 0.67-1.36, p=0.889), without showing significant differences between the study groups; and the heterozygous genotype (TC) was present in 43% (151/346) of the patient group and 52% (155/311) of the controls, (OR 0.64, 95%CI 0.47-0.87, p=0.005) was observed as a protective factor for the association to development to CRC in the sample analyzed.**Conclusions:** The variant rs8720 of the KRAS genes was associated as a risk factor for susceptibility to the development of CRC in the analyzed sample.

Session Title: Cancer Poster Session III

PB5046 Functional annotation and gene mapping of gastric cancer

Authors:

J. Lee, S. Moon, K. Jung, H. Kimma, S. Jee; Yonsei Univ., Seoul, Korea, Republic of

Background: Gastric cancer (GC) is the fifth most frequently diagnosed malignancy and the third leading cause of cancer mortality worldwide. However, the exact etiology of gastric cancer is not clear. **Method:** A large-scale genome-wide scan was performed in gastric cases and controls from KCPS-II cohort. To explore the roles of gene mapping in gastric cancer, a genome-wide association study of the gastric cancer GWAS re-evaluated by the FUMA. **Results:** A total of 65 independent significant single nucleotide polymorphisms (SNPs) and 42 lead SNPs were identified in 39 genomic risk loci. Out of these identified genomic risk loci, 67 candidate causal genes were screened by positional mapping. We identified 26 genes via positional mapping and expression quantitative trait locus (eQTL mapping). In gene-based analysis, THEM6 gene had the lowest P-value (P-value=2.47x10-11), followed by TRIM46 gene (P-value=1.34 x 10-10) and THBS3 gene (P-value=2.59 x 10-10). Among prioritized genes, a gene set consisting with THEM6, PSCA showed strong enrichment related to chr8q24 which was previously reported cancer-related loci. Furthermore, a gene set including THBS3 gene showed strong enrichment related to BREAST_CANCER_1Q21_AMPLICON pathway. Functional annotation of these prioritized genes revealed that chr1q22 and chr8q24 positional pathway was implicated in the onset of gastric cancer. Overall, our findings suggested novel candidate gene and demonstrated that the several pathway has a crucial role in the risk of gastric cancer.

Session Title: Cancer Poster Session I

PB5047 Functional characterization of non-coding *CDKN2B* single nucleotide variants identified in high-risk melanoma families.

Authors:

A. Dennis[1], R. C. Hennessey[1], J. W. H. Wong[2], J. A. Barbour[2], R. Chari[3], K. Jones[1,4], W. Luo[1], M. R. Sargen[1], M. A. Tucker[1], A. M. Goldstein[1], X. R. Yang[1], K. M. Brown[1]; [1]Natl. Cancer Inst., Rockville, MD, [2]The Univ. of Hong Kong, Hong Kong SAR, China, [3]Natl. Cancer Inst., Frederick, MD, [4]Natl. Cancer Inst., Bethesda, MD

Melanoma is a common cancer in which approximately 10% of cases are familial. Of these familial cases, roughly 20-40% are explained by variation in the high-penetrance susceptibility gene *CDKN2A* (which encodes tumor suppressors p14 and p16) and another ~10% by other susceptibility genes. While the remaining families are unexplained by protein-coding variants in known high-penetrance genes, some nonetheless show genetic linkage to the 9p21 locus, which harbors *CDKN2A* and another tumor-suppressor gene, *CDKN2B* (which encodes p15). To examine the apparent predisposition to melanoma in these families, we conducted whole-exome sequencing. Sequencing germline DNA of high-risk melanoma families from the United States identified two families (one with 7 and one with 3 sequenced melanoma patients) harboring distinct cosegregating rare genetic variants clustered closely together at the 9p21 locus, along with a third rare variant found in a melanoma case from an American case-control study. These single nucleotide variants (SNVs) are all within 10 base pairs of each other and are located within the first intron of *CDKN2B*. Notably, these SNVs are found within consensus binding sites for CCCTC-Binding factor (CTCF), which modulates chromatin interactions necessary for gene expression. Consistent with a potential role for altered CTCF binding, ChIP-seq (chromatin immunoprecipitation with sequencing) in fibroblasts from one family showed significant loss of CTCF binding on the familial variant allele, coupled with a slight decrease in the number of interactions across the region where the variant was located. We hypothesized that these *CDKN2B* SNVs may influence melanoma risk because of their ability to disrupt CTCF binding, impacting chromatin interactions necessary for the expression of *CDKN2A* and/or *CDKN2B*. To test this hypothesis, we performed CRISPR-Cas9-based genome editing of immortalized human melanocytes. We created multiple melanocyte clones missing a 107 bp region around this CTCF binding region and also edited in each of these three variants using separate homology vectors with the Lig4 inhibitor SCR7 to promote homology-directed repair. We are presently performing variant characterization (comparing these to wild-type clones), including assessing their effects on expression of nearby genes and determining whether they function via *CDKN2A*, *CDKN2B*, or both. We are also evaluating phenotypic consequences of these variants, including melanocyte proliferation and region-specific chromatin interactions. These results have the potential to shed light on the mechanistic underpinnings of non-coding variants linked to familial melanoma cancers.

Session Title: Cancer Poster Session II

PB5048 Functional copy number alterations as diagnostic and prognostic biomarkers in neuroendocrine tumors.

Authors:

H. Vaughn[1], H. Major[2], E. Kadera[2], K. Keck[3], T. Dunham[2], Q. Qian[2], B. Brown[4], L. Borbon[3], A. Scott[3], A. Belizzi[5], T. Braun[1], P. Breheny[6], D. Quelle[7], J. Howe[3], B. Darbro[2]; [1]Univ. of Iowa Interdisciplinary Graduate Program in Genetics, Iowa City, IA, [2]Stead Family Dept. of Pediatrics, Iowa City, IA, [3]Univ. of Iowa Hlth.care Dept. of Surgery, Iowa City, IA, [4]Univ. of Iowa Dept. of BioMed. Engineering, Iowa City, IA, [5]Univ. of Iowa Dept. of Pathology, Iowa City, IA, [6]Univ. of Iowa Dept. of Biostatistics, Iowa City, IA, [7]Univ. of Iowa Dept. of Pharmacology and NeuroSci., Iowa City, IA

Neuroendocrine tumors (NETs) can arise in multiple organs throughout the body. Sixty percent of NET patients have advanced metastatic liver disease before they are diagnosed, often leading to exploratory procedures to determine primary tumor site. Functional copy number alterations (fCNAs) are DNA copy number changes with concordant differential gene expression. Given that fCNAs are less likely to be bystander genetic lesions, we hypothesized that they could be used as diagnostic (site of origin) and prognostic (event free survival) NET biomarkers. To identify candidate fCNAs, we integrated chromosomal microarray (CMA) and RNA-seq differential gene expression data from 31 pancreatic (pNET) and 33 small bowel neuroendocrine tumors (sbNET) as well as 47 early disease progression (<24 months) and 17 late disease progression (>24 months) patients. Candidate fCNAs that could accurately differentiate these groups were then replicated using fluorescence in situ hybridization (FISH) on formalin-fixed, paraffin-embedded (FFPE) tissues in a larger cohort of 86 pNETs, 114 sbNETs, 75 early disease progression, and 100 late disease progression samples. Logistic regression analysis was performed to determine the predictive ability of these biomarkers as well as the assay performance metrics of sensitivity, specificity, and area under the curve. Our results indicate that fCNAs at chromosomal loci 4p16.3, 5q32, 7q31.2, 9p21.3, 17q12, 18q21.2, and 19q12 may be used as diagnostic and prognostic NET biomarkers and can be utilized in clinical testing to aid in determination of primary tumor site for patients with metastatic liver disease of pNET or sbNET origin and may for the basis for risk stratified therapies in the future.

Session Title: Cancer Poster Session III

PB5049 Gene co-expression networks capture potential biomarkers of UTUC.

Authors:

T. Fu, Y. Lin, Y. Yang, N. Xiang, L. Du, G. Liao, J. Huang, B. Liao; West China Hosp., Sichuan Univ., Chengdu, China

Background Upper tract urothelial carcinoma (UTUC) accounts for 5%-10% of all urothelial malignancies. And the incidence of UTUC in China is much higher than that in European and American countries, accounting for about 31%. Approximately two-thirds of patients have invasive disease at diagnosis. However, little research has been done on UTUCs. **Patients and methods** To better understand the pathogenesis of UTUC and provide reference for its diagnosis and treatment, we conducted a Weighted Correlation Network Analysis (WGCNA) involving the RNA sequencing data of 27 UTUC patients. Multiple linear regression models were used to explore the association between gene modules and clinical indicators. We then imported related modules into Cytoscape 3.7.1. to screen hub genes. Additionally, we performed RNA differential expression analysis, exon mutation analysis, and looked at mutations in hub genes. Enrichment analysis was then conducted based on differentially expressed genes to discover potential pathways regulating UTUC. Finally, bladder cancer data in TCGA public database was used to verify the hub genes. **Results** We classified the genes into 31 modules by unsupervised clustering, with 4 modules (FDR<0.05) significantly associated with tumor infiltration. Then 9 potential biomarkers, including *MSRB3, HSPB2, SCRG1, SYNPO2, FXYD1, PLXNA4, CD163, IFFO1,* and *FPR3,* were discovered by importing these 4 modules into Cytoscape software. Differential expression analysis showed that except for the *FPR3* gene, the remaining 8 genes were underexpressed in UTUC. This has also been verified in public databases. Meanwhile, SNP mutation occurred in *MSRB3, SYNPO2,* and *PLXNA4* genes. Functional enrichment analysis revealed that differentially expressed genes regulate UTUC through pathways such as cell adhesion, transmembrane receptor protein kinase activity, cGMP-PKG signaling pathway, p53 signaling pathway, and cardiomyocyte adrenergic signaling pathway. **Conclusion** Although these 9 genes are new and significant genes in UTUC, they have been proved in public databases. Our results not only promote our understanding of the relationship between the transcriptome and clinical data in UTUC but will also guide the development of targeted molecular therapy for UTUC.

Session Title: Cancer Poster Session I

PB5050 Gene expression analysis of *miRNA 146b* and *miRNA 181b* in Papillary Thyroid Carcinoma

Authors:

R. Khan; Inst. of BioMed. & Genetic Engineering (IBGE) Islamabad Pakistan, Islamabad, Pakistan

Background: Thyroid cancer is considered the most common cancer and is the most recurrent malignancy of the endocrine system. The most frequent type is papillary thyroid cancer (PTC), contributing more than 80% globally with noticeable prevalence in females of Asia, America and Iceland. Genetic insights into molecular pathogenesis of PTC, have been getting a shape of rich library of useful information. The genetic elements known as MicroRNAs (miRNAs) are endogenous non-coding RNAs operating as post-transcriptional regulators involved in development, proliferation and differentiation. They are gaining fame as druggable biomarkers for clinical management of neoplasms. The momentous stackeholding of miRNA family by virtue of gene expression variations in the cancer microenvironment has been witnessed in the thyroid cancer onset, amplification and apoptosis. The growing body of knowledge highlights the modifiable play at the miRNA level harbors potential in lessening the perpetuation of the disease with safe handlers. The genetic information leads to a big highway which can replace the unified yardstick to tailor PTC with the more targeted personalized disease treatment by monitoring the disease risk and aggression modalities. **Objectives:** The study aims to speculate the characteristic involvement of expression level changes in the *miRNA-146b* and *miRNA-181b* as tangible biomarkers for PTC. **Methodology:** The present study was conducted on the PTC in Pakistan, a genetically less explored South Asian country. Specimen of the cancer tissue and healthy normal samples adjacent to the cancer part were collected from the PTC and nodular goiter patients undergoing thyroidectomy. The anthropometric and clinical parameters relevant to each patient were recorded after informed consent. Total RNA was isolated and first strand cDNA was synthesized from all the samples. Gene expression profile for *miRNA-146b* and *miRNA-181b* was done by quantitative Real-Time PCR. Relative gene expression was identified as fold change by utilizing essential gene expression levels. **Results:** The statistically significant changes in the relative expression of *miRNA-146b*; 5 to 20 folds and *miRNA-181b*; 4-60 folds were observed in PTC tissues in comparison to nodular goiter and healthy tissue specimens. **Conclusion:** The boosted gene expression of the *miRNA-146b* and *miRNA-181b* manifests the plausible misregulations in deployment of these molecular musketeers as foes in PTC. This forged maladaptation of the *miRNA-146b* and *miRNA-181b* in cancer microenvironment may warrant analytically, therapeutically and genetically surmountable miRNA targets for PTC clinical management prevention.

Session Title: Cancer Poster Session II

PB5051 Generalizability of PRS313 for breast cancer risk amongst non-Europeans in a Los Angeles biobank

Authors:

H. Shang[1], Y. Ding[2], V. Venkateswaran[2], K. Boulier[2], P. Boodaghi Malidarreh[3], J. Luber[3], B. Pasaniuc[2]; [1]Univ. of California Los Angeles, Los Angeles, CA, [2]UCLA, Los Angeles, CA, [3]Univ. of Texas Arlington, Arlington, TX

Polygenic risk scores (PRS) summarize the combined effect of common risk variants and are associated with breast cancer risk in patients without identifiable monogenic risk factors. One of the most well-validated PRSs in breast cancer to date is PRS313, which was developed from a Northern European biobank but has shown attenuated performance in non-European ancestries. We further investigate the generalizability of the PRS313 for American women of European (EA), African (AFR), Asian (EAA), and Latinx (HL) ancestry within one institution with a singular EHR system, genotyping platform, and quality control process. We found that the PRS313 achieved overlapping Areas under the ROC Curve (AUCs) in females of Lantix (AUC, 0.68; 95 CI, 0.65-0.71) and European ancestry (AUC, 0.70; 95 CI, 0.69-0.71) but lower AUCs for the AFR and EAA populations (AFR: AUC, 0.61; 95 CI, 0.56-0.65; EAA: AUC, 0.64; 95 CI, 0.60-0.680). While PRS313 is associated with Hormone Positive (HR+) disease in European Americans (OR, 1.42; 95 CI, 1.16-1.64), for Latinx females, it may be instead associated with Human Epidermal Growth Factor Receptor 2 (HER2+) disease (OR, 2.52; 95 CI, 1.35-4.70) although due to small numbers, additional studies are needed. In summary, we found that PRS313 was significantly associated with breast cancer but with attenuated accuracy in women of African and Asian descent within a singular health system in Los Angeles. Our work further highlights the need for additional validation in diverse cohorts prior to clinical implementation of polygenic risk scores.

Session Title: Cancer Poster Session III

PB5052 Generation of a functional genomics assay for the clinical re-classification of *PMS2* missense variants of undetermined significance.

Authors:

J. Plowman, E. Matoy, I. Scavuzzo, M. Belshan, H. Feser Stessman; Creighton Univ., Omaha, NE

Lynch syndrome (LS) is the most common hereditary cancer affecting an estimated 1 in 279 people. LS is caused by germline heterozygous mutations in mismatch repair (MMR) genes. While not directly cancer-causing, mutations in LS genes increase the number of somatic mutations over time, increasing cancer risk. *PMS2* is the most frequently mutated LS gene. However, the pathogenicity or functional effect of most missense *PMS2* variants is unknown leaving most classified as variants of uncertain significance (VOUSs). There is a great need to re-classify *PMS2* VOUSs for improved clinical diagnostics and management. We hypothesize that a portion of *PMS2* missense VOUSs will have decreased MMR activity, which will increase cancer risk over time. Our recent identification of a suspected hereditary cancer family carrying a missense variant in *PMS2* (p.Gly29Ala; MAF: 0.000440 across all races) had conflicting interpretations of pathogenicity in ClinVar, prompting the need for an *in vitro* assay by which *PMS2* variants could be functionally assessed in a high-throughput way. CRISPR/Cas9 engineering was used to knockout *PMS2* expression in the HAP1 haploid cell line (MMR proficient) by targeting exon 7 of the canonical transcript. Western blotting for *PMS2* showed full knockout in ~40% of the individual clones. MMR function-quantified by cell viability after 72 hours of 6-thioguanine (6-TG) treatment-showed that two independent *PMS2* KO lines were significantly more resistant to 6-TG compared to the wild-type HAP1 parent line ($p<0.0001$; Tukey's test) indicating reduced MMR activity. Full-length, wild-type (WT) human *PMS2* was cloned into the pCW57.1 Gateway vector and packaged into lentiviral particles. The *PMS2* KO lines were infected, selected with G418, and treated with doxycycline to induce WT *PMS2* expression. Induced WT *PMS2* expression rescued 6-TG sensitivity ($p<0.0001$; Tukey's test) in *PMS2* KO cell lines. Using site-directed mutagenesis on the pCW57.1-WT *PMS2* plasmid, we generated one benign and pathogenic-confirmed in ClinVar-missense variant as well as the p.Gly29Ala variant (VOUS) identified locally in 4 confirmed carriers of a suspected hereditary cancer family. Individual variants were again packaged into lentiviral particles and used to infect *PMS2* KO lines. While the known pathogenic variant conferred significantly decreased MMR sensitivity ($p<0.0001$; Tukey's test), the VOUS was not significantly different from WT *PMS2* suggesting a benign classification. In summary, we present a novel *in vitro* MMR assay for the high-throughput classification of human *PMS2* missense variants that can contribute to the re-classification of VOUSs identified in human patients.

Session Title: Cancer Poster Session I

PB5053 Genetic pathways of temozolomide resistance in glioblastoma.

Authors:

T. Carter[1], R. Valenzuela[1], S. Yerukala Sathipati[2], R. Medina-Flores[1]; [1]Marshfield Clinic, Marshfield, WI, [2]Marshfield Clinic Res. Inst., Marshfield, WI

Glioblastoma (GBM) tumors frequently reoccur despite aggressive surgical resection and temozolomide (TMZ) chemo-radiation, and the acquisition of TMZ resistance remains a barrier to extending survival in GBM. To discover genetic pathways involved in TMZ resistance, bioinformatics analysis was conducted to identify gene interactions predictive of GBM patient survival, considering that patients with shorter survival are more likely to be highly TMZ-resistant cases. Network-based stratification, a bioinformatics approach that has been used previously to identify gene subnetworks in cancer subtypes that are predictive of overall survival, was applied to the somatic variant data of 221 GBM patients who had undergone TMZ chemo-radiation following surgery, as provided in The Cancer Genome Atlas. Differentially mutated genes (DMGs) in each subtype were determined using the Statistical Analysis of Microarrays method, and the over-representation of these genes in biological pathways was assessed by Gene Ontology (GO) term enrichment analysis. Four GBM subtypes were identified, with a median overall survival of 12.8, 16.3, 15.6, and 18.3 months for subtypes 1 (n = 17), 2 (n = 54), 3 (n = 79), and 4 (n = 71), respectively (P = 0.032; log-rank test). The difference in overall survival between subtype 4 and subtypes 1 (P = 0.031) and 3 (P = 0.021) remained significant independent of other factors associated with GBM survival, including age, type of surgery, and radiation treatment, in Cox proportional hazards regression. Subsequently, the discovery of genetic pathways was focused on subtypes 3 and 4, that had a significantly different median overall survival but a similar proportion of subjects with *MGMT* promoter methylation (51.1% and 45.9%, respectively), which is correlated with increased overall survival in GBM. The top 10 GO terms ($P < 0.00001$) were related to mitotic DNA damage checkpoint signaling and telomere formation and maintenance for the 507 DMGs in subtype 3 and to intrinsic apoptotic signaling and DNA damage checkpoint signaling for the 86 DMGs in subtype 4. To validate the presence of somatic variation in the DMGs of subtypes 3 and 4 within other GBM cases, an independent sample of 37 GBM tumors from patients who had been treated with TMZ was exome sequenced. Somatic variants in 39 DMGs from subtype 3 and four DMGs from subtype 4 were detected in 25 and four of the tumors, respectively. In conclusion, DNA damage checkpoint signaling, apoptosis, and telomere regulation are potentially involved in TMZ resistance. The targeting of genes in these pathways could hold promise for developing new strategies to improve TMZ treatment efficacy and overall survival in GBM.

Session Title: Cancer Poster Session II

PB5054 Genetic profiling of human hepatocellular carcinoma by using advanced techniques

Authors:

R. Issa; Inst. of genetic engineering, Cairo, Egypt

In Egypt, Hepatocellular Carcinoma is a major health problem in which the most common cancer- related causes of morbidity and death, account for the fourth most common cancer compared to the sixth most common worldwide.Maintaining genomic integrity, preventing viral infection, and eliminating tumorigenesis are all dependent on cellular responses to DNA damage. Several genes are responsible for cellular responses and repairing DNA damage including TP53 and ATMTP53, also famous as “the guardian of the genome”, is a tumor suppressor gene that plays a key role in organizing cellular responses to various types of stresses by preserving genome safety and preventing the proliferation of mutated cells The ATM plays a critical mechanism as a central role in the repair of Double-strand DNA breaks (DSB). The DSB may be induced by many environmental factors, including HCV. DSB represent the greatest risk for causing genomic instability Circulating tumor DNA, a portion of the cfDNA, can be liberated into peripheral blood from the primary and metastatic tumor cells, which carries a recent image of tumor status and donates important value for molecular studies in a variety of cancers, including HCC In the present study, cfDNA samples of Egyptian HCC patients were sequenced by Ion Torrent NGS to identify the novel and existing variants in TP53 and ATM genes.MethodsClinicopathological data and 5 ml of whole blood were obtained from each 21 primary HCC patients. After extraction of cell-free DNA (cfDNA), the library construction was prepared and sequenced using Next Generation Sequencing (NGS) techniques.ResultsThe age of the studied HCC patients ranged from 48 to 80 years (Mean ± SD = 62.19 ± 9.08), thirteen (61.9%) patients were more than 60 years at diagnosis.Eighteen patients (85.70%) were males while only three patients (14.30%) were females. The Hepatitis C infection was reported in 18 patients (85.71%), and one case had co-infection (HCV and HBV). Four patients (19.05%) had a positive family history for cancers. Medical history was reported by 17 patients in the form of bilharziasis in 13 (61.90%) patients, diabetes in 7 (3.33%) patients and Hypertension in 3 (14.29%) patients.The distribution of somatic mutations in TP53 gene were: missense mutations (50%), intron mutations (21.88%), splicing mutations (9.38%), and 5′ UTR (9.39%), synonymous mutations (3.12%), INDEL (3.12%), and copy number variations (3.12%).On the other hand, the somatic mutations in ATM gene were distributed to be non-synonymous mutations (46.30%), synonymous mutations (22.22%), intron mutations (11.11%), copy number variations (1.85%), and other mutations (18.52%).

Session Title: Cancer Poster Session III

PB5055 Genetic re-assessment of *BRCA*-negative families in the Lynch Legacy Hereditary Cancer Biobank.

Authors:

H. Feser Stessman, J. Plowman, E. Matoy, C. Watson; Creighton Univ., Omaha, NE

Breast cancer remains the most common cancer diagnosed among women and is a leading cause of cancer-related deaths across all races and ethnicities. Approximately 20% of breast cancer cases are attributed to increased family risk; yet variation in *BRCA1/2* can only explain 20-25% of these cases. Until recently, single gene (i.e., *BRCA1/2*) and even single variant testing were common in at-risk family members with little additional follow-up sequencing. For many *BRCA* mutation negative (BRCAX) families, genetic risk remains ill-defined. We performed targeted sequencing of 28 high-risk familial hereditary breast and ovarian cancer (HBOC) genes using a custom molecular inversion probe (MIPs) panel in 246 human samples representing both cancer positive (n=174) and negative (n=72) individuals from 135 BRCAX HBOC families recruited from 1973-2014 by Dr. Henry Lynch. MIP probes were designed to span all RefSeq coding exons plus 10 base pairs of flanking introns to capture both coding and splice variation. Sequencing identified 391 high-quality variants which were annotated using the Ensembl Variant Effect Predictor (VEP) tool and further ranked based on their predicted clinical impact. All annotated pathogenic and variants of undetermined significance (VOUSs) were Sanger validated highlighting a SNV false positive rate <1%. Known pathogenic breast cancer variants in *CHEK2* (p.Thr410MetfsTer15 found in six unrelated families and p.Ser471Phe found in two families) were identified representing ~6% families in the study. While BRCAX was an inclusion criterion for this study, we still identified a pathogenic *BRCA2* variant (p.Met192ValfsTer13) in one family. A portion of BRCAX families could be explained by other hereditary cancer syndromes that increase HBOC risk including Li-Fraumeni Syndrome (gene: *TP53*) and Lynch Syndrome (gene: *MSH6*). Interestingly, these families carried additional VOUSs that may further modify the phenotypes of affected family members. Ten families carried more than one VOUS suggesting the presence of complex multi-variant families; a single VOUS variant was identified in 17 families. Overall, 8.1% of BRCAX HBOC families in our study may be explained by other known pathogenic variants; VOUSs could contribute up to an additional 18.5% of families which will require further functional testing. We conclude that our MIP approach is an inexpensive and scalable multigene screening tool that can be used to genetically re-evaluate suspected HBOC families. Further, genetic mining of historic hereditary cancer biobanks may still have great value in clinically re-defining VOUSs and in discovering novel genetic modifiers of disease.

Session Title: Cancer Poster Session I

PB5056 Genetic study in African ancestry populations identifies 16 novel loci associated with lung cancer risk.

Authors:

M. Betti[1], J. N. Hellwege[1], J. Jaworski[1], S. Zhao[1], J. Rao[2], B. Ryan[3], A. G. Schwartz[4], J. K. Wiencke[5], M. A. Bruce[6], S. Chanock[3], M. C. Aldrich[1]; [1]Vanderbilt Univ. Med. Ctr., Nashville, TN, [2]Univ. of Minnesota, Minneapolis, MN, [3]Natl. Cancer Inst., Rockville, MD, [4]Karmanos Cancer Inst., Detroit, MI, [5]Univ. of California, San Francisco, San Francisco, CA, [6]Univ. of Houston, Houston, TX

Introduction: The heritability of lung cancer is well established, yet prior lung cancer studies have been largely focused on populations of European descent, leaving the genetics of lung cancer in African ancestry persons poorly understood. We performed a genome-wide (GWAS) and transcriptome-wide association study (TWAS) across nearly 7,000 African ancestry individuals, representing the largest lung cancer genetics study performed to date in persons of African ancestry.

Methods: We analyzed data from 6,998 individuals of African ancestry (2,540 cases and 4,458 controls) collected from eight studies across the US. Imputed SNPs were evaluated for an association with lung cancer risk, adjusting for age, sex, smoking status, smoking pack-years, enrollment site, and the first three principal components. Linkage disequilibrium score regression (LDSC) was used to estimate the total narrow-sense heritability (h^2) of lung cancer captured by the GWAS, leveraging a novel linkage disequilibrium (LD) panel constructed using whole genome sequencing data from an independent set of 600 African ancestry individuals. Finally, TWAS was performed using models of genetically predicted expression with SNP covariance matrices of both African and European LD patterns. Significant associations were tested for causality using Mendelian randomization.

Results: We identified 34 loci associated with lung cancer ($p < 1 \times 10^{-6}$), including 11 which were genome-wide significant ($p < 5 \times 10^{-8}$). Of these loci, 18 have been previously associated with lung cancer, while the remaining 16 are novel. Of particular note are novel loci 5q22.1 (closest gene *CAMK4*), 7p22.2 (*SDK1*), 10q11.22 (*ARGHAP22*), 12q14.3 (*LINC02389*), 18q12.1 (*B4GALT6*), and 18q22.1 (*AC007631.1*), all of which were genome-wide significant. The captured h^2 was estimated to be 0.23 (0.10 SE). TWAS identified 8 gene-tissue pairs of suggested significance ($p < 1.85 \times 10^{-6}$) using African ancestry-based covariance matrices, 5 of which were predicted to be causal, and 76 using European ancestry matrices, 21 of which were predicted to be causal. With the TWAS results in lung tissue performed with both African and European ancestry covariance matrices, we also performed gene set enrichment and found a positive enrichment of genes associated with notable metabolic pathways and negative enrichment of genes associated with DNA repair pathways, suggesting potential biological mechanisms mediated by genetically regulated gene expression.

Conclusion: This study suggests novel loci for lung cancer in African ancestry individuals and provokes further studies to identify the genetic underpinnings in diverse ancestry individuals.

Session Title: Cancer Poster Session II

PB5057 Genetic susceptibility to nonalcoholic fatty liver disease and risk for pancreatic cancer: Mendelian randomization

Authors:

S. King[1], S. Veliginti[2], S. Antwi[2], Pancreatic Cancer Cohort Consortium (PanScan), Pancreatic Cancer Case Control Consortium (PanC4); [1]Mayo Clinic Graduate Sch. of BioMed. Sci., Jacksonville, FL, [2]Mayo Clinic, Jacksonville, FL

Background: There are conflicting data on whether nonalcoholic fatty liver disease (NAFLD) is associated with susceptibility to pancreatic cancer. Using Mendelian randomization (MR), we investigated the relationship between genetic predisposition to NAFLD and risk for pancreatic cancer. **Methods**: Data from genome-wide association studies within the Pancreatic Cancer Cohort Consortium (PanScan; cases n=5090, controls n=8733) and the Pancreatic Cancer Case Control Consortium (PanC4; cases n=4,163, controls n=3,792) were analyzed. We used data on 68 genetic variants with four different MR methods (inverse variance weighting [IVW], MR-Egger, simple median, and penalized weighted median) separately to predict genetic heritability of NAFLD. We then assessed the relationship between each of the four MR methods and pancreatic cancer risk using logistic regression to calculate odds ratios (ORs) and 95% confidence intervals (CIs), adjusting for pancreatic cancer risk factors, including obesity and diabetes. **Results**: No association was found between genetically predicted NAFLD and pancreatic cancer risk in the PanScan or PanC4 samples (e.g., PanScan, IVW OR=1.04, 95% CI: 0.88-1.22, MR-Egger OR=0.89, 95% CI: 0.65-1.21; PanC4, IVW OR=1.07, 95% CI: 0.90-1.27, MR-Egger OR=0.93, 95% CI: 0.67-1.28). None of the four MR methods indicated an association between genetically predicted NAFLD and pancreatic cancer risk in either sample. **Conclusion**: Genetic predisposition to NAFLD is not associated with pancreatic cancer risk. Given the close relationship between NAFLD and metabolic conditions, it is plausible that any association between NAFLD and pancreatic cancer might reflect host metabolic perturbations (e.g., obesity, diabetes, or metabolic syndrome) and does not necessarily reflect a causal relationship between NAFLD and pancreatic cancer.

Session Title: Cancer Poster Session III

PB5058 Genetic Testing for Hereditary Cancer Syndromes in Algerian population: Recurrent Germline Pathogenic Variants, Ethical and Social Issues

Authors:

F. Cherbal[1], C. Mehemmai[1], L-L. Saada[1], M. Nehar[1], F. Gouaref[2], K. Bentabak[2], F. Khider[1], A-L. Boumehdi[1], H. Mahfouf[3], F. Zebboudj[4], M. Oukkal[5], M. Maaoui[6], W. Benbrahim[7], K. Boualga[8]; [1]Molecular Genetics Team, LBCM, Faculty of Biological Sci., USTHB, Algiers, Algeria, [2]Oncological Surgery and Liver Transplantation services, EHS Pierre and Marie Curie, Univ. of Algiers 1, Algiers, Algeria, [3]Mohamed El Kolli Publ. Hosp., Academic Med. Oncology Services, Rouiba, Sch. of Med., Univ. of Algiers1, Algiers, Algeria, [4]Mohamed El Kolli Publ. Hosp., Academic Gen. Surgery Services, Rouiba, Sch. of Med., Univ. of Algiers-1,, Algiers, Algeria, [5]Clinic of Med. Oncology Amine Zirout, Univ. Hosp. of Beni-Messous, Sch. of Med., Univ. of Algiers-1, Algiers, Algeria, [6]Bachir Mentouri Publ. Hosp. , Academic Gen. Surgery Services, Kouba, Sch. of medicine, Univ. of Algiers-1,, Algiers, Algeria, [7]Med. Oncology Services, Ctr. Anti Cancer Batna, Univ. Batna 2 Chahid Mostefa Ben Boulaid, Batna, Algeria, [8]Radiation Oncology Services and Lab. for Cancer Res., Anticancer Ctr. of Blida, Blida, Algeria

Background:The effects of genetic counselling and testing for hereditary cancers in Algerian population are largely unknown. To date, the Algerian population is composed of native Berber descent with Sub-Saharan, European and Middle East elements. Genetic testing studies are of interest in admixed populations as most of genetic testing studies have been conducted in individuals of European ancestries. In Algeria, like in most of emerging countries, there is a limited access in clinical genetic services mainly due to the costs, the availability of genetic testing and the lack of public insurance for coverage of genetic services. Here, we report preliminary results of multi-center studies of the effects of genetic counselling and testing on hereditary cancer syndromes in Algerian patients. **Methods**: Since 2008, we conducted three research projects on genetic testing and counselling for hereditary cancer syndromes. We used PCR-Sanger sequencing and multi-gene hereditary cancer panel (Color Genomics) for testing *BRCA1/2* genes in 314 hereditary breast and ovarian cancer patients . We also tested *APC* gene and *MMR* genes in 54 severe familial adenomatous polyposis and 72-suspected Lynch syndrome patients in a national cohort and relatives at risk, respectively. **Results**: We detected six recurrent germline pathogenic variants in *BRCA1* gene: c.83_84del, c.181 T>G, c.798_799del, c.2125_2126insA, Del exon 7 and Del exon 15 and one recurrent germline pathogenic variant in *BRCA2* gene: c1310_1313del. In addition, we detected two recurrent germline pathogenic variants in *APC* gene: c.3927_3931del and c.4728dup and two recurrent germline pathogenic variants in *MLH1* gene: c.1546C>T and *MSH2* gene: c.942+3A>T, respectively. Our results showed that patients with limited background knowledge about genetics expressed positive attitudes towards genetic studies and genetic testing, with the possibility of preventing hereditary cancers in theirs siblings as main advantage. These patients also expressed a strong need for the protection of genetic information. Interestingly, the identification of recurrent germline pathogenic variants in *BRCA1/2* genes in Algerian population by our laboratory have already set up some need and demand from the Algerian patients living abroad. We have got request from these patients for genetic testing of specific or recurrent germline pathogenic variants in *BRCA* genes in their family relatives who are still living in Algeria. **Conclusions**: Continued efforts should be made by our national health care system to allow the access to affordable genetic counseling and testing for Algerian hereditary cancer syndromes' patients and their families.

Session Title: Cancer Poster Session I

PB5059 Genome-wide analysis to assess whether heavy alcohol intake modifies the association between SNPs and pancreatic cancer risk.

Authors:

Z. Ni[1], P. Kundu[2], D. McKean[1], W. Wheeler[3], The Pancreatic Cancer Case Control Consortium, The Pancreatic Cancer Cohort Consortium, L. T. Amundadottir[4], R. Z. Stolzenberg-Solomon[4], A. P. Klein[1,5,6]; [1]Dept. of Oncology, Sidney Kimmel Comprehensive Cancer Ctr., Johns Hopkins Sch. of Med., Baltimore, MD, [2]Dept. of Biostatistics, Johns Hopkins Bloomberg Sch. of Publ. Hlth., Baltimore, MD, [3]Information Management Services, Inc., Silver Spring, MD, [4]Div. of Cancer Epidemiology and Genetics, Natl. Cancer Inst., NIH, Bethesda, MD, [5]Dept. of Epidemiology, Johns Hopkins Bloomberg Sch. of Publ. Hlth., Baltimore, MD, [6]Dept. of Pathology, Sol Goldman Pancreatic Cancer Res. Ctr., Johns Hopkins Sch. of Med., Baltimore, MD

Background: Pancreatic cancer is a leading cause of cancer-related death globally. Risk factors for pancreatic adenocarcinoma, the most common subtype of pancreatic cancer, include common genetic variants and potentially heavy alcohol use. Objective: To assess whether heavy drinking modifies the association between genetic variants (SNPs) and pancreatic cancer risk. Methods: We conducted genome-wide single nucleotide polymorphism (SNP) by heavy alcohol use (greater than three drinks per day) interaction analysis for pancreatic ductal adenocarcinoma in European ancestry populations. The study population included prior genome-wide association studies (GWAS) conducted by the Pancreatic Cancer Cohort and Case-Control Consortia, PanScan I-III GWAS and PANC4 GWAS. The analysis was stratified by study design (case-control and cohort) due to differences in assessment of alcohol intake. There were 3,707 cases and 4,167 controls from the case-control and 1,199 cases and 4,813 controls from the cohort studies. Analyses were conducted using logistic regression with unconstrained maximum likelihood estimation, followed by fixed effect meta-analyses to combine results across GWAS phases. Results: In the case-control studies, a novel region of association on 10p11.22, lead SNP rs7898449 (odds ratio for the interaction $OR_{interaction}$=0.44, 95%CI 0.33-0.57, $P_{interaction} = 2.1x10^{-9}$) was identified. Among the non-heavy drinking participants, each copy of the minor allele A was associated with an increased pancreatic adenocarcinoma risk (OR = 1.16, 95%CI 1.05 - 1.29) compared to the major allele G. In contrast, among heavy drinkers, each copy of the minor allele A was associated with decreased pancreatic adenocarcinoma risk (OR = 0.51, 95%CI 0.37 -0.69) compared to the major allele. This region contains an eQTL for the *NRP1* gene. Of the 20 genomic regions with genome-wide significant evidence of association to pancreatic cancer in our prior studies, we observed that alcohol use potentially modifies the association between genomic variation near *LINC*00673, rs11655237 on 17q25.1 (P = 0.03) and *ABO*, rs505922 (P = 0.04) on 9p34.2 in the case-control studies. Due to the smaller size, there was limited power to detect association in the cohort subset. Conclusion:We identified a novel genomic region that may be associated with pancreatic adenocarcinoma risk in conjunction with heavy alcohol use. This region is located near an eQTL for the NRP1 mRNA, which plays an important role in pancreatic cancer progression. Future studies are needed to confirm the association and better understand how alcohol use may modify pancreatic adenocarcinoma risk.

Session Title: Cancer Poster Session II

PB5060 Genome-wide association study of transient myeloid leukemia in children with Down syndrome

Authors:

Y. Li[1], N. Elliott[2], P. Vyas[2], I. Roberts[2], A. de Smith[1]; [1]Univ. of Southern California, Los Angeles, CA, [2]NIHR Oxford BioMed. Ctr., Oxford, United Kingdom

Background: Children with Down syndrome (DS) have a remarkably high risk, ~150-fold, of developing myeloid leukemia (ML-DS). ML-DS is preceded by a transient preleukemia in newborns with DS that is characterized by increased blast cells and somatic mutations in the ChrX gene *GATA1* that result in a shorter protein isoform (*GATA1s*). In this study, we aimed to investigate the role of genetic variation in the development of *GATA1s* mutations in DS, with the central hypothesis: heritable variation influences the in utero generation of somatic *GATA1s* mutations in neonates with DS.
Methods: The study comprised 436 newborns with DS from the Oxford Down Syndrome Cohort Study, in which targeted deep sequencing of *GATA1* exons 2 and 3 had been previously performed. Whole genome sequencing (WGS) was performed at the Broad Institute and funded by the NIH Gabriella Miller Kids First and INCLUDE projects. Following quality control of WGS data, 425 DS newborns were included in a genome-wide association study (GWAS) of somatic *GATA1s* mutations. Association tests included *GATA1s* mutation as either a binary trait (105 *GATA1s*-positive cases, 320 *GATA1s*-wildtype controls), or continuous trait with variant allele frequencies (VAF) ranging from 0% to 89%. Biological sex, gestational age, and ten principal components were adjusted in the logistic model and the linear model with logit-transformed frequency, respectively. We also assessed genetic variation in the *GATA1* gene itself, on ChrX, in males and females separately.
Results: In the GWAS of *GATA1s* mutation VAF, we identified one genome-wide significant hit, a low frequency variant rs115118904 (p=3.3x10^{-8}, MAF=0.055), in a genomic region at chr2p16.3 previously associated with body mass index. The top signal comprising common variants was located at chr15q26.2, with lead SNP rs2584234 (p=1.5x10^{-7}; allele freq.=0.28) which has previously been associated with lung function. In the GWAS treating *GATA1s* as a binary trait, this SNP was associated with a 2.3-fold increased risk of developing a mutation. In targeted association testing of variants in *GATA1*, we found a distinctive result between males and females with several nominally significant SNPs only in males, including lead variant rs5906709 (p=0.021) that was previously associated with blood cell traits.
Discussion: We performed the first GWAS of transient leukemia in DS, identifying putative risk variants that may influence the development of *GATA1s* mutations in DS neonates, albeit these require validation. Ongoing analyses include examination of variants on the trisomic Chr21, and the pursuit of independent DS cohorts to increase our study sample size.

Session Title: Cancer Poster Session III

PB5061 † Genomic alterations in lung cancer among never-smokers: An ancestry and sex-stratified analysis.

Authors:

A. Schroeder, S. Chatterjee, J. Cadley, C. Falcon, M. Mehine, C. Bandlamudi, Y. Kemel, A. Chatterjee, M. Ladanyi, D. Solit, M. Berger, Z. Stadler, M. Kris, D. Jones, A. Schoenfeld, C. Rudin, K. Offit, S. Mukherjee; Mem. Sloan Kettering Cancer Ctr., New York, NY

Although most lung cancer cases are linked to cigarette smoking, an estimated 10-25% of lung cancer cases occur in never smokers. Lung cancer in never-smokers (LCINS) presents unique molecular characteristics compared to its smoker-induced counterparts. LCINS incidence, mortality, and outcomes vary widely across race/ethnicity. Our study aims to comprehensively assess genomic alterations in LCINS stratified by sex and ancestry to identify clinically actionable alterations and gain insight into targeted therapeutics. We analyzed 6,355 patients with lung cancer who underwent germline and somatic tumor sequencing for 505 genes by MSK IMPACT between 2015 and 2021. Smoking history data was derived using a unique self-reported smoking behavior questionnaire or from electronic health records using machine learning methods. Somatic mutation data was analyzed for recurrent driver mutations, structural alterations, and mutational signatures. In total, 46 genes with recurrent driver mutations observed in >20 cases were tested for sex and ancestry-specific differences. Pathogenicity assessment for germline variants in 90 cancer predisposition genes was performed using ACMG guidelines. In total, 1364 (21.5%) lung cancer patients in our cohort had no smoking history nor a smoking mutational signature. The majority of patients had lung adenocarcinoma (88%) and 70% were female. The prevalence of LCINS varied by ancestry with the highest prevalence in Asian (343/606; 57%), followed by African (75/230; 33%), Ashkenazi Jewish (270/1149; 23%), and European (575/3782;15%). Patients of African ancestry had a significantly earlier age at diagnosis compared to others (58 vs 63 years, p=<.001). Patients of Asian ancestry had higher rates of chromosomal instability as measured by fraction genome alteration (FGA) compared to others (mean FGA 0.21 vs 0.17; p=.005). As previously reported, *EGFR* mutations predominated in never smokers with the highest frequency observed in Asian patients (69.5%) followed by African ancestry patients (56%). In addition to *EGFR, RBM10* mutations were found to be enriched in patients of Asian ancestry compared to others (p<.001). Germline pathogenic variants in DNA damage repair pathway genes were observed in 7% of patients. The prevalence of germline pathogenic variants in *BRCA1* and *BRCA2* was significantly higher in never-smokers compared to ever-smokers (p-value <0.01). Our study suggests that the variation in somatic and germlin*e* mutation frequency is significantly distinct in different ancestry groups. We are currently exploring germline-somatic interactions to elucidate their impact on clinical management and outcomes.

Session Title: Cancer Poster Session I

PB5062 Genomic landscaping of Oral Cancer using Liquid Biopsy: search for an elusive target

Authors:

R. Rawal, D. Yadav; Gujarat Univ., Ahmedabad, India

Oral cancer is a prevalent and potentially life-threatening disease, necessitating the development of robust diagnostic and prognostic tools. Traditional tissue biopsies for molecular profiling have limitations in terms of invasiveness and limited sampling. The emergence of liquid biopsy, a non-invasive approach that analyzes biomarkers in body fluids , including blood, saliva, offers promising opportunities for comprehensive multi-omic profiling in oral cancer. The purpose of the present study was to use bioinformatics to discover concurrent gene signatures from tissue and liquid biopsy sample and define their potential mechanisms in OC.The Gene Expression Omnibus (GEO) database was used to acquire data for OSCC(involving either buccal mucosa or Tongue). Differentially expressed genes (DEGs) in the control and case groups were determined using GeneSpring analysis. Protein-protein interaction (PPI) network and functional enrichment were built for DEGs. Intrinsic disorder property of hub genes were evaluated using PONDR®. qPCR of four selected hub genes (STAT1, MMP9, CXCL8 and ISG15) was carried out in 100 OSCC tissues along with their adjacent normal tissue and exosomal transcripts derived from serum samples. The expression profile of tissue derived transcript correlated well with exosomal transcripts. These genetic alterations in oral cancer, providing insights into dysregulated signaling pathways in oral cancer pathogenesis. The integration of multi-omic data enabled the identification of novel molecular subtypes and the elucidation of complex molecular networks underlying oral cancer. The identified biomarkers hold promise for early detection, risk stratification, treatment selection, and monitoring of therapeutic response in oral cancer patients. In conclusion, the application of liquid biopsy-based genomic profiling in oral cancer has facilitated a deeper understanding of the molecular heterogeneity and complexity of the disease. This approach holds great potential for advancing precision medicine in oral cancer, ultimately leading to improved patient outcomes and personalized therapeutic strategies. Further research and validation studies are warranted to translate these findings into clinical practice.

Session Title: Cancer Poster Session II

PB5063 Genomic profiling of group of Egyptian retinoblastoma patients

Authors:

O. Eid[1], A. Mohamed[2], H. El-Bassyouni[3], I. Fadel[1], R. Mahrous[1], K. Refaat[1], S. Hussein[1], Retinoblastoma group at 57357; [1]Natl. Res. Ctr., Egypt, Giza, Egypt, [2]Natl. Res. Ctr., Cairo, Egypt, [3]Natl. Res. Ctr., Cairo, Cairo, Egypt

Background: Retinoblastoma is a malignant tumor of the retina. About 50% of patients are predisposed to retinoblastoma by a germline *RB1* pathogenic variant. However, retinoblastoma patients can have non-*RB1* germline pathogenic variant in other cancer-associated genes that promote tumorigenesis. Identification of such genetic alterations is essential and may guide the treatment strategy. **Aim:** The aim of the work was to profile germline "hot spot" genomic variants in group of Egyptian retinoblastoma patients. **Results:** Blood sample from forty three retinoblastoma patients were investigated by MLPA and MLPA negative cases by either aCGH or panel NGS for the most common genes involved in retinoblastoma. 20.9% showed *RB1* deletion, abnormal methylation status, or both by MLPA. Moreover, seven case were investigated by aCGH. aCGH analysis detected CNVs that implicated in carcinogenesis, such as CNVs involving 8p11.22, 14p11, 22q11.2. Furthermore, panel NGS analysis was performed using AmpliSeq Custom DNA panel utilizing a list of genes that are available at cancer genetics web. Pathogenic mutations in *RB1* were detected in 10 cases out of 23 cases 43.47%. 10 cases had likely pathogenic mutations in the *RB1*. While, no mutations were detected in 3 cases. **Conclusion:** Our results might contribute to deepen our understanding of the genetic predisposition of this malignancy as well as it might provide valuable tumor biomarker. Further functional validation of our findings is now required.

Session Title: Cancer Poster Session III

PB5064 Genomic, clinical, and pathologic features of *CHEK2* breast cancer

Authors:

T. Lim[1], L. Goldblatt[2], H. Desai[1], G. Kelly[1], R. Hausler[1], B. Wubbenhorst[3], J. Pluta[3], P. Bastian[3], K. D'Andrea[3], A. Doucette[4], P. Gabriel[5,4], Regeneron Genetics Center, Penn Medicine BioBank, A. Cervantes[6], C. Hu[7], P. Shah[1], A. R. Bradbury[1], A. Nayak[8], J. Weitzel[9], F. J. Couch[7], S. M. Domchek[1,10], K. L. Nathanson[3,11,4], K. N. Maxwell[1,11,4,10]; [1]Div. of Hematology/Oncology, Dept. of Med., Univ. of Pennsylvania Perelman Sch. of Med., Philadelphia, PA, [2]Dept. of Med., Univ. of Pennsylvania Perelman Sch. of Med., Philadelphia, PA, [3]Div. of Translational Med. and Human Genetics, Univ. of Pennsylvania Perelman Sch. of Med., Philadelphia, PA, [4]Abramson Cancer Ctr., Univ. of Pennsylvania Perelman Sch. of Med., Philadelphia, PA, [5]Dept. of Radiation Oncology, Univ. of Pennsylvania Perelman Sch. of Med., Philadelphia, PA, [6]Beckman Res. Inst. of City of Hope, Duarte, CA, [7]Dept. of Lab. Med. and Pathology, Mayo Clinic, Rochester, MN, [8]Dept. of Pathology and Lab. Med., Univ. of Pennsylvania Perelman Sch. of Med., Philadelphia, PA, [9]Univ. of Kansas Cancer Ctr., Kansas City, KS, [10]Corporal Michael Crescenz VA Med. Ctr., Philadelphia, PA, [11]Dept. of Genetics, Univ. of Pennsylvania Perelman Sch. of Med., Philadelphia, PA

Background: The DNA damage response gene Checkpoint Kinase 2 (*CHEK2*) has been established as a moderate penetrance gene in breast cancer (BC) development. *CHEK2*+ BC is typically estrogen receptor positive (ER+); however, studies have suggested worse outcomes for *CHEK2*+ BC patients. We thus sought to understand genomic and pathological features of *CHEK2*+ compared to germline mutation negative BCs. **Patients and Methods:** We determined the association of likely pathogenic/pathogenic (LP/P) and low-risk *CHEK2* variants with BC in the Penn Medicine BioBank (PMBB). We then compared clinicopathological characteristics of *CHEK2*+ BC to BC from PMBB patients with no cancer risk gene variants upon whole exome sequencing (WES). Finally, we conducted genomic analyses using WES of breast tumor/normal pairs from 60 *CHEK2*+ patients. **Results:** Among the 22,056 women who underwent WES, a significant but moderate association between LP/P *CHEK2* variants and BC was found (OR: 2.1, 95% CI 1.2-3.7, p = 0.01), while that for low-risk missense *CHEK2* variants and BC was not significant. *CHEK2*+ BCs (n = 99 BC/82 people) were majority ER+ (78%), low/intermediate grade (54%), and diagnosed as Stage II and under (n = 87%). Compared to 1013 PMBB mutation negative BC patients, the *CHEK2*+ cohort had a significantly younger age of diagnosis (median 45 vs 55, p < 0.01) but no differences in invasive hormone receptor profile, histology, grade, tumor stage at diagnosis, distant metastasis rates, and deaths. In the WES cohort, 62% of patients (37/60) had the *c.1100delC* variant, 11 had other LP/P variants, and 12 had low-risk missense variants. Allele-specific loss of heterozygosity was found in 56% of all LP/P and 42% of low-risk missense carriers (p = 0.52), with 19% and 17% each losing the mutant allele, respectively. *TP53* variants were found in 31% of *CHEK2* LP/P carriers, similar to ER+ BC (23%) but lower than TNBC (86%) and germline *BRCA1/2* carriers (67%) from The Cancer Genome Atlas (TCGA). All LP/P, *c.1100delC* only, and low-risk missense *CHEK2* subsets had similar homologous recombination deficiency (HRD) scores, tumor mutational burden, microsatellite instability, and *BRCA* signature (Signature 3). LP/P *CHEK2* HRD scores were lower (mean: 24, 95% CI: 21-28), while Signature 3 (mean: 0.05, 95% CI: 0.03-0.07) was similar compared to the mutation negative ER+, HER2-, or TNBC groups (n = 639) in TCGA. **Conclusions:** These results indicate that *CHEK2*+ BC is similar clinically and genomically to sporadic ER+ BC without evidence to suggest PARP inhibition response. Genomic and pathological features do not suggest an etiology for increased aggressiveness of *CHEK2*+ BC, consistent with outcomes in our clinical cohort.

Session Title: Cancer Poster Session I

PB5065 Germline DNA repair pathway deficiency and clonal TP53/KRAS alterations in pancreatic cancer driving target therapy.

Authors:

S. Innamorato[1,2], E. Pasquinelli[1,2], V. Serio[1,2], L. Adamo[1,2], M. Lista[1,2], L. Loberti[1,2,3], E. Antolini[1,2], G. Castello[1,2], J. Covarelli[1,2], D. Maffeo[1,2], C. Fallerini[1,2], P. Ilardi[1,2,3], M. Mencarelli[3], G. rollo[1,2], S. Minetto[1,2], M. Baldassarri[1,2,3], A. Carrer[1,2,3], A. Pinto[3], C. Lo Rizzo[3], F. Mari[1,2,3], R. Petrioli[4], I. Martellucci[4], S. Marsili[4], S. Miano[4], A. Barone[5], C. Milandri[6], A. Martignetti[7], A. Renieri[1,2,3], F. Ariani[1,2,3]; [1]Med. Genetics, Univ. of Siena, Siena, Italy, [2]Med Biotech Hub and Competence Ctr., Dept. of Med. Biotechnologies, Univ. of Siena, Siena, Italy, [3]Genetica Medica, Azienda Ospedaliera Univ.ria Senese, Siena, Italy, [4]Oncologia Medica, Azienda Ospedaliera Univ.ria Senese, Siena, Italy, [5]Anatomia Patologica, Azienda Ospedaliera Univ.ria Senese, Siena, Italy, [6]Oncologia-Arezzo, Azienda Usl Toscana sud est, Arezzo, Italy, [7]Oncologia - Ospedale di Campostaggia, Azienda Usl Toscana sud est, Siena, Italy

Pancreatic cancer remains one of the most lethal malignancies, with a dismal prognosis and mortality/incidence ratio of 94%. In a fraction of patients BRACA germline variants have been identified but PARP inhibitor clinical trials failed in these patients. Here we want to address the full spectrum of germline variants and interpret the results within the framework of clonal mutations. A cohort of 87 pancreatic cancers who underwent accurate pedigree reconstruction and genetic counselling were enrolled. A virtual panel of 230 cancer driver genes were investigated by exome analysis for germline variant identification, while Next Generation Sequencing (NGS) cell-free DNA (cfDNA) analysis focused on 77 clonal cancer drivers. Germline variants in DNA repair genes (DRGs) were identified in 53% of cases. The homologous recombination (HR) genes (*BRCA1/2*, *PALB2*, *CHEK2*, *ATM*, *SLX4*, *RECQL4*) were found to be the most frequently mutated (46%), followed by base Excision repair (BER) genes (*MUTYH*, *NTHL1*) (14%), mismatch repair (MMR) genes (*MSH2*, *MSH3*, *MSH6*) (11%) and Fanconi anemia (FA) genes (*FAN1*, *SLX4*, *FANCM*) (11%), which were always paired with a second germline mutation in another DRGs. Clinical worsening was marked by paired *TP53* and *KRAS* variants (6 of 19 cases tested). Our results indicated that while a fraction of patients remain target elusive (35%), another fraction (11%) could be treated by immunotherapy with checkpoint inhibitors, and a larger fraction of cases (57%) can be potentially treated by PARP inhibitors: all those within HR or FA deficiency. However, clonal assessment needs to be evaluated as well, as recently KRAS-tumors became druggable by covalent and non-covalent inhibitors. Not addressing clonal evolution could be the primary reason for PARP inhibitor failure in BRACA patients and combo therapy could be envisaged after clonal characterisation by NGS.

Session Title: Cancer Poster Session II

PB5066 Germline genetic analysis of Lynch syndrome using 110K samples from patients with 23 cancer types and controls

Authors:

K. Mizukami[1], U. Yoshiaki[1], M. Endo[1], Y. Iwasaki[1], Y. Kamatani[2], M. Hirata[3,2], K. Shiraisihi[4], T. Kohno[4], I. Imoto[5], K. Matsuda[2], H. Nakagawa[1], Y. Murakami[2], Y. Momozawa[1]; [1]RIKEN, Yokohama, Japan, [2]The Univ. of Tokyo, Tokyo, Japan, [3]Natl. Cancer Ctr. Hosp., Tokyo, Japan, [4]Natl. Cancer Ctr. Res. Inst., Tokyo, Japan, [5]Aichi Cancer Ctr., Nagoya, Japan

Lynch syndrome is a rare familial cancer syndrome caused by pathogenic variants (PV) in the mismatch repair genes: *MLH1*, *MSH2*, *MSH6*, or *PMS2*. Although PVs in these genes are known to be associated with various types of cancers, the current estimated risk of developing the cancers was based mainly on studies with potential selection bias. In this study, we conducted a large-scale case-control analysis to evaluate the risk of developing cancers in individuals with PVs in the four genes and identify their clinical and demographic characteristics. Using targeted sequencing, we analyzed the coding regions of the four genes in over 75,000 patients with 23 cancer types and over 35,000 individuals without cancers in the same population, both of which were sourced from a multi-institutional hospital-based registry, BioBank Japan. Variants predicted to loss of function based on changes in amino acid sequence or registered as pathogenic and likely pathogenic in ClinVar were collectively referred to as PV. We compared the PV frequencies between cases and controls and documented the demographic and clinical characteristics of patients with PVs. In this study, we defined associated genes if they had a high or moderate risk ($OR \geq 1.5$) of development ($p < .05$). We identified a total of 229 PVs across the four genes. The highest carrier frequency of PVs in one of the four genes was observed in endometrial cancer (5.0%), followed by colorectal cancer (1.0%), among cancers with over 100 patients. PVs were associated with the development of several cancers, including colorectal and endometrial cancers. The type of associated cancer and the risk varied depending on the four genes. PVs in *MSH2* had the broadest spectrum of associated cancer types: endometrial (OR 53.6), bladder (26.6), colorectal (16.2), gastric (5.8), and prostate (5.3) cancers. The frequency of patients with PVs was higher in patients with multiple associated cancers. Compared to patients without PVs, the diagnosed ages were significantly younger in patients with PVs in terms of colorectal (*MLH1*, *MSH2*, and *MSH6*), endometrial (*MLH1* and *MSH2*), gastric (*MLH1*), and prostate (*MSH2*) cancers. Family histories of endometrial and colorectal cancers were more frequently observed not only among the same cancer patients with PVs but also among other associated cancer patients with PVs. In conclusion, this study precisely defined the risk of developing cancers caused by PVs in the mismatch repair genes and characterized patients with PVs.

Session Title: Cancer Poster Session III

PB5067 Germline mutations in *RAD51C* and *RAD51D* genes observed in cancer patients in Singapore

Authors:

S-T. Li[1], N. Ishak[1], S. Chan[1], S. Chong[1], Z. Zhang[2], J. Chiang[2], J. Ngeow[2,3]; [1]Natl. Cancer Ctr. Singapore, Singapore, None, Singapore, [2]Natl. Cancer Ctr. Singapore, Singapore, Singapore, [3]Lee Kong Chian Sch. of Med., Nanyang Technological Univ., Singapore, Singapore

Background: *RAD51C* and *RAD51D* play crucial roles in homologous recombination-mediated DNA repair. Germline pathogenic variants in *RAD51C* and *RAD51D* genes have been known to associate with increased risk of developing breast and ovarian cancer. However, the prevalence and clinical significance of germline mutations in *RAD51C* and *RAD51D* among cancer patients in Singapore has not been fully explored. **Methods:** We performed a review on cancer patients who were seen at the Cancer Genetics Service, National Cancer Centre Singapore, from 2015 to May 2023. Patients who underwent germline genetic testing using multi-gene panel including *RAD51C and RAD51D* were reviewed. Patient demographics, clinical characteristics as well as their family history of cancers, were collected and analysed. **Results:** A total of 3121 cancer patients, representing various cancer types, who underwent germline genetic testing were included in our analysis. Germline pathogenic variants (PV) in *RAD51C* and *RAD51D* were identified in 19 (0.6%) and 17 (0.5%) unrelated cancer patients, respectively. The prevalence of germline PVs in *RAD51C* and *RAD51D* genes was 1.9% and 1.2% among 642 ovarian cancer patients, respectively, and 0.6% and 0.4% among 1746 breast cancer patients, respectively. The median age at ovarian diagnosis was 57.5 years for patients with *RAD51C* PVs and 48 years for those with RAD*51D* PVs. Similarly, the median age at breast cancer diagnosis was 49 years for patients with *RAD51C* PVs and 46 years for those with *RAD51D* PVs. Majority of patients harbouring germline PVs in *RAD51C* and *RAD51D* genes presented with a family history of cancer. Specifically, 57.9% of patients with *RAD51C* PVs and 41.2% of patients with *RAD51D* PVs had a family history of breast and/or ovarian cancer. Several recurrent PVs were identified in patients of Chinese ethnicity, including c.394dup (p.Thr132Asnfs*23) (5/19), c.905-2A>C (5/19) and c.571+5G>A (4/19) in *RAD51C* genes and c.270_271dup (p.Lys91Ilefs*13) (14/17) in *RAD51D* gene. **Conclusion:** Our findings provide insights into the prevalence, cancer spectrum and recurrent variants in *RAD51C* and *RAD51D* genes among cancer patients in Singapore.

Session Title: Cancer Poster Session I

PB5068 Germline Panel As Standard Method for Investigating Hereditary Breast and Ovarian Cancer Syndrome in Brazilian Supplementary Health System

Authors:

J. FARIA[1], L. Marco[2], L. bastos-rodrigues[2], J. Assumpção[3], F. Soardi[4], L. Matos[2], G. Cunha[5], A. Ribas[6]; [1]UNIVERSIDADE FEDERAL DE MINAS GERAIS, BELO HORIZONTE, Brazil, [2]Univ.e Federal de Minas Gerais, Belo Horizonte, Brazil, [3]Geneticenter Centro de Genética, BELO HORIZONTE, Brazil, [4]Laboratório Lustosa, BELO HORIZONTE, Brazil, [5]Cetus Oncologia, BELO HORIZONTE, Brazil, [6]Hosp. Felício Rocho, BELO HORIZONTE, Brazil

The Brazilian private health system currently follows the medical literature in which it estimates the prevalence of 10% of pathogenic and/or probably pathogenic variants in women with breast cancer, recommending germline testing in stages through New Generation Sequencing of Genes and Multiplex Ligation-dependent Probe Amplification of the BRCA1 and BRCA2 genes and, in case of inconclusive results, performing a germ panel. Recent studies published worldwide suggest an approximate prevalence of between 10 and 20% of pathogenic and probably germline pathogenic variants in individuals with cancer. Two recent Brazilian studies showed an approximate prevalence of 18% of pathogenic and probably pathogenic variants in individuals at risk of hereditary cancer. The present study evaluated 701 individuals between 2021 and 2022 with Hereditary Breast and Ovary Cancer Syndrome criteria and performed a single test through germ panel, identifying a prevalence of 19.74% of pathogenic and probably pathogenic variants. The results suggest that the standard use of the germ panel results in a 90% gain in the sensitivity of identifying pathogenic and probably pathogenic germline variants and a 67% reduction in the cost and time of investigation.

Session Title: Cancer Poster Session II

PB5069 Germline Testing for Lynch Syndrome:Are we there yet?

Authors:

S. Agnihotri, A. A. Mehta, S. Nathany, H. H. Diwan, D. D. Kumar; Rajiv Gandhi Cancer Inst. And research centre, New Delhi, India

Introduction: Of the genetic cancer risk syndromes, Lynch syndrome is the most prevalent, accounting for 3-5% of all colorectal cancers (CRC) and endometrial malignancies (EC). Finding these individuals facilitates family screening, future surveillance planning, therapeutic optimization, and prognosis prediction. The most practical, economical, and effective screening technique is now being used to screen patients with MSI or MMR deficiencies for other genetic abnormalities. **Objective: T**o assess the MMR/MSI prevalence across all tumor types and the take-up of Lynch syndrome germline testing in India. **Methods:** The adoption of germline Lynch syndrome testing from 2021 to 2015, as well as the MMR and MSI status across all malignancies, were evaluated in this single-center cross-sectional investigation. The state of mismatch repair was assessed by analyzing the somatic tumor using either immunohistochemistry (IHC) for MLH1, PMS2, MSH2 & MSH6 (on Ventana platform) or MSI by fragment length analysis with a Promega kit (SeqStudio/Applied Biosystems).
Peripheral blood was used for the germline test, which looked for single nucleotide variants and indels in the four MMR genes (MLH1, PMS2, MSH2, and MSH6) using a tailored DNA panel and Next-generation sequencing (Ion Torrent).**Results:** Out of 1447 cases showed, a total of 288 (19.9%) showed MMR deficiency (dMMR)/MSI.In 77 of 288 patients (26.7%), a germline lynch test was performed. MLH1 gene mutations were found in 51 of 77 (66.2%) cases analyzed, with a frequency of 60.8%. 60 of 288 cases (20.8%) had insurance coverage but did not undertake germline testing (p=0.0005). A germline mutation was found in 42/61 CRCs tested (68.9%), with the MLH1 gene being most frequently implicated (66.7%) (p 0.0001). **Discussion:** Guidelines recommend germline testing after assessing all CRCs and, more recently, all ECs f r dMMR/MSI to identify Lynch syndrome patients better. When compared to an Australian study that discovered a referral rate of 11% in dMMR CRCs, the current study's referral rate for genetic testing for Lynch syndrome was 26.7%.2In the US, non-Hispanic whites had the highest referral rate (21.2%), while the Hispanic patients 10.9%. **Conclusion:** Genetic testing aids in the early detection of recurrence and the screening of relatives. According to our observations, socioeconomic limitations contribute to India's limited penetration of germline testing. By providing proper counseling, doctors and molecular pathologists can address this unmet need and assist in eliminating the accompanying societal stigma.

Session Title: Cancer Poster Session III

PB5070 Germline whole genome sequencing of children with Down syndrome identifies novel variants associated with risk of acute lymphoblastic leukemia.

Authors:

M. Richard[1], Y. Yu[2], J. J. Junco[1], C-J. Hsu[1], A. J. de Smith[3], T. C. Rosser[4], E. J. Leslie[4], M. Devidas[5], E. A. Raetz[6], M. L. Loh[7], S. P. Hunger[8], J. M. Espinosa[9], J. J. Yang[5], S. L. Sherman[4], C. D. Huff[2], K. Rabin[1], P. J. Lupo[1]; [1]Baylor Coll. of Med., Houston, TX, [2]UT MD Anderson Cancer Ctr., Houston, TX, [3]Univ. of Southern California, San Gabriel, CA, [4]Emory Univ., Atlanta, GA, [5]St Jude Children's Res. Hosp., Memphis, TN, [6]New York Univ., New York, NY, [7]Seattle Children's Hosp., the Ben Town Ctr. for Childhood Cancer Res., Univ. of Washington, Seattle, WA, [8]Children's Hosp. of Philadelphia, Philadelphia, PA, [9]Linda Crnic Inst. for Down Syndrome, Univ. of Colorado Anschutz Med. Campus, Aurora, CO

Background: Children with Down syndrome (DS) have a 10-fold greater risk for acute lymphoblastic leukemia (ALL), primarily B cell precursor ALL, than children without DS. We previously reported ALL susceptibility variants identified by array-based genotyping among children with DS. Here, we leverage whole genome sequencing (WGS) to expand identification of common variants associated with the excess risk of ALL in DS. **Methods**: We performed Illumina 30X germline WGS of 379 individuals with DS-ALL enrolled on Children's Oncology Group trials and 1504 individuals without ALL from DS research cohorts. Variants were detected following Gabriella Miller Kids First best practices, with quality control and ancestry estimation. We conducted autosomal, non-chr21 genome-wide association studies (GWAS) adjusted for sex and top principal components using 1) PLINK for ancestry-specific GWAS followed by trans-ancestry meta-analysis, and 2) GENESIS mixed models for pooled GWAS including all individuals. **Results**: In meta-analysis and pooled analysis, we identified known and novel loci significantly associated with DS-ALL (P<5x10^{-8}). Variants in the *IKZF1* locus were the top known association with DS-ALL, yet each analysis identified different top variants*:* rs58923657 in European ancestry individuals (odds ratio [OR]: 2.08, P =1.1x10^{-10}), rs17133807 in trans-ancestry meta-analysis (OR=1.90, P=7.3x10^{-12}), and rs28462675 in pooled analysis (OR=1.11, P=1.1x10^{-13}). These variants map to similar enhancer regions and transcription factor binding sites (including *PAX5*, *IKZF2*) in ENCODE B-lymphoblastoid cell lines. We validated the *CDKN2A* association (rs3731249, OR=1.23, P=1.5x10^{-9}) and report a novel locus on chr4q13.1 (rs17290452, OR=1.17, P=2.4x10^{-8}), a region also previously associated with susceptibility to *ETV6::RUNX1*+ B-ALL. Our pooled analysis identified multiple novel loci, including 22 significant variants mapping to *PCBP2* (top variant rs1361910621, OR=2.24, P=2.1x10^{-16}), a poly(rC)-binding protein that may function as an oncogenic splicing factor. This *PCBP2* region has strong support for regulatory potential in B cells and has recently been characterized as a novel tumor suppressor in hematologic malignancies. **Conclusions**: Leveraging WGS data on individuals with DS, we validated enhanced genetic penetrance at known ALL predisposition loci among individuals with DS-ALL and identified a novel locus at *PCBP2*. Expansion of these findings is underway by integrating additional germline samples, identifying trisomic variants on chromosome 21, and performing functional validation of identified loci.

Session Title: Cancer Poster Session I

PB5071 Germline-phased de novo assembly of K562 cancer genome from HiFi and HiC data imparts complete cancer genomic rearrangements at single-nucleotide resolution.

Authors:

Y. Guo[1,2], H. Cheng[1,2], H. Li[1,2], C-Z. Zhang[1]; [1]Dana Farber Cancer Inst., Boston, MA, [2]Harvard Med. Sch., Boston, MA

Long-read sequencing has enabled the first complete assembly of a human genome and is rapidly transforming the analysis of human genetic variation. *De novo* long-read assembly is especially powerful for resolving complex structural variation and large segmental duplications, both of which occur frequently as somatic alterations in cancer genomes. However, the presence of whole-chromosome and large segmental duplications in aneuploid cancer genomes poses critical problems for the current assembly algorithms that are designed for haploid or diploid genome assembly. Due to this technical challenge, both the utility and the feasibility of *de novo* cancer genome assembly are unknown.

Here we took these cancer genome characteristics into account and presented the *de novo* assembly of the K-562 genome that preserves the germline phasing, allelic copy number and chromosome structures of this widely used cancer cell line. The assembly was constructed from PacBio HiFi long reads and Hi-C short reads. It was based on the phased assembly generated by a diploid genome assembler (hifiasm) and further refined by haplotype-specific Hi-C contacts and DNA copy number. We resolved previously unknown translocations, even with complex ones such as foldback. We discussed technical challenges due to large DNA deletions (loss-of-heterozygosity) and duplications in aneuploid genomes and described strategies to overcome these challenges in *de novo* assembly. The assembly of the K-562 genome provides a more complete and accurate reference than ever before for the interpretation and correlation of functional genomic and epigenomic data of this widely used cell line. This study further establishes a framework for analyzing cancer genomes by long-read sequencing and *de novo* assembly.

Session Title: Cancer Poster Session II

PB5072 Haplotype-resolved multi-modal analysis of cancer genomes using Oxford Nanopore sequencing.

Authors:

P. Rescheneder[1], S. Mckenzie[2], S. Aganezov[3], P. James[4], S. Juul[5]; [1]Oxford Nanopore Technologies, Oxford, United Kingdom, [2]Oxford Nanopore Technologies, Spokane, WA, [3]Oxford Nanopore Technologies, Inc., Reading, MA, [4]Oxford Nanopore, Oxford, United Kingdom, [5]Oxford Nanopore Technologies, New York, NY

Cancer is a complex and dynamic disease driven by somatic genomic and epigenomic alterations that accumulate over time. These changes give rise to heterogeneous collections of cells or clones, each with distinct (epi)genomic profiles within a single tumour. Accurate characterisation of these changes is crucial for understanding the mechanisms driving the disease, identifying potential therapeutic targets, and personalising treatment strategies. Due to the technological constraints of short-read and array-based approaches, cancer research has historically had a strong focus on detecting small genomic changes like SNPs and small indels (SNVs) as well as the broad characterisation of large-scale copy number changes (CNVs), mostly ignoring other important variant classes and epigenetic modifications. Here we demonstrate how Oxford Nanopore native long-read sequencing enables the direct detection of not only SNVs, but also break point-resolved simple and complex structural variants (SVs and CNVs), identification of DNA modifications like 5-methylcytosine (5mC) and 5-hydroxy-methylcytosine (5hmC), and haplotype phasing of all variant types from a single tumour-normal dataset. We use well characterised cancer cell lines and in-silico benchmarking datasets to assess somatic SNV, and SV calling performance using different sequencing depths, and also demonstrate how Nanopore long reads enable haplotype and clone specific CNV calling. Finally, we use twelve matched tumour-normal pairs from four different tissues to showcase a comprehensive tumour-normal analysis using Nanopore sequencing. This includes the characterisation of complex patterns of somatic SVs in the different cancer samples, the identification of 5mC methylation patterns (e.g. in promotor regions) that are unique to certain tissues, exploring characteristic differences in 5hmC levels between tumour and normal samples as well as the detection of microsatellite instability and other hall marks of cancer.

Session Title: Cancer Poster Session III

PB5073 Hematopoietic cell transplantation for inborn errors of immunity in the context of Lynch syndrome

Authors:

E. Damskey[1], M. Similuk[1], R. Gore Moses[1], J. Yan[1], S. Byers[1], B. Beers[1], H. Orellana[1], M. Kamen[1], M. Setzer[1], R. Ghosh[1], G. Uzel[1], D. Dimitrova[2], J. Kanakry[2], D. Hickstein[2], L. Notarangelo[1], S. Holland[1], M. Walkiewicz[1]; [1]NIH/NIAID, Bethesda, MD, [2]NIH/NCI, Bethesda, MD

Background Allogeneic hematopoietic cell transplantation (HCT) can potentially provide long-term curative treatment for patients with severe immune disorders. Late effects from HCT include increased risk of secondary solid tumors due to genotoxic stress from HCT. Concurrently, relatively common cancer predispositions such as Lynch syndrome lead to an increased cancer risk due to defective DNA repair mechanisms. The extent to which familial malignancy risk may be exacerbated by HCT is largely unknown. We seek to add to this sparse literature by reporting on three cases of HCT recipients with both an immune disorder and Lynch syndrome.
Methods We performed exome and/or genome sequencing on study participants with known or suspected inborn errors of immunity from 2017-2023. We analyzed the data for primary findings related to clinical presentation as well as secondary findings. Cases Patient 1: 40yo Non-Hispanic Caucasian female presented with myelodysplastic syndrome. Exome sequencing identified a previously known heterozygous pathogenic *GATA2* variant (p.Arg398Trp), as well as a heterozygous pathogenic *MSH2* variant (p.Gln337LysfsTer20). The patient underwent HCT with a HLA-matched unrelated donor and is >3 years post-HCT without malignancy detected by high-risk Lynch syndrome screening. Patient 2: 35yo Hispanic Caucasian female with a history of granulomatous lymphoproliferation, sinopulmonary infections, and B-cell non-Hodgkin lymphoma, followed by T-cell lymphoma 9 years later, prompting HCT using an HLA-matched unrelated donor and reduced intensity conditioning (RIC). She died 339 days post-transplant due to progressive lymphoma despite successful engraftment. Exome analysis identified two mosaic *TET2* variants [pathogenic, p.Ser280PhefsTer14; and variant of uncertain significance (VUS), p.His1881Tyr], associated with a lymphoproliferative immunodeficiency. Additionally, a likely pathogenic heterozygous *PMS2* variant (p.Ser46Ile) was identified. Patient 3: 5yo Non-Hispanic Caucasian female with a history of immune thrombocytopenia, neutropenia, and alopecia. Genome sequencing identified a homozygous VUS in *RASGRP1* (p.Leu413Pro), which is consistent with her clinical presentation, as well as a heterozygous pathogenic variant in *MLH1* (p.Leu404ValfsTer12). The patient is currently preparing for RIC HCT. Conclusion These three cases add to the sparse literature on allogeneic HCT in the setting of Lynch syndrome. Longitudinal studies in larger cohorts would clarify risks. Identifying cancer predisposition in a critical minority of HCT patients may inform donor selection, conditioning, and post-HCT risk management.

Session Title: Cancer Poster Session I

PB5074 High incidence of germline mutations in DNA repair genes in bladder cancer

Authors:

L. Adamo[1,2], S. Minetto[1,2], M. Lista[1,2], D. Maffeo[1,2], E. Pasquinelli[1,2], E. Antolini[1,2], J. Covarelli[1,2], G. Castello[1,2], G. Casamassima[1,2], G. rollo[1,2], L. Loberti[1,2,3], A. Carrer[1,2,3], P. Ilardi[1,2,3], F. Mari[1,2,3], M. Mencarelli[3], A. Pinto[3], R. Tita[3], M. Baldassarri[1,2,3], I. Martellucci[4], S. Miano[4], A. Martignetti[5], C. Casi[5], A. Falsetta[5], F. Marrocolo[6], R. Di Rocco[7], M. Bruttini[1,2,3], F. Ariani[1,2,3], M. Lazzeri[8,9], R. Hurle[8,9], A. Renieri[1,2,3]; [1]Med. Genetics, Univ. of Siena, Siena, Italy, [2]Med Biotech Hub and Competence Ctr., Dept. Of Med. Biotechnologies, Univ. Of Siena, Siena, Italy, [3]Med. Genetics, Azienda Ospedaliera Univ.ria Senese, Siena, Italy, [4]Oncology, Azienda Ospedaliera Univ.ria Senese, Siena, Italy, [5]Dipartimento Oncologico USL Sud-Est Toscana-Area Senese, 53036, Poggibonsi, Italy, [6]Dept. of Oncology, San Donato Hosp., Azienda USL Toscana Sud-Est, via P. Nenni, 20, 52100, Arezzo, Italy, [7]Oncology, Ospedale La Misericordia, Grosseto, Italy, [8]Dept. of Urology, IRCCS Humanitas Res. Hosp., Rozzano, Italy, [9]Dept. of BioMed. Sci., Humanitas Univ., Pieve Emanuele, Italy

Bladder cancer is the 10th most commonly diagnosed cancer worldwide. Environmental risk factors such as tobacco smoking and occupational exposure to aromatic amines are well known. A not negligible percentage of superficial bladder cancers recur and show resistance to intravesical treatments (IMC or BCG) evolving into a muscle-invasive disease associated with high mortality. Radical cystectomy with or without neoadjuvant cisplatin-based chemotherapy or immunotherapy is the standard of care but morbidity and mortality remain high. Therefore, there is an unmet clinical need for improving the rate of care and quality of life. Markers that can indicate the rate of response and guide chemo- and/or surgical therapy are missing. Thus, the combination of tumor micro-environment analysis and genetic profiling could lead to new risk stratification for those patients and guide personalized treatment. Genetic predisposition has been thought to be less relevant than in other types of cancer due to little impact from family history and lack of clinically significant genetic changes. Therefore, genetic testing is not currently recommended. We wanted to explore the impact of germline predisposition in a cohort of 21 patients with bladder cancer at different stages. A virtual panel of 230 cancer driver genes from exome sequencing was used. A pathogenic germline variant was identified in 33% of patients, while uncertain significance variants were identified in 33% of patients. DNA repair genes were the most affected genes and included the Homologous recombination genes (*BRCA1/2, ATM, RECQL4*), followed by mismatch repair genes (*MLH1, PMS2, MSH3*). Our results challenge the current recommendation on genetic testing for bladder cancer. Furthermore, they suggest the possible use of PARPi or could predict response to immunotherapy with checkpoint inhibitors. The prognosis of patients with bladder cancer who are unexposed to environmental risk factors is generally poor. Bladder cancer tends to present earlier, be more aggressive and show higher resistance to chemotherapy. Thus, the knowledge of genetic profiling in this subset of patients may represent an improvement in the management of bladder cancer offering specific treatments and exploring new drugs such as PARPi.

Session Title: Cancer Poster Session II

PB5075 Higher breast cancer polygenic risk scores (PRS) are associated with less aggressive tumors and longer survival.

Authors:

D. Kurant, S. Groha, A. Gusev; Dana-Farber Cancer Inst., Boston, MA

Background: Identifying patients who are most likely to respond to treatments is a fundamental goal of precision oncology. Germline Polygenic Risk Scores (PRS) are an estimate of an individual's genetic risk for a trait/disease, but the influence of PRS on clinical outcomes in patients with cancer is largely unstudied. Here we quantified the association between germline PRS and somatic features in 2,058 patients with breast cancer treated at a tertiary cancer center.

Methods: Patients were biopsied as part of routine clinical care, sequenced on a targeted panel of 400-500 known cancer genes, and had common germline polymorphisms imputed based on off-target reads. Over 200 PRS from public GWAS were evaluated for association with survival across multiple anchor points. Patients were restricted to European ancestry and a PRS was computed for each patient for common traits including cancer risk and non-cancer phenotypes. The association between PRS and survival was assessed by a Cox proportional hazards model with sex, age, ancestry, and technical factors as covariates. Survival time was evaluated from diagnosis, biopsy, and first palliative intent therapy. All participants provided informed consent for research.

Results: An ER-positive breast cancer PRS was significantly associated with longer survival with a Hazard Ratio (HR) of 0.89 per standard deviation of PRS ($p = 1.50x10^{-4}$, FDR 0.03). 4 additional (partially correlated) breast cancer-related PRS were nominally significant and exhibited similar HRs. While the strongest association was seen from the time of biopsy, nominally significant ($p<0.01$) associations between the PRS and survival were also observed starting from diagnosis date or the date of first palliative-intent therapy.

We hypothesized that increased germline risk may lead to less aggressive tumors in patients who develop breast cancer and evaluated associations with somatic features. Higher breast cancer PRS was significantly associated with fewer TP53 mutations (OR=0.85, $p=8.7x10^{-5}$). The breast cancer PRS was also nominally associated with subclonal TP53 mutations after restricting to mutation carriers (OR = 1.22, p=0.032). The PRS association with survival remained nominally significant when including TP53 status as a covariate (HR = 0.93, p = 8.85x10-3). Statistical power was insufficient for an interaction analysis.

Conclusions: We demonstrate that higher germline risk correlates with longer survival in a large cohort of breast cancer patients. We hypothesize that higher germline risk may result in the development of less aggressive tumors and lead to better clinical outcomes.

Session Title: Cancer Poster Session III

PB5076 High-throughput DNA methylation profiling in the ovarian cancer susceptibility genes promoters: assessment across a wide spectrum of blood- and tissue-derived samples from patients and healthy individuals.

Authors:

M. Suszynska[1], A. Mostowska[2], J. Dabrowska[2], M. Ratajska[3,4], P. P. Jagodzinski[2], P. Kozlowski[1]; [1]Inst. of Bioorganic Chemistry Polish Academy of Sci., Poznan, Poland, [2]Poznan Univ. of Med. Sci., Poznan, Poland, [3]Med. Univ. of Gdansk, Gdansk, Poland, [4]Dunedin Sch. of Med. Univ. of Otago, Dunedin, New Zealand

Hypermethylation of promoter CpG islands is a generally recognized epigenetic mechanism responsible for gene silencing, constituting, apart from mutations, an alternative mechanism of gene inactivation in cancers. Such epimutations (EM) are somatic events in tumors; however, constitutional EM may occur in some normal tissues (e.g., blood), usually as a secondary event (i.e., with an associated genetic variant) or in a mosaic state, contributing to increased cancer risk.

To investigate the incidence of EM in ovarian cancer (OC) susceptibility genes, we developed a deep bisulfite sequencing assay targeting the promoter regions of 8 OC high- to moderate-risk genes (*BRCA1*, *BRCA2*, *BRIP1*, *RAD51C*, *RAD51D*, *ATM*, *NBN*, and *CHEK2*) and the *HOXA11* gene, which hypermethylation has already been identified as a potential biomarker in several cancers. The amplicon library was prepared from blood-derived DNA samples of OC patients (n=~100) and sex- and age-matched controls (CTRL; n=~100) using a two-step PCR process, and the library was then profiled by next-generation sequencing. Over 1000 individual DNA molecules per region and sample were analyzed to precisely determine the DNA methylation level. To better characterize the methylation profile of selected genes, we extended the analysis with external datasets generated with CpG genotyping arrays deposited at the GEO, UCSC Xena, and GTEx portals, collecting in total an additional ~1800 blood- and tumor/tissue-derived DNA samples from both OC patients and healthy CTRL. Promoter methylation levels of analyzed genes were minimal (except for *HOXA11*) in both blood- and tissue-derived samples. Comparing the mean DNA methylation between OC patients and CTRL revealed no relevant differences in blood-derived samples; however, a significant difference was observed for *HOXA11* in tumor-/tissue-derived samples. In addition, numerous EM with >20% methylation level in *BRCA1* and *RAD51C* was observed in tumors compared to normal tissue. Also, examples of EM with methylation level <20% occurred in *BRIP1* and *RAD51D*; however, these differences may not be representative due to the smaller CTRL size group. In summary, our data improve the understanding of the role of epigenetic silencing in OC, pointing to the particular importance of *HOXA11*, *BRCA1*, and *RAD51C* somatic EM in oncogenesis, while indicating that constitutional EM in OC susceptibility genes are very rare. In addition, the developed assay represents a highly sensitive, cost-effective, and high-throughput tool for precise multisample, multigene DNA methylation analysis, which could become the gold standard in OC cancer research and diagnostics.

Session Title: Cancer Poster Session I

PB5077 HLA heterozygote advantage and risk of colorectal cancer according to molecular tumor subtypes

Authors:

C. E. Thomas[1], Y-Y. Tsai[2], J. R. Huyghe[1], H. Yin[1], R. S. Steinfelder[1], C. Qu[1], L. Hsu[1], S. Thomas[1], D. D. Buchanan[3], M. A. Jenkins[3], S. Gallinger[4], R. Grant[4], P. T. Campbell[5], C. Newton[6], C. Y. Um[6], J. Chang-Claude[7], M. Hoffmeister[7], H. Brenner[7], A. J. Pellatt[8], R. Schoen[9], B. Van Guelpen[10], A. T. Chan[11], S. Ogino[12], M. O. Woods[13], M. J. Gunter[14], S. B. Gruber[15], A. I. Phipps[1], S. L. Schmit[2], U. Peters[1]; [1]Fred Hutchinson Cancer Ctr., Seattle, WA, [2]Cleveland Clinic, Cleveland, OH, [3]The Univ. of Melbourne, Parkville, Australia, [4]Univ. of Toronto, Toronto, ON, Canada, [5]Albert Einstein Coll. of Med., New York, NY, [6]American Cancer Society, Atlanta, GA, [7]German Cancer Res. Ctr., Heidelberg, Germany, [8]Univ. of Utah, Salt Lake City, UT, [9]Univ. of Pittsburgh, Pittsburgh, PA, [10]Umeå Univ., Umeå, Sweden, [11]Harvard Med. Sch., Boston, MA, [12]Brigham & Women's Hosp., Boston, MA, [13]Mem. Univ. Newfoundland, St John's, NL, Canada, [14]Intl. Agency for Res. on Cancer, Lyon, France, [15]City of Hope Natl. Med. Ctr., Duarte, CA

Background: Germline genetic factors central to immunity, such as human leukocyte antigen (HLA) variants, have been associated with tumor immune microenvironment features and response to immune checkpoint inhibitors. HLA heterozygote advantage, where heterozygosity may be related to presentation of cancer neoantigens in more diverse contexts for T cell recognition, may have differential effects by colorectal cancer (CRC) molecular subtype. Methods: Utilizing the resources of the Genetic Epidemiology of Colorectal Cancer Consortium (GECCO) and the Colon Cancer Family Registry (CCFR), classical HLA alleles were imputed from genome-wide array datasets through SNP2HLA. Eleven observational studies including up to 7,241 cases with molecular subtypes for microsatellite instability (MSI), CpG island methylator phenotype (CIMP), and somatic mutations in *BRAF* and *KRAS* genes, and 10,510 controls were analyzed, all of European descent. Associations between HLA class I and class II heterozygosity and risk of developing CRC stratified by molecular subtypes were assessed using case-only logistic regression and multinomial regression adjusted for age, sex, GWAS panel, and the first five principal components of ancestry. Results: A higher number of homozygotes for class II HLA alleles was associated with a higher odds of developing *BRAF*-mutated tumors [Odds ratio (OR) 3 homozygotes vs. heterozygotes = 1.61, 95% confidence interval (CI) (1.08, 2.38)] but not *BRAF*-wildtype tumors [1.06, 95% CI (0.85, 1.32); P heterogeneity = 0.084]. No other overall measure of class I or class II heterozygosity was associated with a meaningful difference in trend; however, some specific HLA allele differences where heterozygote was the reference group were significant for HLA-C by MSI status [MSI-high OR = 0.87, 95% CI (0.69, 1.10); MSI-low/MSS OR = 1.11, 95% CI (0.99, 1.23); P heterogeneity 0.044] and HLA-DRB1 by *BRAF* status [*BRAF*-mutated OR = 1.30, 95% CI (1.04, 1.64); *BRAF*-wildtype OR = 0.97, 95% CI (0.86, 1.09); P heterogeneity 0.025]. Conclusions: HLA heterozygosity association with CRC risk may vary by CRC tumor molecular subtype, where less heterozygosity may be associated with higher risk for *BRAF*-mutated tumors.

Session Title: Cancer Poster Session II

PB5078 Identification of *331G.A* Polymorphism of *PR*gene and *397T.C* and *351A.G* of the *ER alpha*gene as genetic predisposition markers for breast cancer management in yaoundé general hospital.

Authors:

N. NGUEDIA KAZE; Covenant Univ., Ota city, Nigeria

Background : 30% of breast cancer patients in Cameroon usually die a year after diagnosis, since cancer is an accumulation of molecular changes in gene, the *+331 G.A* polymorphism of *PR* gene and *397T.C* and *351A.G* polymorphisms of ER Alpha gene have been investigated as etiological factors. **Methods :** 16 cases diagnosed positive for breast cancer at the Yaoundé general hospital were recruited through the patient file consulting for the case control study. Blood samples were collected from them and 22 healthy women recruited using a questionnaire and a inform concern was signed by each of them. *+331 G.A* polymorphism in the *PgR* gene was identified using NIaIV endonuclease and direct molecular haplotyping was used to determine the relationship between the 2 polymorphisms in the *ER* Alpha gene, *397T.C* and *351A.G* using PvUII and Xbal endonucleases by PCR-RFLP. The data were analyzed using SPSS v20. **Results**: We got a mean age of 53+/-11,08 in our cancerous group with the predominance of invasive ductal carcinoma. We obtained an Odds Ratio of 1.067 with 95% Confidence Interval of 0.940-1.211 and P value of 0.421, making a non-significant association of *331G.A RP* mutation with breast cancer. In addition, we identified the mutant *CG* haplotype allele of the *ER* alpha gene with a predominance in the cancer group exclusively with a frequency of 56.25% with an OR of 2.022 with a 95% CI 0.432 - 9.461 with a p-value of 0,3, not significantly associated with breast cancer. **Conclusion**: Based on our study findings, the specific genetic markers investigated, did not show a significant association with the risk of breast cancer development in our study population.

Session Title: Cancer Poster Session III

PB5079 Identification of a robust microRNA signature for estimating survival time in patients with hepatocellular carcinoma.

Authors:

S. Yerukala Sathipati[1], N. Aimalla[2], M-J. Tsai[3], R. Sharma[2], S. Shukla[1]; [1]Marshfield Clinic Res. Inst., Marshfield, WI, [2]Marshfield Clinic Hlth.System, Marshfield, WI, [3]Hinda and Arthur Marcus Inst. for Aging Res. at Hebrew Senior Life, Boston, MA

Hepatocellular carcinoma (HCC) is one of the most commonly diagnosed liver malignancy and leading cause of cancer-associated mortality worldwide. Most HCCs are resistant to systematic treatments, including surgery, radiation, immune and chemotherapies. Due to the ongoing advancements in the genetic field, identifying microRNA (miRNA) markers in HCC may unravel the disease mechanism and help estimate survival outcome in patients with HCC. Accordingly, we aimed to identify a survival-associated miRNA signature that estimate survival time in patients with HCC. We extracted the miRNA expression profiles of 373 patients with HCC from The Cancer Genome Atlas Database and developed a machine learning-based survival estimation method called HCC Survival Time (HST). The HST method was designed using optimal feature selection algorithm incorporated with support vector regression. HST selected a robust miRNA signature consisting of 32 miRNAs and obtained a mean R and MAE of 0.87±0.02 and 0.73 year, respectively between actual and estimated survival time of patients with HCC. The miRNAs of the identified signature has four diagnostic miRNAs, including hsa-miR-1301-3p, hsa-miR-17-5p, hsa-miR-34a-3p, and hsa-miR-200a-3p, and prognostic miRNAs, including hsa-miR-146a-3p, hsa-miR-200a-3p, hsa-miR-652-3p, hsa-miR-34a-3p, hsa-miR-132-5p, hsa-miR-1301-3p, and hsa-miR-374b-3p, of HCC. The Kyoto Encyclopedia of Genes and Genomes pathway analysis and miRNA-disease association results revealed that identified miRNA signature enriched in hepatitis B pathway. We conclude that survival associated miRNA signature can estimate the prognosis in patients with HCC and further exploration of these miRNAs would help determine mechanism for therapeutic interventions.

Session Title: Cancer Poster Session I

PB5080 Identification of Aurora A as a Novel Oncogene in Hepatocellular Carcinoma and targeting it with CRISPR/Cas9 system

Authors:

m. el-hefnawy, N. Samir, b. soliman, R. Hassan; Natl. research center, cairo, Egypt

Hepatocellular carcinoma (HCC) is one of the most lethal cancers worldwide. There are still challenges for HCC treatments, especially high resistance of the cancer cells to chemotherapy and/or target therapy. The clustered regularly interspaced short palindromic repeats (CRISPR/Cas9) system has demonstrated considerable advantages over other nuclease-based genome editing tools due to its high accuracy, efficiency, and strong specificity. The aim of the present study was to identify the hub genes and underlying pathways of HCC via bioinformatics analyses and using CRISPR/Cas9 system for treatment HCC. The differentially expressed genes (DEGs) were analyzed for functional enrichment pathways and protein-protein interaction (PPI) network was constructed. We knocked out (KO) AURKA(Aurora A) gene using CRISPR-Cas9 system to evaluate its effect on tumor proliferation, migration and apoptosis in HCC. We report that AURKA was highly expressed in liver cancer, showing its promise to be developed further into an anti-cancer agent. We identified differentially expressed genes (DEGs) based on four Gene Expression Omnibus(GEO) datasets. A protein-protein interaction (PPI) network was established to identify the central nodes associated with HCC. Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis of the central nodes were conducted in online tool Database for Annotation, Visualization and Integrated Discovery (DAVID) and using the Cytoscape software to find the 14 top hub genes. This hub genes are : CDK1, CCNA2, CCNB2 ,CCNB1, CDC20, UBE2C, BUB1B, CDKN3, CDC6, AURKA, BIRC5, MAD2L1, MELK, CKS2. AURKA was targeted for Knock out by CRISPR/Cas9 system. We designed AURKA gene specific single guide RNA (sgRNA) with CRISPR/Cas9 system and studied in vitro effects on tumor properties of HepG2. We cloned the sgRNA into pCAG-eCas9 CRISPR vector then transfected into cell culture by lipofectamine transfecting agent. The expression of AURKA was investigated in transfected cell culture by real time PCR and Western blot then using many assays for screening the effect on the tumor proliferation, apoptosis and migration: cell cycle assay, MTT assay, wound healing assay, apoptosis assay. The study identified several genes and the signaling pathways that were associated with tumorigenesis using bioinformatics analyses, which could be potential targets for the diagnosis and treatment of HCC. AURKA was overexpressed in HCC tissues and inhibition the apoptosis process and increasing the cell cycle and cell migration. The study provides evidence that CRISPR/Cas9-mediated AURKA knockout suppresses HCC cell proliferation and migration.

Session Title: Cancer Poster Session II

PB5081 † Identification of circulating epigenetic biomarkers for lesion burden in familial Cerebral Cavernous Malformation by whole-genome methylome profiling.

Authors:

S. Weinsheimer[1], J. Nelson[2], M. C. Mabray[3], B. L. Hart[3], L. Morrison[4], C. E. McCulloch[5], H. Kim[6]; [1]Dept. of Anesthesia and Perioperative Care, Ctr. for Cerebrovascular Res., Inst. for Human Genetics, Univ. of California San Francisco, San Francisco, CA, [2]Dept. of Anesthesia and Perioperative Care, Ctr. for Cerebrovascular Res., Univ. of California San Francisco, San Francisco, CA, [3]Dept. of Radiology, Univ. of New Mexico, Albuquerque, NM, [4]Dept. of Neurology, Univ. of New Mexico, Albuquerque, NM, [5]Dept. of Epidemiology and Biostatistics, Univ. of California San Francisco, San Francisco, CA, [6]Dept. of Anesthesia and Perioperative Care, Ctr. for Cerebrovascular Res., Epidemiology and Biostatistics, Inst. for Human Genetics, Univ. of California San Francisco, San Francisco, CA

Background: Individuals with familial Cerebral Cavernous Malformation (fCCM) present with multiple lesions that can grow in number and size over time which may cause hemorrhagic strokes. Lesion burden is highly variable, suggesting environmental or genetic disease modifiers. Comprehensive DNA methylome profiling has revealed biomarkers for cardiovascular disease and cancer, and may be a useful tool to identify disease monitoring biomarkers for CCM and novel therapeutic targets to stabilize CCM lesions. We hypothesized that the methylome differs between fCCM cases with greater lesion burden compared to cases with lower lesion burden.
Methods: We selected 10 individuals enrolled in the Brain Vascular Malformation Consortium CCM study at the University of New Mexico as representative cases with *CCM1* Q455X genetic mutation and extreme CCM lesion burden (n=5, range 51-619, mean age 47.5y and n=5, range 3-5, mean age 42.0y). Genomic DNA from whole blood was profiled for genome-wide methylation using Whole Genome Bisulfite Sequencing services at Psomagen (~20x coverage). Differentially Methylated Regions (DMR) were identified by comparing the average methylation ratio in fCCM cases with high vs. low lesion count by *t*-test. DMRs with a delta average methylation ratio between the two groups ≥|0.2| and P≤1x10-06 were flagged as both statistically significant and having a large effect size.
Results: Methylome profiles differed between fCCM cases with high vs. low lesion burden with 259 significant DMRs (53% hypermethylated in high lesion burden) mapping within or nearby genes. Ten DMRs mapped to CpG islands for the following genes: *RNA45S5*, *EHBP1L1*, *LBX2-AS1*, *AATK*, *SOX11*, *CDH18*, *NFIC*, *SEMA5B* and *NKX6-2*. DMRs with the largest differences (≥|0.5|) mapped to the promoter *LRRC14B*, 3' untranslated region *EHBP1L1*, exonic region *UACA*, downstream *DDX12P* and *METTL2B*, and intronic regions of the following genes: *MYPN*, *IGFBP7*, *CCSER1*, *CNOT4*, *MAST4*, *ADGRL2*, *ADGRB1*, *OTUD7A*, *KRT72*, *NDUFAF5*, *ADK*, *LMTK2*, *ZC3H6*, *ADCY2*, *SH3RF3* and *PLBD1*. Interestingly, two DMRs associated with lesion burden map to *RASGRF2* and *IGF1R*, which both function in the RAS and MAPK signaling pathways important to CCM pathogenesis.
Conclusions: We performed the first whole-genome methylome profiling of individuals with fCCM and identified several differentially methylated genomic regions associated with lesion burden. DNA methylation modifications of novel genes may serve as a new layer of biological regulation that contributes to CCM pathogenesis. Further studies are needed to validate these findings in a larger fCCM cohort and investigate the role these epigenetic biomarkers have in CCM disease.

Session Title: Cancer Poster Session III

PB5082 Identification of Novel Biomarkers in African American Prostate Cancer (Pca) Samples

Authors:

H. Wei[1], N. Palanisamy[2]; [1]Beckman Coulter Life Sci., Indianapolis, IN, [2]Henry Ford Hlth., Detroit, MI

Identification of novel biomarkers in African American prostate cancer samples Abstract:Introduction: Prostate cancer (PCa) is the most frequent cancer in men in the Western world. In addition, population-based studies demonstrate that African Americans (AA) in the United States have an increased risk of developing prostate cancer and a 4.4% risk of death compared with a 2.4% risk of death in Caucasians. Fusion genes, formed by joining parts of two different genes, are one of the fundamental drivers contributing to the development of PCa. In this study, we aimed to utilize this knowledge and RNA-seq to identify transcriptomic signatures for PCa in AA.Method: We used PCa AA FFPE (N=5) and fresh frozen (N=12) samples in this proof of principle study. First, FFPE RNA was extracted from a single 10 µm curl input using FormaPure XL RNA (Beckman Coulter Life Sciences, Indiana, U.S.). Then, FFPE RNA with DV200 >= 30% and RNA from fresh frozen samples was used for library preparation and sequencing. After sequencing, the raw reads in fastq files were processed using the Trim Galorescript to remove adapter sequences and low-complexity regions. Next, STAR software mapped the trimmed reads to the human genome GRCh38. Finally, the differentially expressed genes (DEGs) between the in-house PCa AA cohort (n=17) and The Cancer Genome Atlas (TCGA) PRAD cohorts were identified using DESeq2. Gene lists were compared using the Venny online tool.Results: We first examined the FFPE RNA yield from 5 PCa AA samples to ensure optimal quality for RNA-Seq workflow. We observed that sufficient yield was obtained from all samples and exceeded the 600 ng required for RNA sequencing. We also assessed the RNA integrity, andfound that the DV200 was above 30% for all of the extracted samples. The average unique reads across all FFPE-derived PCa AA samples were 68 million. Using the TCGA PRAD dataset, we identified 3936 downregulated and 9696 updated genes in PCa AA groups. Commonly regulated protein-coding genes (mRNA) and lncRNA-coding genes (long non-coding genes) were detected in PCa AA and TCGA PCa AA groups. Functional enrichment analysis of the differentially expressed genes revealed a link between top biological processes and associated pathways to the PCa AA group.Conclusion:Our study suggests that these genes might be involved in different networks that lead to uniquely aggressive PCabehavior in AA samples, potentially serving as biomarkers for the racial disparity in PCa progression. This proof of principle study also demonstrates the successful use of FFPE RNA for NGS analysis in cancer research

Session Title: Cancer Poster Session I

PB5083 Identification of sex specific molecular features in glioblastoma

Authors:

Y. Huang[1], Y. Shan[1], W. Zhang[1], C. Printzis[2], L. Pesce[2], B. Stranger[2], R. Huang[1]; [1]Univ. of Minnesota, Minneapolis, MN, [2]Northwestern Univ., Chicago, IL

Glioblastoma (GBM) is a lethal disease with limited treatment options available. Sex differences in GBM are well-documented in clinical settings, with males having twice the risk of developing the disease and increased mortality compared to female counterparts. To investigate the molecular basis of sex differences in GBM, we characterized the somatic genetic alterations in *IDH1* wild-type GBM samples from the GENIE project (n=4530). We also analyzed transcriptomic sex differences in a collection of three independent GBM datasets (n=346) and validated them in three additional independent studies. The impact of differentially expressed genes were assessed through Gene Set Enrichment Analysis (GSEA) on approximately 5000 gene sets to interpret their biological relevance. Given the well-recognized heterogeneity of GBM, we also integrated three single-cell RNA (scRNA) sequencing datasets to decipher the sex differences in cellular composition between male and female tumors. We identified 91 genes exhibiting sex-biased mutation patterns in GENIE GBM samples, including the Androgen Receptor (AR) signaling pathway. The effects of sex on gene expression are modest and primarily driven by the X chromosome (14 out of 16 differentially expressed genes were on the X chromosome). These genes are associated with escape from X-chromosome inactivation (e.g., *KDM5C*, *KDM6A*) and are also sex-differentially expressed (DE) in the healthy human brain according to Genotype-Tissue Expression (GTEx) data. GSEA results suggest increased activity in cell proliferation and neuronal signaling in male GBM samples, whereas females show enrichment in glycoprotein/lipid metabolism and inflammatory response. By deconvoluting the cell-type composition, we identified a higher percentage of pericytes in the male GBM tumor microenvironment compared to those in females, which is consistent with recent spatial-omic discoveries in GBM. By integrating scRNA sequencing data, we found that both differentially expressed genes and pathways show cell-type-specific enrichment. For example, the female-biased gene *GBGT1* is exclusively expressed in the myeloid lineage, while male-biased neuronal pathways are enriched in cancer stem cells. Integrating with the inference of cell trajectories, we observed that male tumors harbor more unconnected, invasive malignant cells than females, which might explain the poor prognosis in male GBM patients. In summary, our results provide a comprehensive multi-level characterization of sex differences in GBM tumors, and these findings will provide novel insights into GBM disease progression and development of new therapy.

Session Title: Cancer Poster Session II

PB5084 Identifying functionally methylated regions in breast cancer

Authors:

H. Xu, V. George; Augusta Univ., Augusta, GA

DNA methylation is an important genomic modification involved in the regulation of gene expression, cell differentiation and developmental processes. Abnormal methylation has been implicated in human diseases such as various forms of cancers. There has been great interest in identifying genes with differentially methylated regions (DMRs) between different biological conditions. With the high-throughput technologies, large-volumes of methylation data at genomic level have been generated. This also poses challenges in data management, analysis, and interpretation. Taking advantage of the high-quality genomic data, various strategies are employed for detecting functional differentially methylated regions (fDMRs) by integrating both methylation data and gene expression data. One such approach is based on unified Bayesian hierarchical modeling. We model DNA methylation and gene expression in a hierarchical way, based on the assumption that fDMRs will show changes in methylation levels, which then lead to changes in gene expression. The evidence of differential methylation will be used as prior information to update the evidence of differential gene expression. Integrating methylation and gene expression data will help address some of the challenges in DMR identification methods and more effectively identify fDMRs, which can help elucidate the genetic mechanisms of complex human diseases. We applied our method to breast cancer data from the Cancer Genome Atlas and identified 6 fDMRs with FDR < 0.05.

Session Title: Cancer Poster Session III

PB5085 Impact of bioinformatics pipelines and prediction tools on the detection of breast cancer susceptibility variants: application to the GENESIS case-control study.

Authors:

F. Lesueur[1], S. Murat El Houdigui[1], E. Girard[1], P. Apostolou[2], S. Eon-Marchais[1], M-G. Dondon[1], J. Beauvallet[1], D. Le Gal[1], R. Olaso[3], C. Baulard[3], M-T. Bihoreau[3], A. Boland-Augé[3], GENESIS study collaborators, J-F. Deleuze[3], D. Stoppa-Lyonnet[4], N. Andrieu[1], N. Servant[1]; [1]Inst. Curie, Inserm U900, Mines ParisTech, Paris, France, [2]Human Molecular Genetics Lab., INRaSTES, Natl. Ctr. for Scientific Res. Demokritos, Athens, Greece, [3]Université Paris-Saclay, CEA, Ctr. Natl. de Recherche en Génomique Humaine, Evry, France, [4]Service de Génétique, Inst. Curie, Université de Paris, Inserm, U830, Paris, France

Epidemiological studies make increasing use of massively-parallel sequencing to identify rare alleles involved in susceptibility to complex traits like cancer. To analyze sequencing data, bioinformatics pipelines perform a chain of data-processing tasks that use tools to annotate and filter variants. Gene-based burden testing is then performed by comparing the aggregate burden of rare, protein-altering variants in each gene between case and control subjects. Here we assessed how the choice of tools for annotating, filtering and interpreting variants affect the results of a case-control study on breast cancer. We examined the consequence of these choices in the GENESIS study where sequencing data for 113 genes were available for 1207 women with familial breast and 1199 controls.
We first compared number and types of variants detected by the original pipeline (Girard et al. 2019, PMID:30303537) and by an upgraded pipeline that uses more recent tools for alignment of bam files (BWA aln was replaced with BWA-MEM), variant calling (the use of the three callers GATK Preprocess, GATK Unified Genotyper and Freebayes was replaced with GATK Haplotype Caller) and annotation (Annovar was replaced with SnpEff and SnpSift). This new pipeline detected less variants (3568 vs 4023) but a higher proportion of loss-of-function variants (LoF) than did the previous one (17.5% vs 9.2%). Especially, the new pipeline was more efficient to detect splice site variants. A total of 3081 variants (68%) were detected by both pipelines of which 43% LoF and 73% missense variants (MV). Despite these differences, analyses aggregating all LoF and MV with CADD score ≥20 confirmed the association of *ATM*, *CHEK2* and *PALB2* with breast cancer, with similar risk estimates (in 2023: OR_{ATM}=2.0 (1.4-3.1), OR_{CHEK2}=2.8 (1.8-4.6), OR_{PALB2}=3.8 (1.8-8.7); in 2019: OR_{ATM}=1.9 (1.3-2.9), OR_{CHEK2}=3.0 (1.9-5.0), OR_{PALB2}=3.5 (1.7-7.5)).
We also assessed the influence of nine ensemble methods other than CADD built on different training and test sets, namely BayesDel, BayesDel no AF, ClinPred, DANN, META-LR, META-SVM, PON-P2, REVEL and VEST4, and found that the number of filtered variants varied from 14% to 66% according to the scoring method. Consequently, additional genes were associated with breast cancer when using some alternative methods.
Hence, sequencing data processing influence the number of variants included in the analyses. Notably, results are impacted by the choice of the prediction models employed for interpretation of deleteriousness of MV. Even if there is no effect on risk estimates, this may impact the list of variants to further investigate in a clinical context.

Session Title: Cancer Poster Session I

PB5086 Impact of panel size on minimum residual disease (MRD) assay performance.

Authors:

C. J. Battey, A. Acevedo, M. LaBella, S. Ganesh, E. Buser, N. Xu, R. Patel, K. Haug, K. Trettin, J. Jasper, J. Risinger, G. Gould, D. Muzzey; Myriad Genetics, Inc., Salt Lake City, UT

Background: Sequencing circulating tumor DNA is a promising method for monitoring cancer treatment response and detecting recurrence, but sensitivity at the low tumor fractions typical of early-stage, post-treatment, and early recurrent tumors is limited by the small number of variants targeted in commercially available assays. Methods: We developed a tumor-informed minimum residual disease (MRD) assay using tumor-normal whole-genome sequencing (WGS) followed by interrogation of select somatic variants in cell-free DNA (cfDNA). We describe theoretical limits on panel designability as a function of tumor purity and sequencing depth using simulations and a discovery cohort of 31 patient samples. cfDNA assay performance was tested by creating serial dilutions of patient plasma before somatic enrichment and sequencing. Simulations and *in-silico* downsampling of target sites were used to test sensitivity as a function of panel size. A bioinformatics pipeline incorporating somatic calling, copy number inference, error assessment, and tumor-fraction estimation was developed to analyze results. Results: Across a range of tissues, tumor-normal WGS identified from hundreds to tens of thousands of somatic variants to inform an MRD capture panel. 97% of samples tested returned at least 300 high-confidence somatic variants, and targeted sequencing demonstrated over 95% positive predictive value in somatic variant calls. In cfDNA dilution experiments, we observed >99% sensitivity at a tumor fraction of 0.01%. Expanding the panel size from 16 to 300 target variants increased sensitivity at 0.005% tumor fraction from approximately 20% to >95%. Our tumor fraction model returned accurate quantification down to the parts per hundred thousand (0.001%) range, even when only a single tumor molecule could be observed at each target site. Conclusions: Increasing the number of targeted variants in an MRD assay increases sensitivity at low tumor fractions, and WGS-driven panel design ensures a sufficient number of interrogated variants while minimizing the failure rate due to panel designability. High sensitivity MRD has the potential to enable earlier recurrence detection while also giving researchers and clinicians new tools for monitoring patient treatment responses.

Session Title: Cancer Poster Session II

PB5087 Impact of sex, age, and ancestry on Apolipoprotein E (*APOE*) risk for Alzheimer's disease (AD): a pooled case-control of Alzheimer's Disease Genetics Consortium (ADGC).

Authors:

O. Garcia Rodriguez[1], L. Wang[1,2], J. Young[1], Alzheimer's Disease Genetics Consortium (ADGC), M. A. Pericak-Vance[1], E. Martin[1], K. Brian[1]; [1]Hussman Inst. for Human Genomics, Univ. of Miami, Miami, FL, [2]Dept. of Publ. Hlth.Sci., Univ. of Miami, Miami, FL

Background/Significance: Previous studies have reported that non-Hispanic white (NHW) females carrying the *APOE ε*2 or ε4 allele differ in risk of developing Alzheimer's disease (AD) when compared to men. However, several studies also indicated that the risk for developing AD may be similar for males and females in carriers of *APOE ε4* but may differ by age. We aimed to comprehensively evaluate the association between sex, age, ancestry, and *APOE* with AD in multiple cohorts. **Materials/Methods:** A pooled case-control study was conducted for 50 independent datasets (n = 42,295 individuals with ages ≥ 55 years) from the Alzheimer's Disease Genetics Consortium (ADGC). Logistic, linear, and Cox regression models were used to study the associations between disease status and independent variables age, sex, self-reported race (as a proxy of ancestry), and *APOE* genotype. A *p-value* less than 0.05 was considered statistically significant. **Results:** Ancestry-specific analyses showed when comparing subjects with *APOE* ε3/ε4 (vs. ε3/ε3), East-Asian (EA) females (*OR*, 5.00; 95% CI: 4.09-6.12) and males (*OR*, 4.20: 3.26-5.42) had strongest *ORs*; whilst the weakest risk difference was for South-Asian (SA) men (*OR*, 1.23: 0.66-2.29). Nevertheless, men and women with *APOE* ε3/ε4 showed no significant difference in the risk of developing AD after accounting for age among ancestries. Age group-specific analyses showed that NHW females (*OR*, 4.66: 4.12-5.27 vs. males: *OR*, 3.99: 3.46-4.60) and African American (AA) females (*OR*, 4.27: 3.09-5.91 vs. AA men: *OR*, 1.98: 1.23-3.16) with *APOE* ε3/ε4 presented a higher risk of developing AD between ages 66-75; whereas EA and SA exhibited this pattern at 55-64. In Hispanics, the relationship was reversed, with males aged 55-65 years carrying *APOE* ε3/ε4 (*OR*, 3.42: 1.22-9.60: vs. females: *OR*, 2.30: 1.18-4.50) having a stronger association with AD risk than females. Nonetheless, these differences were not statistically significant. Case-only and survival analyses also established that NHW, AA, and EA female carriers of APOE ε3/ε4 had an earlier age at the onset of AD than their male counterparts. When comparing *APOE* ε2/ε2 and ε2/ε3 carriers vs. ε3/ε3 carriers, NHW women (*OR*, 0.64: 0.55-0.74) and men (*OR*, 0.50: 0.44-0.57) presented the lowest *ORs* for risk of AD among all ancestries. No differences were found when stratified by age for *APOE* ε2 and AD. Nonetheless, NHW females had later age of onset for AD than NHW males. **Conclusions:** Our findings suggest sex differences by age and ethnicity in the risk for developing AD and age of onset for AD in carriers of *APOE* ε2 and ε4. However, additional local ancestry analyses for APOE are warranted to confirm these findings.

Session Title: Cancer Poster Session III

PB5088 Improving Diagnostic Yields of Cancer Predisposing Variants in Diverse Populations

Authors:

K-l. Huang, A. Elmas, R. Vaza; Icahn Sch. of Med. at Mount Sinai, New York, NY

Cancer risk differs across ancestries. According to the data provided by NCI SEER, cancer incidence per 100,000 individuals is 438 in Whites, 433 in Blacks, 289 in Asian/Pacific Islanders, 320 in American Indian/Alaskan Natives, and 327 in Hispanics from 2011-15. Although factors such as healthcare access or diet may account for some of these disparities, it is likely that a significant proportion of these differences can be attributed to genetic factors unique to each ancestry. In this study, we undertook a comprehensive analysis of the ethnically diverse Mount Sinai BioMe cohort, which includes over thirty-thousand cases with quality-controlled whole-exome sequencing (WES) data. By integrating WES data with electronic health records (EHR), we identified over five-thousand cases or individuals with a family history of cancer. The distribution of known pathogenic variants varies among different types of cancer, and also among different population groups that reflect potential bias in existing databases. Notably, by incorporating protein-truncating variants (PTVs) in ACMG and cancer susceptibility genes, we enhanced the detection rates across genetically-defined populations. Importantly, we developed a rare variant polygenic risk score (rvPRS), which can effectively identify high-risk individuals across varied ancestry groups. Consequently, our study highlights the potential of using a rare variant PRS to more equitably identify high-risk individuals across diverse populations.

Session Title: Cancer Poster Session I

PB5089 In vitro effect of curcumin in combination with chemotherapy drugs in Ph+ acute lymphoblastic leukemia cells

Authors:

U. Santana, U. Santana-Bejarano; Univ. de Guadalajara, Guadalajara, Mexico

Philadelphia chromosome-positive acute lymphoblastic leukemia (Ph+ ALL), is characterized by the t(9;22)(q34q11) that generates the BCR-ABL protein with uncontrolled tyrosine kinase activity. Recently, a connection between BCR-ABL signaling with NF-κB activation mediated by CK2 has been hypothesized. Approximately 95% of patients with Ph+ ALL have the BCR-ABLp190 isoform, which causes aggressive leukemia with a high rate of chemotherapy resistance. Therefore, the use of compounds that could improve the efficacy of chemotherapy drugs is of particular interest. Curcumin is an active chemical in turmeric with antineoplastic potential; it regulates protein-kinases by modulating downstream molecular pathways. The present study evaluated the effect of curcumin in combination with the chemotherapeutic drugs vincristine, imatinib and daunorubicin in the human OP-1 cell line. Several doses of the chemotherapy drugs were examined, and the effects were evaluated following 12, 24 and 48 h of exposure. The interaction between the chemotherapy drugs and curcumin was determined by the dose-effect curve, which generated a combination index (CI); these data were represented in isobolograms. In addition, the individual effect of each drug was compared with its effect in combination with curcumin on cell viability, apoptosis degree, NF-κB activation and gene expression changes. The present study observed that curcumin potentiates the efficacy of vincristine and imatinib, generating an additive/synergistic effect in a dose- and time-dependent manner. These combinations significantly increased the apoptosis degree, decreased the activation of NF-κB and the expression of its regulated genes. Conversely treatment with daunorubicin + curcumin combination produced an antagonistic/additive effect in a dose-dependent manner, and this combination significantly increased the apoptosis degree. However, this effect seems not to be associated with NF-κB activity, as no significant changes were observed in its activation or in the expression of the genes that it regulates. The results of the present study demonstrate that curcumin may be used as an adjuvant agent for chemotherapy in patients with Ph+ ALL.

Session Title: Cancer Poster Session II

PB5090 Initial sequencing and analyses of a new broadly-consented tumor cell line for development of a Genome in a Bottle Tumor/Normal Benchmark

Authors:

J. Wagner[1], G. Rosen[2], J. McDaniel[1], A. Liss[3], J. Zook[1]; [1]Natl. Inst. of Standards and Technology, Gaithersburg, MD, [2]Drexel Univ., Philadelphia, PA, [3]Massachusetts Gen. Hosp., Boston, MA

Here, we describe initial analyses on PCR-Free Illumina WGS and Hi-C sequencing with the goal of developing the first tumor/normal benchmark from the Genome in a Bottle (GIAB) consortium. This effort builds on the widely-used GIAB germline variant benchmarks for seven normal DNA reference materials. We performed sequencing on new cell lines from a broadly-consented pancreatic ductal adenocarcinoma patient and we plan to perform an array of PacBio HiFi, ONT ultralong and duplex, Bionano, and Bioskryb single cell WGS from a large batch of the tumor cell line, along with other technologies in the future.

Using Hi-C for copy number and structural variant analysis, preliminary results identified substantial aneuploidy common in pancreatic tumors, with ~17 large inversions and translocations and 16 chromosomes with extensive loss of heterozygosity due to missing >30% of one copy. Many of these large events appeared to occur in all cells, but some deletions and duplications occur in only a fraction of the cells. The tumor contains the common G12V mutation in KRAS, and interestingly the ~2 Mbp region containing this mutation is likely triplicated.

To assess sources of error in variant calls, we aligned the Illumina PCR-Free data at coverages of ~150x on normal and ~180x on the tumor to GRCh38 then ran Lancet, Mutect2, Strelka2, DRAGEN, and NeuSomatic. We compared Mutect2 calls to Strelka2 using default filtering and found that after excluding difficult regions using the GIAB stratifications more than 90% of SNVs and small indels agree. We categorized disagreements to find that many occur because of systematic sequencing errors, alignment errors, low frequency, and proximity to germline variants. Examining discordant SNVs in difficult regions, most are erroneous somatic variants in regions with loss of heterozygosity that also have mapping errors due to segmental duplications, sequences missing from GRCh38, and/or tandem repeats. After excluding difficult regions, we also found that somatic SNVs fell in 4 primary classes: ~5000 in most cells in diploid regions, ~2000 in most cells in haploid regions, ~800 in only some cells in diploid regions, and ~500 in only some cells in haploid regions.

GIAB is forming an open working group to develop the first authoritative benchmarks from publicly available, broadly-consented tumor-normal cell lines and long read sequencing data will be made public as soon as possible after it is generated.

Session Title: Cancer Poster Session III

PB5091 Insights into the Development of *BAP1* Mutant Uveal Melanomas Following Spontaneous Differentiation of iPSCs into a Self-Formed Ectodermal Multi-Zone primitive eye structure

Authors:

P. Begani, A. Z. Eriksen, T. A. Blenkinsop, A. M. Bowcock; Icahn Sch. of Med. at Mount Sinai, New York, NY

Uveal melanoma (UM) arises from neural crest of the eye and has a mortality rate of 80%. Almost all UMs harbor oncogenic mutations in *GNAQ*, *GNA11* or *CYSLTR2*. However, these do not affect their metastatic competency which is defined by a classic gene expression profile. Class 1 tumors (low risk) usually harbor mutations in *SF3B1* or *EIF1AX*. 84% of Class 2 tumors (high risk) harbor inactivating mutations in BRCA1-Associated Protein 1 (*BAP1*) gene, along with loss of the chromosome 3 homolog harboring WT *BAP1*. *BAP1* encodes a ubiquitin carboxy-terminal hydrolase that regulates multiple cellular processes including chromatin modification, DNA repair and cell cycle control. *BAP1* is also mutated in several other highly aggressive cancers. Cell based studies have shown that loss of *BAP1* function in UM cells leads to a stem-like phenotype. However, the role of *BAP1* in tumorigenesis and metastasis remains largely unknown. We hypothesized that understanding the consequences of *BAP1* loss in development would provide insights into its role in the development of UM and other cancers. We used human iPSCs and induced them to differentiate to form a primordium of four concentric zones termed Self-formed Ectodermal Multi-Zone (SEAM). The SEAM mimics whole eye development and cells within each zone are indicative of lineage. SEAMs were stained for several lineage and ocular markers. *BAP1* localized to Zone 2, which is the developmental analog of neural crest cells. Zone 2 also developed dense melanin pigmentation at around Day 20 of differentiation. Next, using the CRISPR-Cas9 based system, we successfully generated iPSCs with downregulated *BAP1* which were then also differentiated into ocular SEAMs. RNA-sequencing revealed significant alterations in various pathways as a consequence of *BAP1* knockdown. These included melanogenesis, morphogenesis (WNT signaling), ECM assembly and degradation, chromatin modeling and DNA repair. *BAP1* downregulation also led to reduced melanin pigmentation in Zone 2 as well as to decreased BAP1 deubiquitinase activity, consistent with a role for *BAP1* in melanogenesis. Intersection of differentially expressed transcripts from SEAM *BAP1* knockdown with those of primary tumors with *BAP1* loss of function mutations revealed shared altered pathways of cell cycle control, translation regulation, ECM assembly and degradation and EMT. Hence, ocular SEAMs can provide insights into eye development, neural crest differentiation and potentially tumorigenesis and metastasis.

Session Title: Cancer Poster Session I

PB5092 Integrating multiomics data to identify African ancestry-informative markers acting as eQTLs in breast cancer

Authors:

S. Boudeau; Temple Univ., Philadelphia, PA

Ancestry-informative markers (AIMs) are SNPs that are found at relatively higher frequencies in a specific population and can be used to estimate genetic ancestry. Genetic ancestry is becoming more commonly used in biomedical research investigating the genetic basis of diseases. For many cancers, the literature suggests that African ancestry is associated with poor response to treatment, poorer prognosis, and poorer survival. More specifically, African American women with breast cancer are more likely to have more aggressive subtypes of the disease and diagnosed with a later stage of the disease at an earlier age compared to their European American counterparts. Some of this disparity remains after accounting for socioeconomic factors, suggesting a genetic basis for the health disparities observed in many cancers. To evaluate the genetic basis of cancer health disparities in populations of African ancestry, we identified a set of 46,787 African ancestry informative markers (AIMs) using 1000 Genomes Project data from African and European populations. These AIMs were functionally annotated and investigated for potential associations with cancer in the literature, using tools such as VEP, SIFT, and Polyphen. Using multiomics data (RNAseq, genotype, methylation, copy number variation) from The Cancer Genome Atlas and an integrative eQTL modeling approach, we are investigating which of these AIMs act as expression quantitative trait loci in breast cancer. Of the 47K SNPs identified, 0.6% percent was found to be in the coding regions of the genome and AIMS located on chromosomes 8, 10, 11, 15, 18, and 19 were found to be associated with different cancers in the literature including breast and colorectal cancers. Approximately 5K of the AIMs were found in the TCGA breast cancer data and preliminary results identified cis-eqtls in 4 different genes including SERINC2 and ZSCAN23, and trans-eqtls in about 8 genes including CLCN3 and DEDD. These eQTL AIMs are being evaluated for potential association with ancestry-specific differential survival and response to treatment in breast cancer.

Session Title: Cancer Poster Session II

PB5093 Integration of label-free, interpretable image features with spatial molecular profiles from spatial transcriptomics

Authors:

J. Hu[1], Y. Yang[2], J. Huang[1]; [1]Emory Univ., Atlanta, GA, [2]Emory, Decatur, GA

Recent advances in Spatial Transcriptomics (ST) have enabled gene expression profiling while preserving spatial information in tissues. Data from such technologies are typically complemented by histology images of the same tissue section, which are invaluable for examining cellular morphology. The combined gene expression and histological features allow researchers to study the progression of diseases comprehensively. Existing methods for ST data analysis usually employ well-trained deep neural networks to extract features from histology images. Due to the “black box” nature of these methods, these deep features suffer from a lack of biological meanings and are difficult to be integrated with molecular measurements. Novel methods that can extract interpretable, meaningful and robust morphological features are demanded in ST studies.To solve this limitation, we developed a novel approach for Interpretable Morphological features from Unsupervised Segmentation (IMUS). IMUS is fully unsupervised and label-free. Features extracted from IMUS have high interpretability. IMUS starts from unsupervised segmentation. Each cluster from the segmentation is separated as a binary mask, and summary statistics are calculated as features to quantify the pattern in the masks. Next, object detection is performed within each mask, and features for the detected objects are measured. Many features may not be biologically meaningful, e.g., the ones that capture the blurriness of the image. Therefore, IMUS utilized a novel metric to quantify the structural similarity between image features and molecular measurements so that image features with less biological interpretability can be dropped.We have systematically compared IMUS with other popular methods, and IMUS has higher accuracy in segmentation using pathologists’ annotation as ground truth. We also demonstrate that features from IMUS can be integrated with spatial gene expression and identify spatial regions that cannot be revealed when gene expression and histology image is analyzed separately. The interpretability of IMUS’s features also helps to decipher the function of different tissue regions.IMUS is a novel statistical tool to maximumly mine the rich pathology information in ST data. IMUS can be easily applied to studies where training samples are not available and bypass cumbersome labelling steps.

Session Title: Cancer Poster Session III

PB5094 Integration of Large-scale Proteomics and Genomics Data to Discover Risk Proteins in Human Cancers

Authors:

Q. Li[1], Z. Chen[2], J. Ping[2], W. Wen[2], X. Shu[3], J. Yan[4], X-O. Shu[2], Q. Cai[2], J. Long[2], V. Moreno[5,6,7,8], J. Huyghe[9], R. Pai[10], S. Gruber[11], G. Casey[12], U. Peters[9,13], W. Zheng[2], Q. Long[1], X. Guo[2]; [1]Dept. of Biochemistry and Molecular Biology, Univ. of Calgary, Calgary, AB, Canada, [2]Vanderbilt Univ. Med. Ctr., Nashville, TN, [3]Mem. Sloan Kettering Cancer Ctr., New York, NY, [4]Physiology and Pharmacology, Univ. of Calgary, Calgary, AB, Canada, [5]Oncology Data Analytics Program, Catalan Inst. of Oncology (ICO), L'Hosp.et de Llobregat, Barcelona, Spain, [6]Colorectal Cancer Group, ONCOBELL Program, Inst. de Recerca Biomedica de Bellvitge (IDIBELL), L'Hosp.et de Llobregat,, Barcelona, Spain, [7]Dept. of Clinical Sci., Faculty of Med. and Hlth. Sci. and Univ.t de Barcelona Inst. of Complex Systems (UBICS), Univ. of Barcelona (UB), Barcelona, Spain, [8]Consortium for BioMed. Res. in Epidemiology and Publ. Hlth.(CIBERESP), Madrid, Spain, [9]Publ. Hlth.Sci. Div., Fred Hutchinson Cancer Res. Ctr., Seattle, WA, [10]Mayo Clinic Arizona,, Scottsdale, AZ, [11]USC, Los Angeles, CA, [12]Ctr. for Publ. Hlth.Genomics, Univ. of Virginia, Charlottesville, VA, [13]Dept. of Epidemiology, Univ. of Washington Sch. of Publ. Hlth., Seattle, WA

Genomic-wide association studies (GWAS) have identified numerous genetic risk variants associated with various types of cancer. To enhance our understanding of the genetic basis of cancer susceptibility, candidate target genes for the risk variants identified in GWAS have been extensively studied through expression quantitative trait locus (eQTL) analysis. However, the investigation of target proteins for most of GWAS-identified risk variants remains unexplored. In this study, we comprehensively examined previously reported cancer risk variants, and performed additional fine-mapping analyses using summary statistics data from European-ancestry populations from six cancer types: breast (N=247,173), ovary (N=63,347), prostate (N=140,306), colorectum (N=125,478), lung (N=85,716), and pancreas (N=21,536)]. We identified a total of 712 independent risk-related signals for these six cancer types. To identify candidate proteins associated with these risk variants, we conducted a meta-analysis for the results of plasma protein quantitative trait locus (pQTL) analysis from two European-ancestry population studies (N=42,772; PMID: 35501419 and PMID: 34857953). Integrating the findings from these six cancer types, we discovered 259 pQTL associations at a Bonferroni corrected p-value threshold of < 0.05, involving 207 unique proteins associated with 162 risk variants. Among these, 34 proteins were shared by at least two cancer types. Importantly, more than 70% of these 207 proteins had not been previously reported in eQTL analysis or transcriptome-wide association studies in these cancers. Furthermore, over 40% of these proteins were supported by additional evidence based on the analysis of epigenetic data in cancer-related cells and colocalization with GWAS risk signals. Enrichment analyses of these 207 proteins highlighted the prominent involvement of well-established cancer signaling pathways, such as acute-phase responses, IL-6, Natural Killer cell wound healing, and STAT3 signaling pathways. Our study identified a significant number of novel putative susceptibility proteins associated with cancer risk. By shedding light on intricate pathways of genetic variants, target proteins, and signaling pathways related to cancer risk, our findings provide new insights and potential avenues for prevention and therapeutic interventions of these common cancers.

Session Title: Cancer Poster Session I

PB5095 Integrative analysis of RNA-Seq and microarray data for identification of molecular networks and potential biomarkers associated with adverse effects of UVB exposure on the Skin.

Authors:

Y. Jang[1,2], J. Lee[1], J. Han[1], C. Lee[2], H-W. Na[3], Y. Hong[3], H-J. Kim[3], Y. Seo[1]; [1]Dongguk Univ., Goyang-si, Korea, Republic of, [2]The Jackson Lab. for Genomic Med., Farmington, CT, [3]AMOREPACIFIC, Yongin, Korea, Republic of

Ultraviolet B (UVB) from sunlight represents a major environmental factor, which causes toxic effects resulting in structural and functional cutaneous abnormalities of most living organisms. In humans, UVB irradiation acts as a tumor initiator by producing irreversible mutagenic damage. Although numerous studies have indicated the biological mechanisms between UVB exposure and cutaneous manifestations, they have typically originated in a single study, performed under limited conditions. To address this gap, we crawled all publicly accessible expression data on various skin cell types, including skin biopsy data, keratinocytes and fibroblasts exposed to UVB, and we analyzed the biological networks to identify the molecular mechanisms. We identified the three candidate biomarkers (*IL1B*, *CCL2*, and *LIF*) and predicted the inflammatory response and carcinogenesis as major UVB-induced signaling alterations. Furthermore, we confirmed that these three candidate biomarkers contribute to the survival probability of patients with cutaneous melanoma, the most aggressive and lethal form of skin cancer, and that UVB radiation is the main risk factor. Our finding will aid to understand the UVB-induced dermal toxicity and the accompanying molecular mechanisms.

Session Title: Cancer Poster Session II

PB5096 Interactions between breast cancer predisposition genes and known risk factors on breast cancer risk

Authors:

B. Guo[1], A. H. Suger[1], S. Kalia[2], N. Boddicker[3], C. Hu[3], S. Yadav[3], L. R. Teras[4], C. Bodelon[4], J. Lacey[5], C. A. Haiman[6], C. Vachon[3], A. Eliassen[2], D. P. Sandler[7], S. Yao[8], P. Auer[9], A. Trentham-Dietz[10], K. L. Nathanson[11], J. N. Weitzel[12], S. M. Domchek[11], L. Hsu[13], F. J. Couch[3], P. Kraft[2], S. Lindström[1]; [1]Dept. of Epidemiology, Univ. of Washington, Seattle, WA, [2]Dept. of Epidemiology, Harvard Chan Sch. of Publ. Hlth., Boston, MA, [3]Mayo Clinic, Rochester, MN, [4]Dept. of Population Sci., American Cancer Society, Atlanta, GA, [5]Beckman Res. Inst. of City of Hope, Duarte, CA, [6]Keck Sch. of Med., Univ. of Southern California, Los Angeles, CA, [7]Natl. Inst. of Environmental Hlth.Sci., Durham, NC, [8]Roswell Park Comprehensive Cancer Ctr., Buffalo, NY, [9]Med. Coll. of Wisconsin, Milwaukee, WI, [10]Univ. of Wisconsin-Madison, Madison, WI, [11]Univ. of Pennsylvania Sch. of Med., Philadelphia, PA, [12]Natera and Latin American Sch. of Oncology, Sierra Madre, CA, [13]Fred Hutchinson Cancer Ctr., Seattle, WA

Background: Due to the complex interplay between genetic (G) and known reproductive, anthropometric and behavioral ("environmental") (E) risk factors, there is substantial interindividual variability in the risk of developing breast cancer. Previous gene-environment (GE) interaction studies of rare pathogenic variants in breast-cancer predisposition genes have been limited in sample size and the number of interactions assessed.
Methods: We assessed multiplicative and additive GE interactions between five breast cancer predisposition genes (*ATM*, *BRCA1*, *BRCA2*, *CHEK2*, and *PALB2*) and 14 established risk factors (age at menarche, parous, number of births, age at first full-term pregnancy, breastfeeding duration, age at natural menopause, height, pre-and post-menopausal body mass index (BMI), use of menopausal hormone treatment, use of oral contraceptives, history of benign breast disease, smoking and alcohol consumption) in a sample of women drawn from the Cancer Risk Estimates Related to Susceptibility (CARRIERS) Consortium and the UK Biobank, with 28,745 breast cancer cases and 102,997 controls. We assessed both multiplicative and additive interactions using multivariable logistic regression models, adjusting for age at recruitment and study site. Analyses were further stratified by estrogen receptor (ER) status. We used meta-analysis to pool results and assess the heterogeneity of interaction effects between CARRIERS and UK Biobank.
Results: No interaction was significant after adjusting for multiple comparisons ($P<1.2 \times 10^{-4}$) and 28 interactions showed nominal significance at $P<0.05$. Among these, twelve were significant for overall breast cancer, and ten and six were significant for ER+ and ER- breast cancer, respectively. The strongest suggestion of multiplicative interaction was observed for nulliparity and BRCA2 (interaction OR = 3.10, 95% CI: 1.36 - 7.08, P = 7.29 × 10-3), suggesting that those with a pathogenic/likely pathogenic (P/LP) variant in BRCA2 who are nulliparous might have a particularly high risk of breast cancer. For age at first birth and ATM, we observed negative multiplicative (interaction OR = 0.71, 95% CI: 0.54 - 0.93, P = 0.01) and additive interactions (RERI = -0.55, 95% CI: -0.98 - -0.12, P = 0.01), suggesting that the risk increase associated with an older age at first full-term pregnancy is attenuated among women with a P/LP variant in ATM.
Conclusions: The associations between P/LP variants in known predisposition genes and breast cancer may be modified by reproductive factors. Larger studies encompassing diverse populations are needed to confirm these findings.

Session Title: Cancer Poster Session III

PB5097 Interlaboratory assessment of NIST test materials for lentiviral vector copy number and integration site measurements

Authors:

H-J. He[1], Z. He[1], A. Whale[2], C. Foy[2], A. Li[3], B. Schroeder[3], E. Stelloo[4], H. Feitsma[4], G. Dijksteel[4], L. Turner[5], Y. Fan[5], B. Dropulic[6], O. Slessareva[6], J. Kandell[7], U. Lakshmipathy[7], P. Welch[8], S. X. Senderovich[9], D. Steenmans[10], M. Raimo[10], W. Vogels[10], Y. Ding[11], M. Sheldon[11], S. Ahmed-Seghir[12], S. Miller[1], S. Lund[1], S. Maragh[1], J. Zook[1], J. Elliott[1], S. Lin-Gibson[1], K. Cole[1]; [1]Natl. Inst. of Satndards and Technology, Gaithersburg, MD, [2]Molecular and Cell Biology, Natl. Measurement Lab., LGC, Teddington, United Kingdom, [3]Mission Bio, South San Francisco, CA, [4]Cergentis, Utrecht, Netherlands, [5]Azenta, South Plainfield, NJ, [6]Caring Cross, Gaithersburg, MD, [7]Thermo Fisher Scientific, San Diego, CA, [8]Castle Creek Biosci, Exton, PA, [9]GSK, Stevengage, United Kingdom, [10]Glycostem Therapeutics, AB Oss, Netherlands, [11]Sampled, Piscataway, NJ, [12]Genomic Vision, Bagneux, France

Lentiviral vectors (LV) have emerged as powerful tools for stable gene delivery for both cell and gene therapies. As an effective measure to reduce the genotoxic and tumorigenic potential caused by uncontrolled integration, the US Food and Drug Administration recommends the integrated lentiviral vector copy number (VCN) shall be less than 5 copies per cell. Reference or control materials are essential for accurately measuring VCN, a critical quality attribute to the safety and efficacy of gene modified cellular products. We conducted an interlaboratory assessment to determine the utility and suitability of NIST test materials to serve as reference materials or controls for the VCN measurements. The test materials comprise of five human genomic DNAs extracted from clonal cell lines with N=1 (VCN1), N=2 (VCN2), N=3 (VCN3), and N=4 (VCN4) copies of the integrated provirus and the parental Jurkat cell line (VCN0), or fixed cell samples from these clonal cell lines. A total of twelve laboratories from four countries participated in this study, composed of two National Metrology laboratories and ten companies. Quantitative real-time PCR (qPCR; four laboratories), digital PCR (dPCR; seven laboratories using droplet, two of them also using chamber dPCR), Next Generation Sequencing (NGS; targeted amplicon NGS, targeted locus amplification NGS, and single cell NGS), and molecular combing methods were used to make VCN measurements. All laboratories were able to identify the five blinded VCN samples by either qPCR, dPCR or NGS assay. In all cases no integrated LV was identified in the VCN0 sample. Using qPCR and dPCR, VCN1-3 samples were identified correctly within 10.2% of the expected VCN. Greater variability in the estimates for the VCN4 sample between the laboratories using both qPCR and dPCR. All three NGS methods reported correct copies and positions of integration upon aligning on the same reference genome build. Interestingly, one integration site in the VCN4 sample reported by all three methods was different from the site originally published, but was confirmed to be correct by additional whole genome sequencing analyses. The consensus integration sites were achieved after the optimization of assays and bioinformatics pipelines. The cell-based test material was also used to evaluate the assay performance of a new direct visualization technology called molecular combing. This interlaboratory assessment supports the utility and suitability of the NIST VCN test materials in the analytical parameters of different VCN measurement methods to ensure quality assurance.

Session Title: Cancer Poster Session I

PB5098 Interplay of germline risk loci, somatic variants, and structural alterations in the development of pancreatic cancer

Authors:

J. Levine, L. Ali, K. McKenna, S. Finkelstein, V. Timmaraju; Interpace Dx, Pittsburg, PA

Germline variants can predispose an individual to an increased risk of pancreatic cancer, a disease with the 2nd highest mortality rate in the US. However, with a low incidence of occurrence compared to its mortality rate, understanding the germline mutations' role in concert with acquiring somatic mutations is critical to early detection and diagnosis, allowing intervention. To address this, we developed a hybrid capture-based next-generation sequencing method to detect: germline mutations in key risk loci, somatic single nucleotide variants, insertion-deletions in cognized PC oncogenes, and tumor suppressor genes, genome-wide copy number analysis, and repeat expansion alterations. We compared precancerous pancreatic cysts (n=66) and pancreatic cancer (n=22) to elucidate the presence of variants in pancreatic cancer germline risk loci and acquired somatic alterations. Most notably, we discovered a change in heterozygosity of the *TP53_P72R* germline variant, which, when combined with a TP53 oncogenic mutation, co-occurred with multiple structural variants in both cohorts. These results further support germline mutation screening to be combined with somatic variant detection for early diagnosis and accurate risk assessment for both the primary patient and first-degree relatives.

Session Title: Cancer Poster Session II

PB5099 Interrogating immuno-oncological interactions in the tumor microenvironment

Authors:

C-Y. Chen, B. Patterson, N. Fernandez, S. Kaushal, L. Colbert, J. He; Vizgen, Cambridge, MA

Single-cell RNA sequencing is a powerful tool for assessing cancer heterogeneity. However, single-cell RNA-seq inevitably destroys the spatial information within tumor microenvironments, preventing the study of the complex tumor niche interactome during disease progression. The Vizgen MERSCOPE® Platform is built on Multiplexed Error-Robust Fluorescence in situ Hybridization (MERFISH) technology and enables direct molecular and cell atlasing in intact tissue with subcellular resolution. Here, we used MERSCOPE to probe immuno-oncological interactions in patient-derived cancer tissues. Using a custom 500-gene panel designed to assess immune cell states, canonical cancer signaling pathways and immune-oncological interactions occurring within the tumor microenvironment, we spatially profiled the gene expression of >1 million cells across human lung cancer samples. The MERFISH data showed strong correlation when compared with block-matched RNA-seq data, demonstrating the high accuracy of our data capture. To map the cell types and characterize their molecular signatures within the tumor microenvironments, we employed a deep learning-based segmentation algorithm, and resolved distinct gene expression patterns and spatial locations in fibroblasts, endothelial cells, and multiple cancer cell populations within the 3-dimensional tissue context of the tumor tissue. The expression of CXCL9, CXCL10, and CXCL11 were used to identify immune hubs and neighboring cell types which can provide key spatial insights into potential immuno-oncological interactions occurring across these lung cancer samples. We quantified the overlap of various cell types in immune neighborhoods and found that macrophages and CD4+ T cells spatially overlap with immune hubs, while B cells and monocytes have low spatial overlap with the hubs; these differences suggest a proximal-to-distal polarization of the tumor interactome. Lastly, we identified three major cancer clusters with distinct spatial distributions and relationships with epithelial cells, which may provide further insights into tumorigenesis and metastasis. These findings demonstrate how the MERSCOPE Platform is a powerful research tool providing insights into the complex immuno-oncological microenvironments in tumor samples.

Session Title: Cancer Poster Session III

PB5100 Intersecting predisposition germline variants from TCGA with the *All of Us* Research Program reveals the importance of diverse cancer datasets.

Authors:

B. Bates, K. Bates, S. Boris, M. Davis, M. Bailey; Brigham Young Univ., Provo, UT

The last two decades of cancer research have benefited tremendously from the advent of next-generation sequence technologies. The Cancer Genome Atlas (TCGA) has led the way in this genomics revolution by molecularly characterizing 10,000 tumors from 33 different cancer types. Among the biomarkers discovered by TCGA are cancer predisposition mutations, or germline mutations that contribute to cancer development. Despite this large effort, TCGA is a case set, which is heavily biased toward cancer selection and is primarily composed of participants of European genetic ancestry. Consequently, while many predisposition germline variants have been identified in TCGA, their impact on diverse populations remains understudied. The All of Us Research Program provides an excellent dataset for validating the effect of the predisposition variants from TCGA in a cohort that is both more diverse and unbiased by cancer selection. Using the All of Us Controlled Tier Dataset v6, which contains 70,666 people with both whole-genome sequences and insurance billing code data, we investigated the connection between predisposing germline variants and cancer occurrence across 6 genetic ancestries: European (EUR), African (AFR), Admixed American (AMR), East Asian (EAS), South Asian (SAS), and Middle Eastern (MID). We identified 22,372 patients with billing codes suggesting a cancer-related phenotype (15,925 EUR, 3166 AFR, 2639 AMR, 362 EAS, 189 SAS, and 91 MID), making the All of Us Research Program a meaningful cancer dataset. We validated established predisposition variants in EUR participants, showing that the presence of predisposition variants was associated with higher frequency of cancer occurrence ($p < 0.001$, relative risk = 1.12). However, this result did not replicate in any other genetic ancestry. Additionally, the presence of predisposition variants did not correlate to earlier age at onset in any genetic ancestry. Further analyses establish the relationship between predisposition variants and other clinically available data. By embracing diversity and harnessing the power of comprehensive datasets, we illuminate the interplay between genetics, ancestry, and cancer risk.

Session Title: Cancer Poster Session I

PB5101 Intra-tumor heterogeneity of HPV integration and its association with focal genomic instability in oropharyngeal cancer.

Authors:

N. Sasa[1], H. Eguchi[1], H. Tanaka[1], T. Fukusumi[1], M. Suzuki[1], Y. Takenaka[1], Y. Okada[2,3,4,5], H. Inohara[1]; [1]Osaka Univ. Graduate Sch. of Med., Suita, Japan, [2]Osaka Univ. Graduate Sch. of Med., Osaka, Japan, [3]Graduate Sch. of Med., The Univ. of Tokyo, Tokyo, Japan, [4]RIKEN Ctr. for Integrative Med. Sci., Yokohama, Japan, [5]Immunology Frontier Res. Ctr. (WPI-IFReC), Osaka Univ., Suita, Japan

Human papillomavirus (HPV) integration has been considered as one of the driver factors for cancer progression, but its pattern and role in carcinogenesis is still unclear. Whole genome sequencing data of 14 Japanese HPV16+ OPSCC and 13 HPV+ OPSCC from Pan-Cancer Analysis of Whole Genomes were used to identify integrations and somatic mutations. A total of 254 integration breakpoints (BPs) in 20 of the 27 tumors (74.1%) were detected. We estimated the cancer cell fraction of integration BPs and found that HPV integration also shows intra-tumor heterogeneity, indicating that nearly half of the integration events occurred after carcinogenesis. Clonal BPs were more likely to occur in the *E1* ($P = 0.036$), confirming that the disruption of the *E1* gene may result in abnormal negative regulation of the *E6/E7* oncogenes and promote OPSCC carcinogenesis. Subclonal BPs were less likely to occur in the *E6* ($P = 0.045$), suggesting that the *E6* may also play an important role in the process of cancer evolution after carcinogenesis. Four states of the HPV genome were identified: (1) episomal-only (7/27, 25.9%), (2) integrated-only (4/27, 14.8%), (3) clonally-mixed (episomal + clonally integrated) (11/27, 40.7%), and (4) subclonally-mixed (episomal + subclonally integrated) (5/27, 18.5%). Since nearly half of these cancers have developed with just episomal copies of HPV, integration itself is not essential for carcinogenesis and can occur during and after carcinogenesis. Interestingly, the *E6/E7* were conserved in all four cancers with integrated-only HPV, while the *E2* and *E1* were disrupted or deleted. This confirms that the constitutive expression of *E6/E7* is essential for the carcinogenesis of HPV+ OPSCC. These results suggest that some integrations may be drivers of carcinogenesis with disruption of *E1/E2*, as in these four cancers, and others may arise randomly due to genomic instability, as in the cancers with the *E6/E7* deleted integrants. The number of BPs of structural variations (SVs) occurred in the overlapping 100 kb regions flanking integration BPs (ITG regions) was positively correlated with the number of integration BPs (Spearman's rank correlation $\rho = 0.78$; Permutation test $P = 9.9 \times 10^{-5}$). The genomic instability that causes clonal integrations during carcinogenesis is considered to cause clonal SVs, mainly large duplications, and subclonal integrations and SVs such as large deletions. Clonal SVs, mainly large duplications, were also observed in the regions where only subclonal integrations occurred, suggesting that large duplications occur before integrations.

Session Title: Cancer Poster Session II

PB5102 Investigating cell-type specific regulation of *IRF4* via a lung cancer risk-associated pleiotropic variant

Authors:

J. Yin[1], A. Kane[1], E. Long[2], J. Feng[1], C. Gräwe[3], H. Patel[1], J. Byun[4], C. Amos[4], K. Brown[1], J. Xia[5], M. Vermeulen[3], J. Choi[1]; [1]Natl. Cancer Inst./NIH, Bethesda, MD, [2]Chinese Academy of Med. Sci. and Peking Union Med. Coll., Beijing, China, [3]Radboud Inst. for Molecular Life Sci., Oncode Inst., Radboud Univ., Nijmegen, Netherlands, [4]Baylor Coll. of Med., Houston, TX, [5]5Dept. of BioMed. Sci., Creighton Univ., Omaha, NE

A cross-ancestry genome-wide association study (GWAS) in European, East Asian and African populations have identified a locus associated with lung cancer risk at 6p25.3 tagged by rs12203592. Subsequent eQTL colocalization followed by cell-based DNA damage assays identified *IRF4* as a plausible susceptibility gene in this locus. rs12203592 was previously associated with other traits, including pigmentation traits, melanoma and other skin cancers, blood molecular traits, and smoking cessation. Importantly, GTEx and primary melanocyte eQTL (n = 106) data indicated that rs12203592 is an eQTL in multiple tissue types but in opposite allelic directions. Namely, the lung cancer risk-associated T allele is correlated with higher *IRF4* expression in lung and whole blood tissues but lower expression in skin tissues and melanocytes. Pairwise cross-tissue eQTL colocalization identified rs12203592 as the common causal variant of *IRF4* eQTLs between lung tissue and melanocyte, skin, and whole blood (colocalization posterior probability > 0.8 for all three). rs12203592 was previously shown to drive an allelic enhancer function in melanocyte and blood cell lines, and the pigmentation trait-associated C allele binds TFAP2A together with a lineage-specific regulator. To further characterize the context-specific role of rs12203592 in *IRF4* gene expression we first identified allele-specific binding proteins by mass spectrometry using the nuclear extracts from three cancer cell lines representing lung, melanocyte, or lymphocyte lineage. Among them, the NF-kB family proteins were prioritized as potential regulators of *IRF4* expression through rs12203592-T risk allele in A549 lung cell line. We further confirmed the allele-dependent binding of rs12203592 to NF-kB2 recombinant protein using electromobility shift assays, as well as antibody-supershift using A549 nuclear extract. Next, we performed expression-genotype interaction analyses using GTEx and melanocyte eQTL data. Consistent with the previous report of NF-kB pathways regulating *IRF4* in blood cell types, most of the *NF-κB* family members displayed a correlation with *IRF4* expression in lung, skin, blood, and melanocyte. However, their interaction with rs12203592 was mainly observed in lung tissues ($P < 0.025$ for *NF-kB2*, *RELB*, and *cREL*). To validate our observation and demonstrate the molecular interaction between NF-kB and *IRF4* through rs12203592, we are performing chromatin immunoprecipitation under stimulation of NF-kB proteins in lung cells. In summary, we identified NF-kB pathway proteins as potential transcriptional regulators of *IRF4* expression via rs12203592 in a context-specific manner.

Session Title: Cancer Poster Session III

PB5103 Investigating the relationship between sleep-wake characteristics and risk of colorectal cancer: A Mendelian Randomization approach.

Authors:

K. Santiago[1], T. Rogne[1], N. S. Redeker[2], X. Llor[3], A. Dewan[1]; [1]Yale Sch. of Publ. Hlth., New Haven, CT, [2]Sch. of Nursing, Univ. of Connecticut, Storrs, CT, [3]Yale Sch. of Med., New Haven, CT

Colorectal cancer is the third most diagnosed malignancy and the estimated second leading cause of cancer-related death in the United States. There are several well established modifiable risk factors for colorectal cancer (CRC), however, few studies have examined whether sleep-wake characteristics are associated with CRC. In observational studies, long sleep duration has been associated with an increased risk of CRC, while evidence suggests an inverse association of morning chronotype (i.e., individuals known as early risers) with CRC. To further investigate these relationships, we conducted Mendelian Randomization (MR) analyses to assess the effect of insomnia, chronotype, short sleep duration, and long sleep duration on CRC risk. A total of 536 independent SNPs associated with insomnia at the genome-wide significance level ($P < 5 \times 10\text{-}8$) were used as instrumental variables from a genome-wide meta-analysis of 2,365,010 individuals. For chronotype, a total of 148 SNPs from 403,195 adults of European ancestry from the UK Biobank were used as instrumental variables. Lastly, a total of 25 SNPs associated with short sleep duration and 8 SNPs associated with long sleep duration from 446,118 UK Biobank participants of European ancestry served as instrumental variables. Summary-level data on colorectal cancer was obtained from FinnGen (5,458 cases) and a meta-analysis of CRC from subjects of both European and Asian ancestry (100, 204 cases). In the meta-analysis of the two sources, genetic liability to short sleep duration (<7 hours) was associated with an increased risk of colorectal cancer (OR 4.62, 95% CI: 3.42-5.82). Consistent with extant literature, morning chronotype was associated with a decreased risk of CRC (OR 0.92, 95% CI: 0.82-1.03). In sex-specific analyses, genetic liability for insomnia in females was associated with a significantly increased risk for CRC (OR 1.62, 95%CI: 1.06-2.18), while the genetic effect of insomnia in men did not result in a significantly increased risk (OR 1.49, 95%CI: 0.98-1.99). Evidence from these MR analyses support the findings that morning chronotype reduces risk of CRC and support the hypothesis that chrono-disruption, in this case short sleep duration, significantly increases CRC risk. Given these preliminary results, future directions include multivariable MR analyses to adjust for potential confounding and an exploration of mediating pathways such as that of the metabolome. This study highlights the importance of designing interventions that promote healthy sleep behaviors such as limiting screen time and promoting a regular sleep schedule to reduce the risk of CRC.

Session Title: Cancer Poster Session I

PB5104 Investigating vitreous cytokines in posterior uveal melanoma in relation to tumor occurrence, size parameters, and gene expression profiling-based prognostication.

Authors:

Y. Demirci[1], L. Tang[1], C. Ozgonul[2], S. Weber[3], J. Sundstrom[3], H. Demirci[2]; [1]Univ. of Pittsburgh, Pittsburgh, PA, [2]Univ. of Michigan, Ann Arbor, MI, [3]Penn State Univ., Hershey, PA

Uveal melanoma (UM) is a rare but aggressive cancer with a high rate of metastasis. Metastatic risk prediction in UM is currently done by using clinical information, and when available, tumor biopsy-based information. While tumor biopsy-based molecular analysis (i.e., chromosomal aberration, driver mutation, or gene expression profiling) is known to improve UM prognostication, a search is underway for a liquid biopsy-based approach (e.g., ocular fluid-based molecular analysis) that can overcome the limitations of direct tumor biopsy. Vitreous humor fills the posterior eye cavity, and its composition can reflect the state of posterior eye layers. Given that UMs predominantly occur in posterior eye segment (>90% in choroid), vitreous is expected to be enriched in tumor-derived molecules because of its proximity to these posterior tumors. Molecular analysis of the vitreous may thus provide useful insights into UM biology and inform new clinical interventions. Inflammatory tumor microenvironment (TME) is associated with poor prognosis in UM, and cytokines constitute key soluble factors/molecules of the TME. In this study, we aimed to gain further insights into the TME in UM arising from the choroid (CM) by examining the vitreous cytokines in 32 eyes, including 18 with CM (prior to plaque radiotherapy or following enucleation) and 14 without CM. We employed a 42-plex cytometric bead immunoassay to measure vitreous cytokines and used median fluorescence intensity values as relative quantification of cytokine abundance. We used Wilcoxon rank sum test to compare vitreous cytokine levels between CM and non-CM groups and across different prognostic categories within the CM group (classified as having high or low metastatic risk using tumor biopsy-based gene expression profiling (GEP)). We used Spearman's rank correlation to assess the relation between vitreous cytokine levels and tumor size parameters. We identified twenty-six vitreous cytokines significantly upregulated in CM-bearing eyes compared to CM-free eyes. Within the CM group, we detected six differentially expressed vitreous cytokines in eyes with GEP Class 2 (high risk) vs. Class 1 (low risk) tumors. Principal component analysis of these six vitreous cytokines suggested a potential separation between the CM subgroups representing different tumor GEP-based prognostic categories. The vitreous levels of these six plus multiple other cytokines showed correlations with the tumor size. In summary, we identified several UM-relevant vitreous cytokines, which may serve as new therapeutic targets or biofluid-based prognostic biomarkers upon further validation in larger and longitudinal studies.

Session Title: Cancer Poster Session II

PB5105 Investigation of cancer related genes in patients with Ollier disease and Maffucci syndrome

Authors:

I. Babler-Madrid[1], R. Paulo Martin[2], E. Wohler[3], N. Sobreira[4]; [1]Johns Hopkins Univ., Orlando, FL, [2]Johns Hopkins Med., Towson, MD, [3]Johns Hopkins Univ., Baltimore, MD, [4]Johns Hopkins Univ, Crownsville, MD

Ollier disease (OD) and Maffucci syndrome (MS) are newly recognized cancer susceptibility syndromes. Our recent work shows that ~50% of patients with these syndromes develop a malignancy. Most commonly chondrosarcomas (~30%), gliomas (~10%), and, in patients with MS, vascular malignancies (~7%). Additional shared clinical features include enchondromas with resulting bone deformities, and other skeletal abnormalities. In addition, patients with MS have vascular overgrowth anomalies. Here we present an analysis based on a burden test using a multisample VCF file containing whole genome sequencing (WGS) data from 639 samples (73 cases and 566 controls), and whole exome sequencing (WES) data from 1855 samples (55 cases and 1800 controls). FASTQ files were aligned to the reference genome (GRCh38) with the Burrows-Wheeler Alignment (BWA 0.5.10) resulting in SAM/BAM output. GATK best practices were applied in order to obtain the multisample VCF file. First, variants were filtered based on the gene symbols of 1202 genes described in the GeneDx oncology panels and COSMIC (cancer gene census). Only single nucleotide variants (SNVs) were analyzed, and the final multisample VCF was annotated with ANNOVAR. SNVs were filtered based on RefSeq and gnomAD annotations to include functional (missense, nonsense, stop-loss, and splicing variants) rare (MAF $\leq$ 0.01) SNVs. A contingency 2x2 table containing the sum of individuals presenting at least one functional rare SNV was built for each gene mutated in at least 5 cases. 135 genes were mutated in 5 or more cases. Out of the 135 genes, three were statistically significant based on a Fisher's exact test where $p<0.05$ was considered statistically significant. The variants in these three genes will be further evaluated based on their gnomAD frequency, familial segregation, variant type, in silico prediction, ClinVar and HGMD classification, and ACMG classification to determine if any of these genes should be further investigated functionally. Additionally, the strongest candidate genes will be further investigated in unrelated cohorts of patients with isolated gliomas, chondrosarcomas, osteosarcomas and Ewing's sarcoma to determine if they are also involved in these malignancies.

Session Title: Cancer Poster Session III

PB5106 Investigation of the effect of FCRL gene family expression levels on prognosis in chronic lymphocytic leukemia patients.

Authors:

K. Gündüz[1], E. Coskunpinar[1], S. Erdem[2], M. Yenerel[2]; [1]Univ. of Hlth.Sci., Istanbul, Turkey, [2]Istanbul Univ., Istanbul, Turkey

Background: Chronic lymphocytic leukemia (CLL) is a malignancy of CD5+ B cells characterized by the accumulation of small, mature-appearing neoplastic lymphocytes in the blood, bone marrow, and secondary lymphoid tissues, resulting in lymphocytosis, leukemia cell infiltration in the bone marrow, lymphadenopathy, and splenomegaly. Although the disease shows a heterogeneous clinical course, some of the diagnosed patients are asymptomatic and live for many years without progression, while some show rapid progression even in the early stages. Therefore, it is crucial to predict the prognosis of the disease at the time of diagnosis to identify high-risk patients. *FCRL*, called the Fc receptor-like protein family, encodes type 1 transmembrane proteins with tyrosine-based signaling properties expressed by B cells. It is thought that *FCRL* family members may function to regulate innate and adaptive humoral immunity by positively and/or negatively affecting B-cell signaling. **Aim:** To reveal candidate biomarkers that may be effective in predicting the prognosis of the disease by examining immunophenotypic markers, chromosomal abnormalities, and the expression levels of *FCRL* gene family members (*FCRL1*, *FCRL2*, *FCRL3*, *FCRL4*, *FCRL5*) in CLL patients. **Methods:** 48 patients diagnosed with CLL between November 2021 and June 2023 were included in our study. Immunophenotypic parameters by flow cytometry and chromosomal abnormalities by FISH method (del13q14, del11q22, trisomy12, del17p13, del14q32 and del6q23) were analyzed from the peripheral blood samples. *FCRL* gene family members (*FCRL1*, *FCRL2*, *FCRL3*, *FCRL4*, *FCRL5*) and *ZAP70* gene expression levels were determined by qRT-PCR using RNA samples isolated from whole blood. **Results:** It was determined that there was a difference between the two groups in the fold regulation rates of *FCRL2* and *FCRL5* genes, but there was no statistical significance between the groups (respectively; FR-*FCRL2*= 2.23, p-*FCRL2*= 0.196498; FR-*FCRL5*= 2.41, p-*FCRL5* = 0.156453). When the expression of these genes in normal tissues is examined, it is known that they are mostly expressed in lymph nodes and both genes play a role in the cell surface signaling receptor pathway. **Conclusion:** It is thought that *FCRL2* and *FCRL5* genes, which are in the cell surface signal receptor pathway, can be considered as good prognostic markers in CLL.

Session Title: Cancer Poster Session I

PB5107 Karyotype prediction in pediatric acute leukemias using long-read sequencing.

Authors:

J. Lin, T. Alexander, C. Jones, J. Wang; Univ. of North Carolina at Chapel Hill, Chapel Hill, NC

Throughout the past few decades, the 5-year survival rate for pediatric cancers in high-income countries (HICs) has greatly increased in comparison to that of low-to-middle-income countries (LMICs). This disparity can be attributed to advances in cancer detection and treatment more readily available in HICs. Acute lymphoblastic leukemia (ALL) accounts for most pediatric cancers and diagnosis-dependent risk stratification greatly affects clinical prognosis. While short-read sequencing technologies have proven useful in diagnosing genomic subtypes of pediatric leukemias, long-read sequencing (specifically, Oxford Nanopore Technologies) has the benefits of relative affordability, flexible implementation, and better detection of genomic structural variations. Therefore, Nanopore sequencing may be a viable diagnostic resource in LMICs to subtype ALL. Our previous work demonstrated that gene expression profiles derived from whole transcriptomic sequencing (using Nanopore RNA-Seq) can be used to classify major lineages of leukemia (B-ALL vs T-ALL vs AML) and clinically relevant genomic subtypes of B-ALL such as *ETV6-RUNX1* and *BCR-ABL1*. Building on these efforts, we show there is the potential to additionally infer aneuploid subtypes. From a cohort of 137 ALL samples, we show that Nanopore RNA-seq shows signals consistent with known karyotypes. Our results show minor allele frequency (MAF) distributions with qualitative peak shifts correlated with chromosomal copy number variation. Additionally, gene expression distributions on a chromosome level show elevations and decreases consistent with copy number gains and losses, respectively. Using features derived from MAF distributions and gene expression levels per chromosome, we tested various classification models (from python's scikit-learn machine learning library) to predict chromosomal copy number in our cohort. Initial results show a best overall accuracy of 86%, albeit biased toward overrepresented diploid classifications. By testing our model exclusively on aneuploids and collating results at the sample level, we see a best overall accuracy of 86% with better discrimination between ploidy (near haploid, low hypodiploid, diploid, and hyperdiploid). Ongoing challenges stem from low coverage, error rate and class imbalance in our dataset. Despite these challenges, our work shows Nanopore RNA-seq has diagnostic potential in the hope of developing diagnostic, point-of-care tests for pediatric patients in LMICs.

Session Title: Cancer Poster Session II

PB5108 Learning drug resistance mechanisms enables AI to generate rational anti-cancer combination therapies

Authors:

M. Sherman[1], L. McDaniel[1], O. Priebe[1], E. Farias[1], T. Ideker[2], A. Yaari[1]; [1]Serinus BioSci.s Inc, New York, NY, [2]Univ California, San Diego, San Diego, CA

Administering therapies in combination improves survival of cancer patients, but combination therapies are challenging to develop due to the complexity of identifying clinically effective and non-toxic drug pairings out of the billions of possibilities. We designed a deep learning framework to generate rational combinations by 1) learning clinically relevant genomic mechanisms of resistance for a drug of interest and then 2) identifying existing compounds that will counteract these resistance processes. The key innovation is the model's architecture that, by mimicking tumor cell proteomic organization, learns an embedding space where each dimension reflects an interpretable relationship between a cellular process and drug response, conditioned on a tumor cell's genetic, epigenetic, and/or transcriptomic state.

.

We applied the framework to learn genetic correlates of response and resistance for 1,950 drugs and 17,386 CRISPR gene knockouts screened in human cancer cell lines. The learned embeddings reflected known and novel gene-drug relationships. For example, *RAF* inhibitors (N=6), *MEK* inhibitors (N=3), and *BRAF* CRISPR knockout colocalized ($P=1.7\times10^{-11}$) due to their dependency on hyperactivation of RAF signaling; conversely, the *CDK4/6* inhibitor palbociclib shared key genetic dependencies with *CDK6* and *CDK2* loss-of-function but, surprisingly, not *CDK4* loss-of-function. To demonstrate clinical applicability of the framework, we retrospectively predicted response of 70 ER+ metastatic breast cancer patients treated with palbociclib based on the learned resistance embedding. Patients with predicted response in the bottom standard deviation (N=13) had significantly shorter treatment duration than patients in the top standard deviation (N=9) (15.2 months vs 21.7 months, log-rank test P=0.05) with no progression free survival benefit relative to patients not treated with palbociclib (P=0.35).

.

Finally, we implemented a zero-shot learning paradigm to score drug pairings for potential synergy based on their learned genetic resistance embeddings. The approach recovered clinically successful combinations such as RAF+MEK inhibitors, and in a systematic evaluation of 986 previously screened combinations, predicted synergy scores were significantly correlated with experimentally derived scores (Spearman correlation $P=6.6\times10^{-4}$). The framework has proposed dozens of new combinations that are now being experimentally tested. These results highlight the promise of biologically tailored deep learning to learn complex cellular processes and accelerate development of anti-cancer combination therapies.

Session Title: Cancer Poster Session III

PB5109 Leveraging polygenic scores to reveal the interplay of serum bilirubin, smoking, and cancer risk in a diverse Los Angeles biobank

Authors:

V. Venkateswaran[1], A. Bhattacharya[2], K. Boulier[1], E. Petter[1], Y. Ding[1], B. Pasaniuc[1]; [1]UCLA, Los Angeles, CA, [2]Univ. of California, Los Angeles, Los Angeles, CA

Recent studies identify serum bilirubin (SB) as a metabolic hormone with potent antioxidant effects, linking low SB levels with cancers, and metabolic and cardiovascular diseases. Additionally, tobacco smoking is reported to contribute to low SB levels. Thus, the associations between SB and cancers are theorized to be secondary to interactions with tobacco smoking. Using electronic health records data (EHR) on ~400,000 individuals and a polygenic score for SB on ~60,000 individuals, we examine the interplay of SB with head and neck cancer (HNC) and tobacco smoking within the UCLA ATLAS biobank, a diverse EHR-linked biobank with extensive de-identified phenotypic and demographic information.

We find that SB is inversely correlated with smoking (Linear coefficient: -0.02, CI [-0.0250, -0.0152]) and HNC (Linear coefficient: -0.16, CI [-0.19, -0.13]) after adjusting for age, sex, and self-identified race and ethnicity. Further, in a group of propensity score-matched HNC cases and controls (2040 cases and controls), matched on patient age, sex, smoking history, and self-identified race and ethnicity, we find a similar inverse association with bilirubin (Linear coefficient: -0.16, CI[-0.18,-0.12]).

Lastly, we used a polygenic score (PGS) for ‘total bilirubin’ from the PGS catalog (PGS002160) as a genetic fixed point to identify the direction of the associations between SB, smoking, and HNC. We imputed the PGS in European genetic ancestry individuals in ATLAS and validated the predictive ability against observed total bilirubin from the EHR (Linear coefficient: 0.22, CI[0.21, 0.23]). Next, we tested the association of the validated bilirubin PGS with smoking history and HNC respectively, finding no significant associations (Linear coefficients: -0.004, CI [-0.0116,0.0035] and 0.0002, CI [-0.0014, 0.0018], respectively).

This study is the first reported evaluation of the associations between SB, smoking, and HNC. Our results suggest that low SB is likely secondary to HNC or a common unidentified factor that influences control over both HNC risk and SB and is independent of smoking.

Session Title: Cancer Poster Session I

PB5110 Long-read sequencing of cervical cancer reveals the structure of DNA ends during Breakage-Fusion-Bridge events

Authors:

M. Dean[1], I. Rodriguez[2], N. M. Rossi[2], A. Keskus[3], Y. Xie[2], T. Ahmad[4], H. Lou[5], J. Godinez Paredes[2], V. Mukhina[6], D. Gaykalova[6], L. Malik[7], K. Billingsley[7], C. Blauwendraat[7], M. Kolmogorov[7]; [1]NCI, Frederick, MD, [2]NCI, Rockville, MD, [3]NIH, BETHESDA, MD, [4]NCI, Bethesda, MD, [5]NCI - Frederick, Rockville, MD, [6]Univ MD Sch. of Med., Baltimore, MD, [7]NIH, Bethesda, MD

Cervical cancer is the most common cause of cancer death in low-resource countries, and there are no targeted therapeutic approaches. We characterized 22 cervical cancer cell lines using long-read DNA and RNA sequencing to provide improved models. We identified the human papillomavirus (HPV) types, HPV integration sites, and cancer driver mutations. Cervical cancer cell lines have recurrent chromosomal alterations, and structural variation analysis revealed telomeric deletions associated with DNA inversions (breakage-fusion-bridge (BFB) cycles). BFB events are a common mechanism of chromosomal alteration in cancer, but, to our knowledge, this is one of the first analyses of BFB events using long-read sequencing. Analysis of the inversion sites revealed staggered ends consistent with exonuclease deletion of the DNA after breakage. Some BFB events are complex, involving the insertion of DNA from another chromosome or local rearrangements. BFB events are thought to resolve by adding a telomere at the chromosome break, but we did not find telomere sequences at any BFB site. BFB events have been associated with chromothripsis; however, we only observed this in 1/12 BFB breakpoints. Three cell lines have a chr11q BFB event, with *YAP1/BIRC2/BIRC3* gene amplification. *YAP1* amplification is associated with a 10-year earlier age of diagnosis and is three times more common in African Americans. The major molecular subtypes of cervical cancer are represented in this panel and will allow the development of targeted therapies. In summary, using long-read sequencing, our data provide new insight into a critical type of chromosomal alteration in cancer.

Session Title: Cancer Poster Session II

PB5111 Lung cancer in ever- and never-smokers: Findings from multi-population GWAS studies.

Authors:

Y. Li[1], X. Xiao[1], J-R. Li[1], G. Fernandes[1], M. Zhu[2], Y. Han[1], J. Xia[3], C. Cheng[1], C. Amos[1], INTEGRAL-ILCCO Lung Cancer Consortium; [1]Baylor Coll. of Med., Houston, TX, [2]Nanjing Med. Univ., Nanjing, China, [3]Creighton Univ., Omaha, NE

Clinical, molecular, and genetic epidemiology studies displayed remarkable differences between ever- and never-smoking lung cancer. Association analysis stratified by smoking behavior has the potential to identify novel variants that confer risk in only ever- or never-smoking group and were missed by prior main-effect association studies. We conducted a stratified multi-population (European, East Asian and African-American) association study in 44,823 ever-smokers and 20,074 never-smokers. Five independent novel loci, including *GABRA4* from ever-smoking and *LRRC4C* and *LCNL1* from never-smoking lung cancer, were identified with association evidence from two or three populations ($P < 5x10^{-8}$). In addition, we also validated the lung cancer risk effect for known variants at *VTI1A* and *ACVR1B* in never-smoking African American women. Further functional analysis provided multiple lines of evidence suggesting the variants affect lung cancer risk through excessive DNA damage (*GABRA4*) or cis-regulation of gene expression (*LCNL1*). For the variants with association effect in ever-smoking lung cancer (including the known variants), we examined their risk effect among never-, light- (packyear <= 20) and moderate-to-heavy- (MtoH) (packyear > 20) smokers with European ancestry. 9 out of 12 independent variants had increased lung cancer risk in MtoH-smokers compared with light-smokers. We also observed high correlation (79.19%, P=0.01) between ever- and never-smoking lung cancer suggesting the common variants shared by them. The high heritability (9.80%) of ever-smoking lung cancer suggests certain signaling pathways induced by smoking behavior in lung carcinoma; the low heritability (2.62%) of never-smoking lung cancer is aligned with the uncommon variants identified in this disease subtype.

Session Title: Cancer Poster Session III

PB5112 Meta-analysis of Rare Loss of Function Variant Data from the UK Genetic Prostate Cancer Study (UKGPCS) Identifies Genes Associated with Risk of Aggressive Prostate Cancer.

Authors:

E. Saunders[1], T. Dadaev[1], M. N. Brook[1], S. Wakerell[1], K. Govindasami[1], R. Rageevakumar[1], N. Hussain[1], A. Osborne[1], D. Keating[1], A. Lophatananon[2], K. Muir[2], UKGPCS Collaborators, A. C. Antoniou[3], R. A. Eeles[1,4], Z. Kote-jarai[1]; [1]Inst. of Cancer Res., London, United Kingdom, [2]Univ. of Manchester, Manchester, United Kingdom, [3]Univ. of Cambridge, Cambridge, United Kingdom, [4]Royal Marsden NHS Fndn. Trust, London, United Kingdom

Prostate Cancer (PrCa) is the second most frequently diagnosed cancer among men globally, and a substantial cause of mortality. Despite this, the majority of PrCa patients experience indolent, slow-developing disease, although a subset develop aggressive forms with poorer prognosis. Several candidate genes have been proposed in which rare germline mutations may increase PrCa risk, however due to sample size limitations only the *BRCA2* gene has been robustly validated to date, alongside less consistent evidence in support of a handful of additional genes. In order to maximise achievable power for the detection of genes associated with risk of aggressive PrCa, we have accumulated a cohort of 6,809 PrCa cases, through a meta-analysis of rare variant sequencing data from men in the UK Genetic Prostate Cancer Study (UKGPCS), sequenced as part of six previously reported studies. We curated a set of 10 genes (*ATM*, *BRCA1*, *BRCA2*, *CHEK2*, *MLH1*, *MSH2*, *MSH6*, *MUTYH*, *NBN* and *PALB2*) that had been sequenced in all samples available for the meta-analysis as well as having been previously proposed as candidate moderate penetrance genes contributing towards PrCa susceptibility or aggressiveness. We examined association between rare putative loss of function variants in each gene and aggressive disease (defined as cause of death recorded as PrCa, or at diagnosis any of metastatic disease, stage T4, or both stage T3 and Gleason score >7), whilst secondary analyses examined association with individual phenotypic indicators of aggressiveness (presence of metastases, nodal spread or Gleason score >7). We observed significant evidence for association between mutations in 3 genes (*ATM*, *BRCA2* and *NBN*) and risk of aggressive PrCa at the time of diagnosis after adjusting for multiple testing ($P<0.005$). A further 2 genes (*MLH1* and *MSH2*) demonstrated nominally significant evidence ($P<0.05$) for association with two or more of the phenotypes examined in the primary and secondary analyses. Additionally, through Cox proportional-hazards modelling and Kaplan-Meier survival analyses, we found evidence that carriers of germline mutations in the 5 candidate genes associated with aggressive phenotypes experience shorter all-cause (HR=2.24, 95% CI 1.91-2.63, $P=3.4\times10^{-23}$) and PrCa-specific survival (HR=2.34, 95% CI 1.93-2.84, $P=5.1\times10^{-18}$) than patients who do not carry a mutation. These observations support the role of rare pathogenic germline mutations in risk of aggressive PrCa and could help define the panel of genes for which sequencing may be informative for the identification of men at elevated risk of PrCa with poorer prognosis.

Session Title: Cancer Poster Session I

PB5113 Metabolic liver cancer: Associations of rare and common germline variants in one-carbon metabolism and DNA methylation genes

Authors:

S. Antwi[1], M. Heckman[1], L. White[1], I. Yan[1], V. Sarang[2], K. P. Lauer[3], J. Reddy[1], F. Ahmed[3], S. Veliginti[1], E. D. Mejías Febres[4], R. I. Hatia[5], P. Chang[6], L. Izquierdo-Sanchez[7], L. Boix[8], A. Rojas[9], J. Banales[7], M. Reig[8], P. Stål[10], M. Romero-Gomez[9], A. G. Singal[11], D. Li[6], M. M. Hassan[6], L. Roberts[2], T. Patel[1]; [1]Mayo Clinic, Jacksonville, FL, [2]Mayo Clinic, Rochester, MN, [3]Mayo Clinic, Rochester, FL, [4]Univ. of Puerto Rico Med. Sch., San Juan, Puerto Rico, [5]The Univ. of Texas MD Anderson Cancer Ctr., Houston, TX, [6]The Univ. of Texas MD Anderson Cancer Ctr., HOUSTON, TX, [7]Biodonostia Hlth.Res. Inst. - Donostia Univ. Hosp., Univ. of the Basque Country (UPV/EHU), San Sebastian, Spain, [8]BCLC Group, Liver Unit, ICMDM, IDIBAPS, Hosp. Clinic of Barcelona, Univ. of Barcelona, Barcelona, Spain, [9]Inst. of Biomedicine of Seville (IBiS), Virgen del Rocio Univ. Hosp./CSIC/Univ. of Seville, Seville, Spain, [10]Karolinska Univ. Hosp., Karolinska Inst., Stockholm, Sweden, [11]Univ. of Texas Southwestern Med. Ctr., Dallas, TX

Background: Animal studies implicate one-carbon metabolism and DNA methylation genes in hepatocellular carcinoma (HCC) development in the setting of metabolic perturbations. We investigated this in humans by assessing the associations between common and rare variants in these closely related biochemical pathways and risk for metabolic HCC in a multicenter international study. **Methods**: Targeted exome sequencing of 64 genes was performed among 556 metabolic HCC cases and 643 cancer-free controls with metabolic conditions. Multivariable logistic regression was used to calculate odds ratios (ORs) and 95% confidence intervals (CIs), adjusting for multiple comparisons. Gene-burden tests were used for rare variant associations. Analyses were performed in the overall sample and among non-Hispanic whites. **Results**: Among non-Hispanic whites, presence of rare functional variants in *ABCC2* was associated with 7-fold higher risk of metabolic HCC (OR=6.92, 95%CI: 2.38-20.15, p=0.0004), and this association remained significant when analyses were restricted to functional rare variants observed in ≥2 participants (cases 3.2% vs. controls 0.0%, $p=1.02 \times 10^{-5}$). In the overall multiethnic sample, presence of rare functional variants in *ABCC2* was nominally associated with metabolic HCC (OR=3.60, 95%CI: 1.52-8.58, p=0.004), with similar nominal association when analyses were restricted to functional rare variants observed in ≥2 participants (cases 2.9% *vs.* controls 0.2%, p=0.006). A common variant in *PNPLA3*-rs738409[G] was associated with higher HCC risk in the overall sample ($p=6.36 \times 10^{-6}$) and in non-Hispanic whites (p=0.0002). **Conclusions**: Rare functional variants in *ABCC2* are associated with susceptibility to metabolic HCC in non-Hispanic whites. *PNPLA3*-rs738409 is also associated with metabolic HCC risk.

Session Title: Cancer Poster Session II

PB5114 Methods to characterize mosaic and engineered variants in HG002 using HG002 T2T diploid assembly

Authors:

A. Abdulkadir[1], C. Daniels[1], M. Cleveland[2], N. Olson[2], N. Dwarshuis[2], M. de Mars[1], J. Zook[3]; [1]Med. Device Innovation Consortium (MDIC), Arlington, VA, [2]Natl. Inst. of Standards and Technology (NIST), Gaithersburg, MD, [3]Natl. Inst. of Standards and Technology, Gaithersburg, MD

The Somatic Reference Sample (SRS) Initiative is a program led by the Medical Device Innovation Consortium (MDIC) to improve the accuracy, reliability and transparency of NGS-based oncology tests by developing well characterized somatic variant reference samples and benchmarks. The pilot project involves using CRISPR to engineer individual cancer variants into the well characterized Genome in a Bottle Consortium HG002 cell line. Currently, the Telomere-to-Telomere Consortium (T2T) is working towards a highly accurate diploid assembly of the HG002 cell line. In advance of having the engineered SRS materials, we will evaluate whether the HG002 T2T assembly can be used to detect on and off-target edits in the SRS materials. To do this, we first will simulate the products of in-vitro SRS engineering by conducting in-silico genome editing of real HG002 data, resulting in ten edited genomes that each contain a clinically relevant somatic (cancer) variant and simulated "off-target" edits. An unedited HG002 genome will be treated as the normal sample input while each edited HG002 genome will be treated as the tumor input for somatic variant calling. We hypothesize that mapping reads to the paternal or maternal haplotypes will reduce biases in mapping and somatic variant calling, particularly around structural variants and complex variants. We will evaluate this hypothesis by comparing variants called using the current human genome reference (GRCh38) as a reference vs T2T-HG002 as a reference for different somatic and mosaic variant callers. By evaluating the in-silico edit results using the HG002 T2T assembly, we aim to help comprehensively capture a wide range of human genetic variation and fine tune SRS analyses.

Session Title: Cancer Poster Session III

PB5115 Methylation Detection in Cancer Cells with Long Read Sequencing Technology

Authors:

J. Zhang, B. Wong, L. Liu, P. Collins, N. Fitzpatrick, S. Wood, J. Pearson, N. Waddell; QIMR Berghofer Med. Res. Inst., Brisbane, Australia

DNA methylation of cytosine is an epigenetic mechanism that can transform cell gene expression. It plays important roles in biological processes in normal cells and aberrant DNA methylation is a hallmark of cancer cells. Traditional methods for profiling cytosine methylation in DNA rely on bisulfite conversion. The Oxford Nanopore long-read sequencing technologies provide distinct advantages by directly sequencing DNA molecules allowing the detection of base modifications in real-time. Long reads are also ideal for phasing haplotypes which enable the identification of allele-specific methylation events. Several tools have been developed to detect cytosine methylation status using the raw sequencing signals from the Oxford Nanopore sequencing technologies. However, the reliability of prediction of these approaches has not been benchmarked in cancers in which tumour purity and heterogeneity have large effects on the results.Here, we evaluate the performance of multiple tools using two cancer cell lines (COLO829, HCC1937) and matched non-cancer cell lines. Existing EPIC Array data of the normal-tumour paired cell lines were used as a gold standard for assessment. To understand the effect of tumour purity on cytosine methylation detection, we physically simulate tumour purity by mixing DNA from tumour and non-tumour cell lines to mimic different tumour purities (100%, 80%, 60%, 40%, 20%, 0%) for Nanopore sequencing. Subsequently, we test the best thresholds to use for each tumour purity and assess the point where the signal from the tumour becomes in-detectable. Our results include the improvement of both base calling and modification calling from using R9.4 chemistry to the latest R10.3 chemistry.Together, this study systematically evaluates the modification base calling of different tools in cancer cells with Nanopore long-read sequencing and exams the impact of purity on detection. It will serve as a guide for the best practice in tumour cytosine methylation detection in tumour tissue samples.

Session Title: Cancer Poster Session I

PB5116 MiR-206 expression in the differential diagnosis of pediatric sarcomas.

Authors:

A. Ahmed[1], L. Yermakov[2], G. Elizabeth[3], M. Zapata-Tarres[4], R. Cardenas-Cardos[5], L. Velasco-Hidalgo[5], C. Corcuera-Delgado[5], R. Rodriguez-Jurado[5], L. García-Rodríguez[6], A. Parrales[3], T. Iwakuma[3], M. Farooqi[3], T. Flatt[3]; [1]Seattle Childrens Hosp., Seattle, WA, [2]Univ. of Washington, Seattle, WA, [3]Childrens Mercy Hosp., Kansas City, MO, [4]Res. Coordination Mexican Inst. of Social Security Fndn., Mexico City, Mexico, [5]Inst. Natl. de Pediatría, Mexico City, Mexico, [6]Univ. de Monterrey, Monterrey, Mexico

Background: MicroRNAs (miRNAs) are short non-coding RNAs that function as post-transcriptional regulators of genes involved in vital physiologic processes within the cell. Their dysregulation has been shown to play a role in development and progression of cancer. Expression patterns of miRNA can be used to discriminate tumor types and to predict outcome. Materials and Methods: A total of 108 tumors in children have been analyzed for microRNA expression including 32 osteosarcoma (OS), 26 Ewing sarcoma (EWS), and 50 rhabdomyosarcoma (RMS) cases, collected from two Children's hospitals, in the USA and Mexico. RNA, retrieved from formalin-fixed paraffin-embedded tumor tissue scrolls from de-identified patients, was used for experiments on the NanoString multiplex platform (nCounter Human miRNA Expression Assay kit). The expression of 827 human miRNAs have been sequenced for clinically relevant miRNAs. Results were analyzed for read distribution percentages and data visualization through heat maps and volcano plots and compared via fold changes and p-values generated using statistical methods. Results: In comparison to EWS and OS, the RMS tumors had a total of 56 differentially expressed miRNAs [>3 fold change and $p<0.01$]. Among them, miR-206 was the most significantly overexpressed and miR-29a-3p most significantly downregulated in RMS compared to EWS. miR-206 was also among the most significantly upregulated miRNA in RMS compared to OS. miR-206 had the highest significant overexpression (>55 fold change, $p<0.0001$) in RMS and exhibited the highest sensitivity and specificity in differentiating RMS from EWS and OS. A set of eight miRNAs can differentiate embryonal from alveolar RMS, including let-7e-5p which was most significantly overexpressed in embryonal RMS [>2 fold increase, $p<0.01$]. A total of 47 miRNAs were differentially expressed in EWS versus OS. There were no significant racial or ethnic differences in miRNA expression. Conclusions: Knowledge of differentially expressed miRNA helps in further understanding of pediatric tumor biology. Specific miRNAs show potential in differentiating the common pediatric sarcomas and in differentiating between two most common subtypes of rhabdomyosarcoma.

Session Title: Cancer Poster Session II

PB5117 MiSeVis: A R-Shiny based platform for visualizing disease candidate genes with mutated protein structures caused by missense variants and predicting drug targets of candidate biomarkers.

Authors:

I. Hou; The John Cooper Sch., The Woodlands, TX

Background: Ovarian cancer is one of the leading killers in the US for women. The American Cancer Society (ACR) ranks ovarian cancer as the fifth leading cause of cancer deaths among women. This cancer is inheritable, with higher chances if the patient has a family history of ovarian cancer. The Cancer Genome Atlas (TCGA) is a joint-effort project by NCI and NHGRI that generates genomic data for the cancer patients and matches normal data samples with these patients. Missense mutations are Single Nucleotide Polymorphisms (SNPs) that modify the amino acids. A missense mutation can disrupt protein structures, possibly altering the function of the affected protein. There is an urgent need to discover potential genes with pathogenic repercussions and visualize the alterations in their 3D protein structure. This study developed a user-friendly platform for automatically prioritizing cancer genes by integrating multiple databases and visualizing changes in protein structure.
Methods and Results: Using my previous research results, I created a convenient platform called MiSeVis using R-Shiny by combining all analyses steps. MiSeVis has a protein visualization function where the user can choose a PDB file and visualize the 3D protein structure based on the file. The user can also choose between mutated and non-mutated files and examine the exact locations of the mutations with different imaging options. The platform has a function in the tool that allows the user to view the contents of both FASTA and PDB files and input their own files as well. As a case study, I visualized a gene Prosaposin (PSAP) with a mutation at amino acid position 469. The software clearly showed how the protein structure is altered and reported its drug targets.
Conclusions: In this study, I implemented a R-Shiny based tool - MiSeVis that allowed the user to visualize protein structure changes and report potential therapeutic drug targets for a given gene. By using the developed platform, I successfully identified genes that have pathogenic consequences in ovarian cancers. MisSeVis can simultaneously visualize the structures of both wild-type and mutant proteins, making it a useful tool for identifying candidate genes resulting from variant-causing structural alteration. It could guide medical researchers to determine the most effective treatment based on the precise alterations in the protein structures that make up a person's particular collection of biomarkers. Scientists can also use these specific biomarkers to develop drugs that specifically target them.

Session Title: Cancer Poster Session III

PB5118 Molecular and clinical features of high-grade serous ovarian cancers in BRCA1 and BRCA2 mutation carriers

Authors:

M. Walters[1], S. J. Winham[1], J. M. Cunningham[2], S. M. Armasu[1], H. J. Atkinson[1], B. M. McCauley[1], A. Mishra[1], C. Wang[1], A. V. Bhagwate[1], E. L. Goode[1]; [1]Mayo Clinic, Rochester, MN, [2]Mayo Clinic & Fndn, Rochester, MN

Background: Pathogenic germline mutations in BRCA1 and BRCA2 represent one of the most significant genetic risk factors for high-grade serous ovarian cancer (HGSOC). However, much remains unknown about clinical and molecular correlates for affected mutation carriers. Methods: HGSOC cases with pathogenic germline mutations in BRCA1 (N=117) or BRCA2 (N=92) were identified through comprehensive gene-based sequencing. Histopathology was confirmed using IHC-assisted review (WT1 and p53). A pathogenic mutation was defined based on annotation of 'likely pathogenic' or 'pathogenic' mutations of potential clinical significance in ClinVar. Medical record abstraction provided age at diagnosis, disease stage (early, advanced), extent of residual disease (any, no macroscopic disease), and neoadjuvant treatment (yes, no). Cases were followed for at least five years. RNA extracted from fresh frozen tissue (>70% tumor) was interrogated with gene expression for estimation of tumor molecular subtype (C1.MES mesenchymal, C2.IMM immunoreactive, C4.DIF differentiated, C5.PRO proliferative). Tumor cores of archival tissue on tissue microarrays were scored by extent of CD8+ tumor infiltrating lymphocytes (TILs; none, low, medium, high). Results: Overall, mean age at HGSOC diagnosis was 58.5 (range, 28 - 84), 92.5% were advanced stage, and 23% received neoadjuvant treatment. Surgeries left no macroscopic disease for 59.2% of cases, and five-year overall survival was 47.1%. As expected, older age, advanced stage, and macroscopic disease were associated with shorter survival time ($p< 0.05$). BRCA1 mutation carriers were more likely than BRCA2 mutation carriers to be fully surgically resected with no macroscopic residual disease (63.6% v 53.3%). However, among cases with chemo-naïve tumors, BRCA1 mutation carriers were more likely than BRCA2 mutation carriers to have C1.MES molecular subtype tumors (29.4% v 19.7%); this finding is inconsistent with the premise that C1.MES tumors may be more challenging to resect. BRCA1 mutation carriers were less likely than BRCA2 mutation carriers to have C2.IMM molecular subtype tumors (7.1% v 19.7%), although frequencies of CD8+ TILs were similar among both mutation carrying groups. There was no difference in survival between BRCA1 and BRCA2 mutation carriers after adjustment for age, stage, residual disease, molecular subtype, and CD8+ TILs ($p=0.69$). Conclusion: Laying the foundation for larger scale analyses, results suggest a potentially complex interplay between tumor molecular subtype, surgical resectability, and immune features in HGSOC that differs by germline BRCA1 and BRCA2 mutation status.

Session Title: Cancer Poster Session I

PB5119 Molecular characterization of appendiceal epithelial cancer using integrated single-cell RNA sequencing

Authors:

L. Bui-Raborn[1], X. Cao[2], J. wang[3], F. Meng[2], M. Feng[2], L. Arvanitis[2], R. Mannan[2], M. Raoof[2], N. Banovich[1]; [1]Translational Genomics Res. Inst., Phoenix, AZ, [2]City of Hope, Duarte, CA, [3]city of hope, Monrovia, CA

Appendiceal cancer develops from the cells of the appendix and is a rare cancer with 1-2 cases per 1 million individuals. Epithelial appendiceal malignancy starts from the lining cells of the appendix, in most cases it leads to the accumulation of mucins, appendix rupture, and spread inside the abdominal cavity. Epithelial appendiceal cancer is characterized by low-grade or high-grade invasive adenocarcinomas or appendiceal mucinous neoplasm. Without a reliable blood or urine test, diagnosing appendiceal cancer at its early stages is challenging and the classification of cancer grades can only be done on tissue specimens by a pathologist.To comprehensively characterize the cell types and molecular mechanisms driving cellular remodeling in appendiceal epithelial cancer, we performed an integrated scRNA-seq study on 127,000 cells from 16 appendix samples (11 peritoneal metastases samples and 5 healthy controls). We identify 38 distinct cell types/cell states with 9 being cancer-specific. Overall, we observe a remarkable shift in the transcriptional profile of many cell types in different pathology groups compared to control samples. Specifically, cancerous CD4+ T cells display lower expression levels of naive markers (TCF7, SELL, LEF1) while the expression of regulatory markers like IL2RA and IL4R is upregulated in cancerous Tregs compared to normal cells. PDL1-related immune checkpoints genes such as TNFRSF4, CD28, and CTLA4 display dysregulation in tumor T cells compared to controls. We identify three subtypes of cancer-associated fibroblasts (CAFs) - iCAFs, myCAFs and apCAFs, in the cancerous samples, but not in the healthy controls. Consistent with the histology classification of appendix biopsies, we report a highly expressed MUC5B+ cell cluster in the mucinous goblet-cell adenocarcinoma samples. Ligand/receptor analyses show an upregulation of many signaling pathways, including Midkine and PTN pathways, in the immune cells and CAFs derived from mod/high-grade and goblet-cell adenocarcinoma samples. Together, our study provides high-resolution insights into the complexity and plasticity of different appendiceal cancer pathologies and is a valuable resource for the development of biomarkers to classify appendiceal cancer grades.

Session Title: Cancer Poster Session II

PB5120 Molecular landscape of non-driver genes in myeloproliferative neoplasms through next generation sequencing; insights to reveal in Pakistan

Authors:

R. Ahmed[1,2], M. Rashid[2], M. Usmani[3], S. Mukry[2], U. Zaidi[2]; [1]Univ. of Karachi, Karachi, Pakistan, [2]Natl. Inst. of Blood Disease and Bone Marrow Transplantation, Karachi, Pakistan, [3]Univ. of Maryland, Baltimore, MD

Background: Discoveries of driver mutations in myeloproliferative neoplasms (MPNs) have filled the diagnostic gap however non-driver genes also play an important role in the phenotype of the disease. Mutations in driver genes *JAK2, CALR* and *MPL,* defining phenotypic alterations, aid in differential diagnosis of MPNs subcategories polycythemia (PV) vera, essential thrombocythemia (ET) and primary myelofibrosis (PMF). However, a triple negative MPNs group also exists with wild type driver genes that exhibits the diagnostic and prognostic utility of comprehensive mutational analysis of non-driver genes. Therefore, in order to understand the complex pathophysiology of MPNs beside driver mutations; defects in non-driver genes are the subject of extensive current research. This study is the first to evaluate the molecular landscape of non-driver genes in MPNs patients from Pakistan. **Methods:** This cross sectional study was designed within the ethical boundaries mentioned in the declaration of Helsinki. A cohort of fourteen MPNs patients (eight ET, five PMF and one PV) was investigated by the next generation sequencing, using 333 cancer gene panel, to attain comprehensive genomic analysis. Diagnosis of PV, ET and PMF was done according to the WHO diagnostic criteria. Basic demographical and haematological parameters, *JAK2*V617F (ARMS PCR), *CALR* and *MPL* status (Sanger sequencing) were reviewed from the patient's medical record. Chi square test was run in SPSS 22.0 to check association of non-driver genes with sub categories of MPNs. **Results:** Among 333 oncology related genes, possible pathogenic variations were identified in 2.1% of analyzed genes (7/333). *TP*53 and *KIT* were the only known non-driver genes in MPNs which were mutated in this cohort. The highest frequency (85.7%) was found of *UGT1A1* gene variant **28* with 71.4% heterozygous genotype (*1/*28) and 14.2% homozygous genotype (*28/*28). Second most common (64.2%) detected gene variants were of *MTHFR* with WT/A1298C, A1298C/A1298C and C677T/A1298C genotypes 28.5%, 28.5% and 7.1%, respectively. Frequency of *TP*53 substitution c.215C>G was 57.1% and *XRCC1* Q399R was 42.8%. *KIT* CNV was 42.8% whereas *KIT* substitution c.1924A>G was 7.1%. The frequency of *DPYD* genotype *9A/c.496A>G/ IVS1 0-15T>C and *2A/*9A/c.496A>G was 21.4%. The lowest frequency (7.1%) was observed of *CYP*2D6 *4/*41. *KIT* was significantly (*P*=0.026) frequently mutated in PMF patients (4/5). **Conclusion:** A distinct molecular landscape of non-driver genes was observed in MPNs from Pakistan and most of the genes variants detected belonged to drug metabolizing pathways.

Session Title: Cancer Poster Session III

PB5121 Multidimensional mutational scanning of the human insulin receptor gene to accelerate diagnosis and enable translational studies

Authors:

V. Aslanzadeh[1], G. Brierley[2], K. Matreyek[3], G. Kudla[1], R. Semple[1]; [1]The Univ. of Edinburgh, Edinburgh, United Kingdom, [2]Anglia Ruskin Univ., Cambridge, United Kingdom, [3]Case Western Reserve Univ., Cleveland, OH

Loss-of-function variants in the human insulin receptor gene, *INSR*, lead to severe insulin resistance, which is often fatal in infancy or childhood when variants are biallelic. More than 150 pathogenic mutations have been described, but few have been functionally studied, posing challenges for accurate genetic diagnosis. It also impedes patient stratification for novel candidate treatments such as peptide, antibody, and small molecule ligands that activate the receptor through non-canonical mechanisms. Discriminating pathogenic alleles from insignificant variants, and establishing which are expressed and responsive to atypical ligands, and achieving this at scale, will be critical for implementing translational studies. To address this challenge, we used plasmid-based saturating mutagenesis coupled to massively multiplexed Flow-Seq-based assays in mouse embryo fibroblasts in which endogenous insulin and IGF-1 receptors had been knocked down or out respectively. This permitted interrogation of the effect of 14,000 missense variants (around 75% of all possible variants) in the human INSR extracellular domain. Our analysis identified variants that reduce or eliminate receptor cell surface expression, and variants that maintain cell surface expression but exhibit selectively impaired insulin binding and/or signalling. A small number of variants also conferred gain of function in the assays used. We also assayed binding of two bivalent anti-receptor monoclonal antibodies that we have previous shown to act as partial agonists at some expressed mutant receptors. This identified a group of mutations with greater signalling in response to antibody than to insulin, clustering in known insulin-binding domains. These mutations are promising candidates for therapeutic targeting with such anti-receptor monoclonal antibodies. Our assays confirmed previous studies of characteristics of limited numbers of disease-causing variants, agree well with available molecular structures of the insulin receptor, and provide a multidimensional sequence-function map for INSR variant interpretation. The empirical data are currently being compared to outputs of a panel of computational predictors of functional effects of coding variants, and a prediction tool to aid in clinical diagnosis of rare pathogenic INSR variants is being developed. This will also enable rapid identification of variants with attributes amenable to activation by non-canonical INSR ligands, and will inform stratification of variants in population-based reverse genetic studies.

Session Title: Cancer Poster Session I

PB5122 Mutational processes of tobacco smoking and APOBEC activity generate protein-truncating mutations in cancer genomes

Authors:

J. Reimand[1,2], A. T. Bahcheli[1,2], K. Cheng[1,2], K. N. Al-Zahrani[3], M. Slobodyanyuk[1,2], D. Pellegrina[2], D. Schramek[1,3], N. Adler[2]; [1]Univ. of Toronto, Toronto, ON, Canada, [2]Ontario Inst. for Cancer Res., Toronto, ON, Canada, [3]Lunenfeld-Tanenbaum Res. Inst., Toronto, ON, Canada

Mutational signatures represent a footprint of tumor evolution and its endogenous and exogenous mutational processes. However, their functional impact on the proteome remains incompletely understood. We analysed the protein-coding impact of single base substitution signatures in 12,341 cancer genomes from 18 cancer types. Stop-gain mutations (SGMs) were strongly enriched in the signatures of tobacco smoking, APOBEC cytidine deaminases, reactive oxygen species (ROS), and less-frequent carcinogenic signuatures. These mutational processes affect specific trinucleotide contexts to substitute serine and glutamic acid residues with stop codons and thus provide a mechanistic model of the SGM-generating mutational processes. The signature-associated SGMs are enriched in cancer hallmark pathways and dozens of known and putative tumor suppressor genes including TP53, FAT1, and APC, tying together the mutational processes and positive selection in cancer genomes. Tobacco-driven SGMs in lung cancer correlate with lifetime smoking history and highlight a preventable lifestyle determinant of these harmful mutations. In breast cancer, the burden of APOBEC-driven SGMs is correlated with gene expression of APOBEC3 enzymes, associating these mutations with the expected molecular pathway. In summary, our study exposes SGM expansion as a genetic mechanism by which endogenous and carcinogenic mutational processes contribute to protein loss-of-function, oncogenesis, and tumor heterogeneity, providing potential translational and mechanistic insights.

Session Title: Cancer Poster Session II

PB5123 Mutational spectrum of Korean patients with T-cell acute lymphoblastic leukemia

Authors:

H. Kim[1], S. Jo[2], H. Kim[1], S. Oh[3], I. Kim[3]; [1]Pusan Natl. Univ. Hosp., Busan, Korea, Republic of, [2]Seegene Med. Fndn., Busan, Korea, Republic of, [3]Pusan Natl. Univ. Yangsan Hosp., Yangsan, Korea, Republic of

Introduction: The landscape of genetic alterations implicated in the pathogenesis of T-cell acute lymphoblastic leukemia (T-ALL) has been revealed in recent years, but there are limited data that have thoroughly analyzed these variant patterns in Korean patients. We investigated the results of gene panel tests performed on Korean patients with T-ALL. **Methods:** The study included 19 Korean patients newly diagnosed with T-ALL between January 2016 and December 2021, eight of which were classified as early T-cell precursor ALL (ETP-ALL) patients. We performed customized multigene panel tests targeting at least 101 genes with next-generation sequencing. Clinical data and other laboratory findings including immunophenotypes, cytogenetics, and bone marrow studies were also reviewed. **Results:** A total of 78 clinically significant or reportable mutations were identified in 40 genes, and all patients harbored at least two mutations (range 2-8 per patient). The median of variant allele frequencies (VAFs, %) and bone marrow blasts at diagnosis (%) were 43% (range 3% to 96%) and 79% (range 21%-92%), respectively. Variants in NOTCH signaling genes were found in 11 patients (58%), including all non-ETP ALL patients. Mutations related to the JAK-STAT pathway, epigenetic factors, and PI3K-AKT pathway were detected in 32% (6/19), 26% (5/19), and 11% (2/19) of patients, respectively, and most patients (79%, 15/19) had genetic alterations implicated in two or more pathways. **Conclusion:** Using multigene panel testing, we detected clinically significant genetic mutations in all T-ALL patients in this study, including those with normal cytogenetics. Despite the small number of patients, this study provides insight into the mutational spectrum implicated in the leukemogenesis of T-ALL in Koreans.

Session Title: Cancer Poster Session III

PB5124 Myelodysplastic syndromes progression: A comprehensive expression-based *in silico* analysis.

Authors:

B. McKinney, J. Ferreira de Vasconcellos; James Madison Univ., Harrisonburg, VA

Myelodysplastic syndromes (MDS) are a heterogeneous group of clonal disorders characterized by ineffective hematopoiesis in the bone marrow leading to cytopenias in the peripheral blood and predisposition to secondary Acute Myeloid Leukemia (sAML). The current overall survival of sAML is 5 months in patients younger than 60 years of age and 4.7 months in patients ≥ 60 years of age. As a result, the discovery of novel biomarkers and treatment strategies to prevent the progression from MDS to sAML are urgently needed. Here we investigated novel critical molecular pathways that support MDS-to-sAML transformation. Using a publicly available dataset from the Gene Expression Omnibus database (https://www.ncbi.nlm.nih.gov/geo/) we generated a list of differentially expressed genes comparing the datasets from 164 MDS to 202 sAML patients to systematically identify molecular pathways that are predicted to support MDS-to-sAML transformation in the bone marrow niche. In addition, we performed pathway analyses using Metascape v3.5.20230101 (https://metascape.org) and ShinyGO 0.77 (http://bioinformatics.sdstate.edu/go/). Our results identified 232 differentially expressed genes (log2 fold change: ≥1.5 or ≤ -1.5) where 206 genes were up-regulated and 26 genes were down-regulated. In addition, the gene expression signature from MDS compared to sAML datasets identified enriched molecular pathways, such as regulation of hemopoiesis, hematopoietic stem cell differentiation, immune system development and skeletal system development. Overall, these results have an important positive impact because they address aspects of unique importance to MDS transformation, including fundamental insights in the disease biology that will be investigated as potential biomarkers and/or new targets for therapeutic intervention.

Session Title: Cancer Poster Session I

PB5125 Noncoding and germ line mutations in African American esophageal squamous cell carcinoma

Authors:

R. Wadleigh[1], H. Erkizan[2]; [1]DC VA Med. Ctr., Washington, DC, [2]Inst. for Clinical Res., DC VAMC, Washington, DC

Esophageal Squamous Cell Carcinoma (ESCC) ranks among the most lethal cancers and is particularly aggressive among African Americans (AA). To conduct an extensive mutation screening, we performed whole exome sequencing in 10 paired AA ESCC and control tissues. Mutations included nonsynonymous mutations in multiple genes and copy number variations. Here, we report on further analysis of the WES data to identify noncoding region and germ line genetic alterations. Late-stage disease showed 340 genomic mutations, two thirds of which occurred mostly in the 3`UTR of differentially expressed genes (DEGs) suggesting association of dysregulation at the 3`UTR with metastasis. AA ESCC related changes in promoter regions were detected, and 70% of these mutated sites were revealed in the RNA polymerase (*POLR2A*) binding site. In addition, germ line variants were displayed 21 DNA damage repair (DDR) genes. *RAD52*, a gene involved in repair of DNA double-stranded break, that has been reported previously to exert a role in genetic susceptibility for squamous cell cancer of the upper aerodigestive tract, carried >20 different mutations in 10 control tissues. *BLM*, which encodes a DNA helicase that has been reported as a genome stabilizer displayed a missense mutation with a high deleteriousness (CADD) score. Five other germ line variants were presented in *TP53*. We propose that the combined effect of these genetic and coding region mutations that exert a large impact on poor prognosis in AA ESCC may yield targets for developing candidate therapeutics.

Session Title: Cancer Poster Session II

PB5126 Novel genetic loci associated with cirrhosis and primary liver cancer in the Million Veteran Program suggest potential therapeutic relevance for EPHA2 and JAK2 signaling

Authors:

M. Vujkovic[1], T-H. L. Schwantes-An[2], D. Dochtermann[3], B-L. Chang[4], M. G. Levin[4], C. V. Schneider[5], D. Gill[6], S. Mukherjee[4], A. Verma[4], A. Mezina[4], S. Pyarajan[7], P. Devineni[3], L. Bastarache[8], A. Lewis[8], K-M. Lee[9], C. C. Teerlink[9], P. S. Tsao[10], D. J. Rader[4], B. F. Voight[4], J. A. Lynch[9], A. Shaked[4], M. A. Hoteit[4], T. R. Morgan[11], Million Veteran Program, D. E. Kaplan[1], K-M. Chang[1]; [1]Corporal Michael J. Crescenz VA Med. Ctr., Philadelphia, PA, [2]Indiana Univ. Sch. of Med., Indianapolis, IN, [3]VA Boston Hlth.Care System, Boston, MA, [4]Univ. of Pennsylvania Perelman Sch. of Med., Philadelphia, PA, [5]Univ. Hosp. RWTH Aachen, Aachen, Germany, [6]Imperial Coll. London, London, United Kingdom, [7]VA Boston Hlth.care System, Boston, MA, [8]Vanderbilt Univ. Med. Ctr., Nashville, TN, [9]VA Salt Lake City Hlth.Care System, Salt Lake City, UT, [10]VA Palo Alto Hlth.Care System, Palo Alto, CA, [11]VA Long Beach Hlth.care System, Long Beach, CA

Chronic liver disease (CLD) is a leading cause of death worldwide due to associated cirrhosis and hepatocellular carcinoma (HCC). Regardless of underlying causes, a subset of patients with CLD will progress to cirrhosis, and subsequently a subset of cirrhotic patients will develop HCC. We aimed to better define the genetic architecture of liver disease progression by disentangling the genetic contributions to cirrhosis and HCC, and to identify causative genes and downstream signaling pathways for potential pharmacologic manipulation. We conducted two multi-ancestry genome-wide association studies (GWAS) of all-cause cirrhosis and HCC in the Million Veteran Program (MVP) biobank. The cirrhosis GWAS with 10,931 cirrhosis cases and 162,603 non-cirrhotic CLD controls (66% European, 22% African American, 12% Hispanic) identified 14 independent genome-wide significant loci, including 4 known loci (*PNPLA3, HSD17B13, SERPINA1, TMS6F2*). Eleven loci are known to promote metabolic and liver disease and implicate genes with differential liver cell type expression. A target-based screening for drug repurposing using DrugBank data identified JAK2 inhibitors as a potential therapeutic target for cirrhosis. For the HCC GWAS, we performed a genome-wide meta-analysis of 5,047 HCC cases and 11,092 cirrhotic controls from several studies and identified 3 novel (*EPHA2, MYC, RAB28,* and *SAMD5*) and 3 known loci (*PNPLA3, TM6SF2,* and *TERT).* A drug screen identified EphA2 signaling as having a potential therapeutic role in HCC. We therefore performed drug target MR analysis of genetically-proxied EphA2 signaling using instrumental variables based on liver tissue gene expression data from the GTEx project and circulating plasma protein levels from the deCODE consortium, respectively. In contrast to previous reports of HCC suppression by EphA2 inhibition, we showed that genetically-proxied increased EphA2 levels were associated with reduced HCC risk in analyses both considering gene expression (OR 0.62, 95%CI 0.53-0.71, $P=2x10^{-11}$) and plasma protein levels (OR 0.59, 95%CI 0.40-0.87, P = 0.008). These findings were corroborated by high colocalization at the EPHA2 locus for gene expression, protein concentration, and risk of HCC. In summary, we provide novel insights into the pathophysiology of cirrhosis and HCC and demonstrate both distinct and shared genetic risk loci for the progression of chronic liver disease to cirrhosis and HCC. Our drug target MR supports that EphA2 signaling plays a role in HCC, however the directionality of effect warrants further study, particularly in the context of ongoing clinical investigation of EPHA2 inhibitors for treating HCC.

Session Title: Cancer Poster Session III

PB5127 *NSD2* expression driven by the t(4;14) translocation disrupts the DNA methylation landscape in multiple myeloma

Authors:

R. Chavez[1], D. R. Powell[1], K. Lakhani[1], A. Cook[2], M. Hamilton[2], G. Mulligan[2], D. Auclair[3], J. J. Keats[4], P. M. Vertino[5], L. H. Boise[1], S. Lonial[1], K. N. Conneely[1], B. G. Barwick[1]; [1]Emory Univ., Atlanta, GA, [2]Multiple Myeloma Res. Fndn., Norwalk, CT, [3]AstraZeneca, Waltham, MA, [4]Translational Genomics Res. Inst., Phoenix, AZ, [5]Univ. of Rochester, Rochester, NY

The t(4;14) chromosomal translocation, which places the histone methyltransferase gene NSD2 under the regulatory control of an enhancer of IGH, is observed in 15-19% of multiple myeloma patients and is associated with poor prognosis. The alteration gives rise to a transcriptionally distinct subtype of myeloma, but the mechanisms by which the translocation alters gene expression and leads to high-risk disease are poorly understood. It is known that H3K36me2, which is deposited by NSD2, is widely distributed at high levels in t(4;14) myeloma. Additionally, it has been shown that the PWWP domains of de-novo DNA methyltransferases recognize H3K36me2. To better understand how increased NSD2 activity and widespread H3K36me2 lead to epigenetic and transcriptional dysregulation, we generated DNA methylation (DNAm) data by whole genome bisulfite sequencing on 415 samples from the MMRF CoMMpass study and integrated these data with existing genetic, transcriptional, and clinical data. A differential methylation analysis between samples of the t(4;14) subtype and samples of other subtypes indicated t(4;14) was the most epigenetically distinct form of myeloma and identified 2,075,489 differentially methylated loci ($p < 1e\text{-}9$). The vast majority of these loci (92%) were hypermethylated in t(4;14) samples and were largely located in heterochromatic, late replicating regions of the genome. These hypermethylated loci were significantly enriched within several classes of transposable elements (TEs) from the RepeatMasker database ($1.02 < OR < 1.59$, $p < 0.001$) but roughly evenly distributed between gene body and intergenic DNA (51% to 49%, respectively). This pattern was specific to NSD2 expression as analysis of myeloma cell lines with NSD2 ablation showed a marked decrease in DNAm. Gene set enrichment analysis on differentially expressed genes in t(4;14) patient samples showed downregulation of biological processes including translation, immune signaling, and oxidative phosphorylation, demonstrating biological implications of this aberrant epigenetic programming. These results indicate that t(4;14) driven NSD2 expression and resulting widespread H3K36me2 disrupts the DNAm landscape leading to a transcriptionally distinct and high-risk form of multiple myeloma.

Session Title: Cancer Poster Session I

PB5128 Olink Insight and Human Disease Blood Atlas to uncover human disease proteome and accelerate adoption of proteomics.

Authors:

M. Förster[1], M. Rucevic[1], M. Bueno Álvez[2], M. Uhlen[2], O. Caster[3], H. Andersson[3], F. Edfors[2], L. Fagerberg[2]; [1]Olink Proteomics, Waltham, MA, [2]Sci. for Life Lab., Dept. of Protein Sci., KTH Royal Inst. of Technology, Stockholm, Sweden, [3]Olink Proteomics, Uppsala, Sweden

Despite decades of outstanding research and the impact of genomics, cancer remains an overwhelming burden on global healthcare resources and challenge for the scientific community. Next generation proteomics technologies are increasingly emerging to better understand cancer biology and identify early diagnostic, prognostic, and therapeutic response biomarkers essential to drive development of new, more effective therapies. To progress cancer research more rapidly, data sharing and free access to generated data remains an essential step forward. Here, comprehensive proteome profiling, using Olink Explore platform, was used to measure 1,463 high and low-abundant proteins in plasma, collected at the time of diagnosis and before treatment, from more than 1,500 patients representing 15 common cancer types. The obtained results were used as a foundation for establishment of the Olink Insight platform, an open-access digital knowledge data resource to accelerate adoption of proteomics in the research community. Olink Insight comprises key features such as pathway browser, panel and biomarker selection, and proteomics publication explorer. In Olink Insight, we are creating a collection of proteomic profiles for some of our most important diseases, starting here with cancer. Results can be interactively explored through a variety of analyses and visualizations, including differential expression analysis, pathway enrichment, pathway annotation with hexmaps, and predictive protein groups identified by machine learning. Using a complementary multi-step statistical approach for the Human Protein Atlas (HPA) platform, a panel of 83 proteins was identified that can discriminate 12 different types of human cancer with extremely high accuracy. The plasma profiles for all measured proteins across 12 types of cancers are available in HPA. Olink Insight and the Human Disease Blood Atlas represent a significant step towards uncovering the human disease proteome and will be a valuable resource for researchers in many areas of medicine and biology.

Session Title: Cancer Poster Session II

PB5129 Optical genome mapping for genome-wide structural variation analysis in hematologic malignancies: results of a prospective study and impact on diagnosis and management

Authors:

T. Sahoo[1], S. J. Wicks[1], J. Hauenstein[1], A. Mylavarapu[1], B. Matthews[1], M. Younan[1], K. Hovanes[1], M. Naik[2], N. Ojha[2], R. Arora[2], V. Khushoo[2], V. Babu[2], N. Dhakate[2], P. Jain[2], P. Jain[2], G. Babu[2], S. Rao[2], S. Jadhav[2]; [1]Bionano Lab., San Diego, CA, [2]Hlth.Care Global, Bengaluru, India

Structural variations (SV) play a key role in the pathogenesis of hematologic malignancies. Standard-of-care (SOC) cytogenomic methods, including chromosome karyotyping (KT) and fluorescence *in situ* hybridization (FISH), have inherent limitations, while next-generation sequencing (NGS) technologies have limited ability to detect most SVs. Optical genome mapping (OGM) is a high-resolution technique that overcomes these challenges and allows detection of all classes of SV. Recent studies, mostly retrospective, demonstrate the diagnostic value of OGM for both hematologic malignancies and constitutional genetic disorders. This prospective study, utilizing OGM as a validated laboratory developed test (LDT), included 115 samples from a broad variety of hematologic malignancies undergoing active clinical workups and management. OGM was successful for 107 of 115 samples (93%; eight failed); 76 had abnormal and 31 normal results by OGM (71% and 29% respectively); 48% of abnormal OGM results were classified as complex. Karyotyping and FISH results were available for 45 and 52 samples, respectively. Of the 22 cases with an abnormal KT, 20 were also abnormal and concordant by OGM and identified additional SV in the majority; of the 28 cases with abnormal FISH, 23 were abnormal and concordant by OGM and majority of samples had additional findings. Guideline and SOC result-based disease risk stratification was possible for 50 cases at initial diagnosis with 19 graded as high, 15 as intermediate and 16 as low-risk; the disease risk stratification was reevaluated for 61 cases after completion of OGM analysis and resulted in a change to 28 as high, 13 as intermediate and 28 as low-risk. A final risk assessment (no change, down- or upgrading of risk status) to direct clinical management in 67 cases resulted in 13 as upgraded, 2 as downgraded and 52 with no change in risk status. Additionally, the change in therapeutic management based on addition of OGM results (and likely change in risk status) was also evaluated for 92 and revealed initiation of more intensive therapy in 6 and no change in 86 cases. In summary, OGM identified a case abnormality rate of 71%, a change in disease risk stratification in 22% of cases, and a change in clinical management in 6.5% of cases. The impact of OGM on disease risk stratification and patient management is in alignment with previously published studies and not only provides improved diagnostic assessment but also adds significant prognostic information for better patient management and care. Also, this study suggests a significant role for OGM in providing results with early diagnostic and prognostic potential compared to conventional cytogenetics.

Session Title: Cancer Poster Session III

PB5130 Pan-cancer analysis reveals roles of retrotransposon-fusion RNAs

Authors:

J. Park[1,2], B. Lee[3], Y. Choi[4], A. Voshall[2], E. Maury[1], Y. Kang[5], Y. Kim[3], J-Y. Lee[6], H-R. Shim[6], S-C. Kim[6], B. Hoang[6], D-W. Kim[6], M. Kim[6], E-J. Choi[6], K. Ha[7], J. Choi[5], Y. Kim[6], W-Y. Park[3], E. Lee[2,1]; [1]Harvard Med. Sch., Boston, MA, [2]Boston Children's Hosp., Boston, MA, [3]Samsung Med. Ctr., Seoul, Korea, Republic of, [4]Chonnam Natl. Univ. Med. Sch., Hwasun, Korea, Republic of, [5]KAIST, Daejeon, Korea, Republic of, [6]Yonsei Univ., Seoul, Korea, Republic of, [7]New Drug Dev. Ctr., KBioHlth., Cheongju-Si, Korea, Republic of

Transposons make up about half of the human genome, and some full-length elements create novel insertions in germ cells and somatic tissue of both healthy and diseased individuals. These genetic elements have the potential to generate chimeric transcripts consisting of transposons and non-transposons, which may result in the production of abnormal proteins or immunogenic molecules. However, accurate identification of these transposon fusion events from short RNA-seq reads is challenging due to the repetitive nature of transposon sequences and read alignment errors near exon-intron junctions. To address this, we have developed a computational pipeline, rTea (RNA Transposable Element Analyzer) to detect various types of transposons-fusion transcripts from reference and non-reference, i.e., polymorphic or somatic transposon insertions from RNA-seq data. We applied rTea to analyze 10,257 cancer samples across 34 cancer types, as well as 3,088 normal tissue samples from the TCGA/ICGC, our unpublished colorectal cancer cohort, and the GTEx consortia. We realigned all RNA-seq reads for unified processing on Google Cloud Platform (GCP) using GenomeFlow, a tool to design scalable distributed processing architectures to optimize computational resources and reduce compute costs. We identified 30,016 fusions with an average of 203 events per normal sample, particularly abundant in the testis. We also identified 52,277 cancer-specific fusions that were not detected in the corresponding normal tissues, with an average of 30 events per cancer sample. We found that the somatic cancer fusions were enriched in known cancer genes, suggesting their involvement in tumorigenesis. Furthermore, we discovered distinct splicing hotspots and DNA methylation changes associated with fusions from different families of source transposons. Our in silico immunogenicity modeling and experimental validation confirmed that several peptides derived from transposon fusions in cancers bind to MHC-I and activate CD8+ T cells to a comparable extent to EBV viruses. Our findings highlight the potential of endogenous retroelements as novel therapeutic targets and a significant source of neoantigens. rTea is available at https://github.com/ealeelab/rtea.

Session Title: Cancer Poster Session I

PB5131 † Parent-of-origin assignment of allelic variants without parental data across multiple hereditary cancer syndromes

Authors:

K. Schrader[1], V. Akbari[2], V. Hanlon[3], K. Dixon[4], K. O'Neill[5], A. Roston[4], L. Cordova[6], Y. Shen[5], A. Sharma[7], K. Wong[7], J. Senz[8], Y. Wang[8], D. Chan[8], A. Fok[8], J. Nuk[8], Q. Hong[4], D. Regier[8], A. Virani[9], L. Lefebvre[9], F. Feldman[8], M. Marra[10], S. Sun[7], S. Yip[8], P. Lansdorp[8], S. Jones[8]; [1]BC Cancer / Univ British Columbia, Vancouver, BC, Canada, [2]Dept. of Med. Genetics, Canada's Michael Smith Genome Sci. Ctr., BC Cancer, BC, Canada, [3]BC Cancer Res. Ctr., Vancouver, BC, Canada, [4]Univ. of British Columbia, Vancouver, BC, Canada, [5]BC Cancer, Canada's Michael Smith Genome Sci. Ctr., Vancouver, BC, Canada, [6]Univ British Columbia, Vancouver, BC, Canada, [7]Hereditary Cancer Program, BC Cancer, Vancouver, BC, Canada, [8]BC Cancer, Vancouver, BC, Canada, [9]Dept. of Med. Genetics, Univ. of British Columbia, Vancouver, BC, Canada, [10]BC Cancer, as part of the Provincial Hlth.Services Authority, Vancouver, BC, Canada

Background: Although genetic testing can inform a patient's inherited risk for disease, predicting which side of the family an autosomal variant comes from is a key limitation of current clinical technologies. Parent-of-origin-aware genomic analysis (POAga) using Oxford Nanopore Technologies long-read sequencing combined with Strand-seq enables assignment of any autosomal variant to either parent with 99% accuracy using only the blood sample of the proband. Knowledge of predicted parental segregation is essential in risk management, variant curation, and facilitating cascade genetic testing. We sought to assess POAga in common Tier 1 hereditary cancer conditions such as hereditary breast and ovarian cancer (HBOC) and Lynch syndrome due to germline pathogenic variants in *BRCA1*, *BRCA2*, *MLH1*, *MSH2*, *MSH6*, *PMS2*, and *EPCAM* on chromosomes 17, 13, 3, 2, 7, and 2, respectively, rarer syndromes with parent of origin effects such as hereditary paraganglioma and pheochromocytoma due to paternally inherited pathogenic variants in *SDHD* and *SDHAF2* on chromosome 11 and other genes predisposing to breast cancer and gastrointestinal malignancies, such as *ATM*, *PALB2*, and *CDH1* on chromosomes 11 and 16.**Method:** To determine the analytic validity of POAga, blood samples from carriers of pathogenic variants in *ATM*, *BRCA1*, *BRCA2*, *CDH1*, *MLH1*, *MSH2*, *MSH6*, *PMS2*, *EPCAM*, *PALB2*, *SDHD* and *SDHAF2* of differing age, sex, ethnicity, and cancer status with known parental segregation are currently being ascertained and subjected to Strand-seq and long read sequencing to determine parent-of-origin according to previously described methods (Akbari V, Hanlon VCT, et al. Cell Genom. 2022 Dec 21;3(1):100233.) under an REB approved protocol.**Results:** Thus far 43 informative samples have been been ascertained for POAga across multiple cancer susceptibility
genes, *SDHD* (n=19), *SDHAF2* (n=2), *BRCA1* (n=4), *BRCA2* (n=6), *MSH2* (n=2), *MSH6* (n=5), *PMS2* (n=4), and *EPCAM* (n=1). Sequencing is ongoing for 24 samples. Analysis is complete for 19 samples in *SDHD* (n=18) and *SDHAF2* (n=1), which has demonstrated 100% concordance with known clinical segregation.**Conclusion:** Results to date support the ability of POAga to correctly infer germline variant parent-of-origin using blood samples from hereditary cancer patients with known parental segregation. Ongoing ascertainment and analysis will shed light on the real-world accuracy of POAga and potential to clinically translate this transformational advance in genetic assessment.

Session Title: Cancer Poster Session II

PB5132 PDPR gene variant predisposing to familial papillary thyroid cancer

Authors:

T. Nieminen[1], P. Brock[2], M. Sevigny[1], S. Liyanarachchi[2], W. Li[3], D. F. Comiskey[2], H. He[2], S. Saarinen[1], A. Yilmaz[2], M. Ringel[4], S. Ollila[5], P. Peltomaki[1]; [1]Univ. of Helsinki, Helsinki, Finland, [2]The Ohio State Univ., Columbus, OH, [3]The Ohio State Unversity, Columbus, OH, OH, [4]The Ohio State Univ., columbus, OH, [5]Univ Helsinki, Helsinki, Finland

BACKGROUND: Thyroid cancer is the 12th most common cancer in the USA, with papillary thyroid cancer (PTC) as the predominant subtype. PTC clusters in families in 5 to 15 % of all thyroid cancer cases. No established highly penetrant predisposing variants for familial PTC exist, despite intensive research efforts. **AIM**: Our aim was to identify novel genes and mechanisms underlying PTC susceptibility. **MATERIAL AND METHODS**: Our previous investigation on 17 PTC families led us to conduct a closer analysis on one PTC family (64557) with whole-genome sequencing (WGS) data available on three PTC-affected individuals. Our analysis identified *PDPR* (Pyruvate Dehydrogenase Phosphatase Regulatory Subunit) as a promising candidate gene predisposing to PTC. We subsequently screened 323 sporadic thyroid cancer cases and 12 Familial adenomatous polyposis (FAP) individuals with secondary thyroid cancer for *PDPR* variants from whole-exome sequencing (WES) and WGS data, respectively. We studied the functional and metabolic consequences of the *PDPR* variants by creating PDPR mutant cell lines using CRISPR-Cas9. Finally, we evaluated the Cancer Genome Atlas (TCGA) data for *PDPR* expression in thyroid cancer**. RESULTS**: We found truncating splice donor variants in PTC-affected family members from 64557 (NM_017990.4:c.361+1G>C) and in a sporadic case, SL250416 (NM_017990.4:c.443+1G>C). Nine other sporadic cases as well as a FAP-PTC patient revealed constitutional *PDPR* variants of the missense type. The *PDPR* c.361+1G>C variant segregated perfectly with patients with PTC (n=4) and with a parotid cancer patient (n=1) from family 64557, and was absent in unaffected members (n=2). RNA sequencing data (TCGA) of thyroid cancer tumor with matching normal pairs (n=59) showed significantly lower *PDPR* gene expression ($p<0.0001$) compared to matching normal thyroid tissue. PDPR mutant cells presented with elevated phosphorylation of pyruvate dehydrogenase and displayed metabolic changes. **CONCLUSIONS**: Our finding of a novel truncating *PDPR* germline variant in a three-generation PTC family, together with constitutional (and somatic) *PDPR* variants in additional cohorts, suggests a role for *PDPR* in PTC predisposition. PDPR is a regulatory protein of the Pyruvate dehydrogenase complex (PDC), and our functional data suggests dysregulation of energy metabolism as a potential mechanism mediating the pathogenicity of *PDPR* loss-of-function variants. Full characterization of the role of *PDPR* in thyroid cancer tumorigenesis warrants further investigations.

Session Title: Cancer Poster Session III

PB5133 Performance of Somatic Structural Variant Calling in Lung Cancer using Oxford Nanopore Sequencing Technology

Authors:

L. Liu[1,2], J. Zhang[1], S. Wood[1], F. Newell[1], C. Leonard[1], K. Nones[1], L. T. Koufariotis[1], A. J. Dalley[2], H. Chittoory[2], F. Bashirzadeh[3], J. H. Son[3], D. Steinfort[4], J. P. Williamson[5], M. Bint[6], C. Pahoff[7], P. T. Nguyen[8], S. Twaddell[9], D. Arnold[9], C. Grainge[9], P. T. Simpson[2], D. Fielding[3], J. V. Pearson[1], N. Waddell[1]; [1]QIMR Berghofer Med. Res. Inst., Brisbane, Australia, [2]The Univ. of Queensland, Brisbane, Australia, [3]Dept. of Thoracic Med., The Royal Brisbane & Women's Hosp., Brisbane, Australia, [4]Dept. of Thoracic Med., Royal Melbourne Hosp., Melbourne, Australia, [5]Dept. of Thoracic Med., Liverpool Hosp. Sydney, Sydney, Australia, [6]Dept. of Thoracic Med., Sunshine Coast Univ. Hosp., Birtinya, Australia, [7]Dept. of Thoracic Med., Gold Coast Univ. Hosp., Southport, Australia, [8]Dept. of Thoracic Med., Royal Adelaide Hosp., Adelaide, Australia, [9]Dept. of Respiratory and Sleep Med., John Hunter Hosp., Newcastle, Australia

Background: Lung cancer is a heterogeneous disease and the primary cause of cancer-related mortality worldwide. At diagnosis, over 70% of patients have metastatic disease, with only one-third of patients having an earlier-stage disease that is potentially curable with multimodality therapy. Somatic mutations are important biomarkers for selecting targeted therapy and predicting outcomes for lung cancers. The identification of somatic mutations through genome sequencing of tumors has revolutionized cancer research, playing a crucial role in uncovering driver events, tumorigenesis mechanisms, potentially actionable targets, and intra- and inter-tumor heterogeneity. A key driver of many cancers is somatic structural variants (SVs). Genomic studies in lung cancer have been conducted using short-read sequencing. Emerging long-read sequencing (LRS) technologies, such as the Oxford Nanopore Technologies (ONT), which sequences DNA directly to provide simultaneous whole genome sequencing and methylation profiling is a promising approach to studying somatic SVs and tumor heterogeneity. However, the analysis of LRS genome data is challenging with no consensus on how to process data and call somatic mutations. **Methods:** In this study, we examine the reliability of ONT LRS to detect somatic SVs using a panel of lung cancer samples. We comprehensively benchmark the performance of 3 sequence aligners and 6 SV computational tools. The SV tools include three generic SV callers (sniffles2, cutesv, svim) and three somatic SV callers (delly, nanomonsv, savana). We investigated factors affecting the variant callers and evaluated their performance against high-confidence somatic SV events identified from short-read whole genome sequencing. **Results:** Our results show that different combinations of aligners and variant callers influence somatic SV detection. In particular, the choice of SV callers has a significant influence on somatic SV detection in terms of variant type, size, sensitivity, and accuracy. The recall rate for somatic SV callers (~89%) is higher than generic SV callers (~65%). We will provide a recommendation on the most appropriate computational tools for the detection of somatic SV events from LRS. **Conclusion:** This research enables the development of standardized, efficient, and accurate computational pipelines for long-read somatic SV detection. Better characterization of somatic SV events will contribute to our understanding of cancer development and inform personalized cancer treatment and will have a valuable impact on the field of cancer research.

Session Title: Cancer Poster Session I

PB5134 Plasma Proteomic Signature, Clonal Hematopoiesis and Risk of Myeloid Neoplasm in 46,237 Healthy Individuals

Authors:

D. Tran[1], J. Beeler[1], J. Liu[1], B. Wiley[1], I. Chan[1], M. Kramer[1], A. Batchi-Bouyou[1], X. Zong[1], M. Walter[1], S. Chlamydas[2], F. Ferraro[1], B. Busby[3], D. Link[1], Y. Cao[1], K. Bolton[1]; [1]Washington Univ. in St. Louis, Saint Louis, MO, [2]Olink Proteomics, Frankfurt, Germany, [3]DNAnexus, Mountain View, CA

The presence of clonally expanded mutant hematopoietic stem cells, clonal hematopoiesis (CH), is known to predispose to hematologic malignancy, particularly myeloid neoplasms (MN). However, the overall risk of developing MNs among individuals with CH is low (<0.5% per year) and our poor ability to predict who with CH will progress to MN is a barrier to prevention strategies. The human proteome reflects the downstream impact of complex interactions between genetic and epigenetic regulation of biologic systems. We hypothesized that plasma proteomic markers might be associated with MN risk, and if so, could inform our understanding of the mechanisms promoting CH and MN development. We jointly characterized CH and plasma proteomic profile among 46,237 individuals in the UK Biobank with whole exome sequencing data of the blood and plasma proteomic profiling using the Olink Explore 1536 platform (1,463 unique proteins). Cox proportional hazard models were used to assess the prospective association between the level of each plasma protein and risk of MN; while controlling for age, sex, genetic ancestry, smoking status, blood counts, and sample collection year. The relative contribution of clinical, CH, and proteomics features in prediction of incident MN was assessed using lasso regression analysis with 100-fold cross validation. A total of 115 participants developed MN at least 3 months after study entry (median 83 months). We identified 115 proteins significantly associated with incident MN (FDR-corrected p-value <0.05), many of which (N=34) were also significantly associated with CH. Most of the top 25 associated MN proteins have known roles in immune cell function. These include known regulators of the adaptive immune system (AMIGO2, LAG3, SEMA7A), innate immune system (GCNT1, ICAM, CD244, CLEC6A, FCGR2A, VCAM1, SIGLEC6, LY9, AXL), leukocyte trafficking (CXCL11 or I-TAC, SDC4 or Syndecan-2, ADGRE5 or CD97), hematopoietic stem cell self-renewal (SIRBP1, THPO), and platelet activation (GP1BA, PEAR1). Inclusion of proteomics features significantly (AUC=0.85, p=5e-9) improved prediction of incident MN beyond clinical factors and CH (AUC=0.80). In summary, we show that plasma proteomic markers predict both CH and risk of progression to MN, improving MN risk prediction beyond basic clinical and CH features. These data highlight the promise of protein and genetic biomarker integration for early diagnosis of MN and further support the role of immune cell regulation in early leukemogenesis.

Session Title: Cancer Poster Session II

PB5135 Polygenic risk score and age at onset in breast cancer: Insights from the Korean population.

Authors:

J. Hwang[1], Y-J. Lee[2], S. Lee[2], T-K. Yoo[2], K. Nguyen[3], J. Lee[2], J. Hopper[3], J. Sung[1]; [1]Seoul Natl. Univ., Seoul, Korea, Republic of, [2]Asan Med. Ctr., Seoul, Korea, Republic of, [3]Univ. of Melbourne, Carlton, Australia

Background: Compared to Europeans, East Asian populations, particularly in South Korea, exhibit a higher incidence of early-onset breast cancer (BC), occurring before 40 years of age. Alarmingly, this trend continues to escalate. Early-onset BC imposes significant challenges to both preventative measures and clinical management due to its poorer prognosis. It is broadly accepted that early-onset BC carries a higher genetic load, especially with regards to high-penetrance genes such as *BRCA1/2* mutations. However, our understanding of potential polygenic burdens in early-onset BC remains incomplete. This study thus sought to explore the association between polygenic risk scores (PRS) and age-dependent risks of BC within the Korean population, using data from Korean Genome and Epidemiology Study (KoGES), a nationwide population-based biobank in Korea.**Methods**: The study population consisted of 46,323 Korean women, including 67 early-onset and 575 late-onset BC cases. We calculated PRS using a validated 313-SNP PRS (PRS313) model derived from European ancestry. Logistic regression analysis was performed to calculate the odds ratios (ORs) per standard deviation (SD) increase in PRS and their corresponding 95% confidence intervals (CIs) across different age groups.**Results and Interpretations**: Our analysis revealed an age-dependent relationship between PRS and BC risk in the Korean population. For the younger age group (age 40 or earlier), each SD increase in PRS was associated with a higher risk of breast cancer (OR=1.49, 95% CI: 1.17-1.91) compared to the older age group (OR=1.39, 95% CI: 1.28-1.51). Furthermore, comparing the highest PRS quintile with the lowest, BC risk in the younger age group increased 3.63-fold (95% CI: 1.64-9.16), while it increased 2.69-fold (95% CI: 2.04-3.60) in the older group.This study demonstrates the significance of considering PRS in understanding the age-related risk of breast cancer within the Korean population. Our findings suggest that preventative strategies for early-onset BC would be more effective if they incorporate the patients' polygenic profiles.

Session Title: Cancer Poster Session III

PB5136 Population specific breast cancer risk variants associated with estrogen-receptor negative breast cancer in women of African ancestry on chromosome 3q.

Authors:

H. M. Ochs-Balcom[1], L. Preus[1], Z. Du[2,3], R. C. Elston[4], G. Jia[5], X. Guo[5], Q. Cai[5], J. Long[5], J. Ping[5], M. K. Bolla[6], J. Dennis[6], A. M. Dunning[7], D. F. Easton[7], J. Palmer[8], M. Garcia-Closas[9], D. Huo[10], W. Zheng[5], C. Haiman[3], The Ghana Breast Health Study Team, The African-ancestry Breast Cancer Genetic Consortium; [1]Dept. of Epidemiology and Environmental Hlth., Sch. of Publ. Hlth.and Hlth.Professions, Univ. at Buffalo, Buffalo, NY, [2]Publ. Hlth.Sci., Fred Hutchinson Cancer Res. Ctr., Seattle, WA, [3]Dept. of Preventive Med., Keck Sch. of Med. and Norris Comprehensive Cancer Ctr., Univ. of Southern California, Los Angeles, CA, [4]Dept. of Population and Quantitative Hlth.Sci., Case Western Reserve Univ., Cleveland, OH, [5]Div. of Epidemiology, Dept. of Med., Vanderbilt Epidemiology Ctr., and Vanderbilt-Ingram Cancer Ctr., Vanderbilt Univ. Sch. of Med., Nashville, TN, [6]Ctr. for Cancer Genetic Epidemiology, Dept. of Publ. Hlth.and Primary Care, Univ. of Cambridge, Cambridge, United Kingdom, [7]Ctr. for Cancer Genetic Epidemiology, Dept. of Oncology, Univ. of Cambridge, Cambridge, United Kingdom, [8]Slone Epidemiology Ctr., Boston Univ., Boston, MA, [9]Div. of Cancer Epidemiology and Genetics, Dept. of Hlth.and Human Services, Natl. Cancer Inst., NIH, Bethesda, MD, [10]Dept. of Publ. Hlth.Sci., Univ. of Chicago, Chicago, IL

Background. Expansion of genome-wide association studies across population groups is needed to improve our understanding of shared and unique genetic contributions to breast cancer. We performed an association study guided by *a priori* linkage findings from African ancestry (AA) relative pairs. Methods. We performed a fixed-effect inverse-variance weighted meta-analysis under a significant AA breast cancer linkage peak on chr3q26-27 in 9,241 AA cases and 10,193 AA controls. We examined associations with risk of overall breast cancer as well as estrogen receptor (ER)-positive and negative subtypes (47,737 SNPs in total). We performed validation of chr3 rare variant associations using whole genome sequencing (WGS) data obtained from 1340 cases and 675 controls included in the African-ancestry Breast Cancer Genetic Consortium (AABCG). Results. In our chr3 region, we identified two associations with ER-negative disease (rs184090918, OR=3.70, $p=1.23x10^{-5}$; rs76959804, OR=3.57, $p=1.77x10^{-5}$) that exceeded the multiple testing significance threshold. In the AABCG validation study, the SNP associations were significant and effect sizes were larger (rs184090918, OR: 6.66, 95% CI: 1.43, 31.01; rs76959804, OR: 5.24, 95% CI: 1.70, 16.16). These two SNPs are upstream to open chromatin ENSR00000710716, a regulatory feature that is actively regulated in mammary tissues. Conclusions. Chromatin state data show it is strongly transcribed and enhanced in mammary tissue, providing evidence that variants in this chr3 region may have a regulatory role in our target organ. These two identified chr3 SNPs, which are monoallelic in non-African ancestry population groups, may represent population specific breast cancer risk variants. Further functional characterization at this locus is needed. We are conducting additional analyses to replicate the findings. Our study provides support for the discovery of breast cancer risk variants using prioritization of variants based on linkage evidence.

Session Title: Cancer Poster Session I

PB5137 Preclinical evaluation of JAK2 specific investigational oligonucleotide for the treatment of MDS/PV

Authors:

B. Przychodzen; Vanda Pharmaceuticals, Washington DC, WA

Myeloproliferative disorders (MPD) are clonal hematopoietic stem cell malignancies with cytokine independency or hypersensitivity. Polycythemia vera (PV), an acquired MPD characterized by increased blood cell mass and hematocrit and leukocyte count, is associated with incidental myelofibrosis (MF). MDS is characterized by cytopenia and the presence of morphological dysplasia of precursor and mature bone marrow blood cells. PV and MDS leave patients at risk for progression to acute myeloid leukemia.

Abnormal cytokine signaling due to an aberrant JAK2-STAT pathway has a vital role in PV and MDS pathogenesis. JAK2 mutations can result in hematologic malignancies, where hyperactive signaling of the JAK2-STAT pathway promotes tumor cell proliferation, invasion, and angiogenesis. Increased JAK2 kinase activity is observed in hematologic malignancies; somatic $JAK2^{V617F}$ gain-of-function mutations are found in at least 95% of PV patients and is implicated in MDS cases. We hypothesize that preventing JAK2 transcription by ASOs mediated exon masking of the *JAK2* intron-exon junction will result in reduced *JAK2* mRNA, and thus JAK2 protein, by providing nonsense mediated decay (NMD) in the reading frame.

This study utilized a HEL cell line harboring the V617F *JAK2* gain-of-function mutation. We designed a series 19mer ASO targeting *JAK2* exon-intron junctions (*JAK2* intron-exon junction providing NMD in the reading frame, steric, non-RNAse H1) and tested these at a range of concentrations. The ASOs were designed with a PS 2'-O-methoxyethyl backbone and prioritized based on *in silico* binding affinity and limited off-target binding. HEL cells underwent ASO treatment (1µM) and 72-hour incubation.

We observed significant JAK2 protein decrease (~50%) in ASO-treated samples compared to untreated samples. JAK2 qPCR results confirmed 40-60% of target transcript. STAT5 phosphorylation status further confirmed this effect, and we report a 35% pSTAT5 reduction. Furthermore, this ASO showed limited off-target effects *in silico*. siRNA and CRISPR knockout lysate were used as controls. Our preclinical data support this ASO as a highly specific JAK2 agent, affecting direct levels of JAK2 as well as downstream STAT signaling.

Around 50-60% of primary MF patients harbor the $JAK2^{V617F}$ gain-of-function mutation. ASOs offer the capability to directly target *JAK2* mutations with high precision and effectively reduce JAK2 protein production, without off-target kinase effects. The ability to reduce JAK2 protein may alleviate the disease burden that patients with hematologic malignancies face, resulting in a higher quality of life through prevention and treatment.

Session Title: Cancer Poster Session II

PB5138 Prevalence of germline pathogenic variants in cancer predisposition genes in a population-based study of renal cell carcinoma

Authors:

P. Harraka[1], F. Bruinsma[2,3,4], G. Severi[5,6], S. Jordan[7], T. Dudding[8], D. Park[9,10], B. Pope[1,9,11,12], J. Steen[1], R. Milne[1,3,4], G. Giles[1,3,4], I. Winship[13,14], K. Tucker[15], M. Southey[1,3,16], T. Nguyen-Dumont[1,3,16]; [1]Precision Med., Sch. of Clinical Sci. at Monash Hlth., Monash Univ., Clayton, Australia, [2]Burnet Inst., Melbourne, Australia, [3]Cancer Epidemiology Div., Cancer Council Victoria, Melbourne, Australia, [4]Ctr. for Epidemiology and Biostatistics, Melbourne Sch. of Population and Global Hlth., The Univ. of Melbourne, Parkville, Australia, [5]Université Paris-Saclay, UVSQ, Inserm, CESP U1018, "Exposome & Heredity" team, Gustave Roussy, Villejuif, France, [6]Dept. of Statistics, Computer Sci. & Applications "G. Parenti" (DISIA), Univ. of Florence, Florence, Italy, [7]Sch. of Publ. Hlth., The Univ. of Queensland, Brisbane, Australia, [8]Sch. of Med. and Publ. Hlth., The Univ. of Newcastle, Newcastle, Australia, [9]Melbourne Bioinformatics, The Univ. of Melbourne, Melbourne, Australia, [10]Dept. of Biochemistry and Pharmacology, Univ. of Melbourne, Melbourne, Australia, [11]The Univ. of Melbourne Ctr. for Cancer Res., Victoria Comprehensive Cancer Ctr., Melbourne, Australia, [12]Dept. of Surgery (Royal Melbourne Hosp.), Melbourne Med. Sch., Faculty of Med., Dentistry and Hlth.Sci., The Univ. of Melbourne, Melbourne, Australia, [13]Dept. of Med., the Univ. of Melbourne, Melbourne, Australia, [14]Genomic Med. and Family Cancer Clinic, Royal Melbourne Hosp., Parkville, Australia, [15]Hereditary Cancer Ctr., Nelune Comprehensive Cancer Ctr., Prince of Wales Hosp., Sydney, Australia, [16]Dept. of Clinical Pathology, Melbourne Med. Sch., The Univ. of Melbourne, Parkville, Australia

Introduction Renal cell carcinoma (RCC) is the most common form of kidney cancer with a mortality rate of 25% within 5 years. RCC is a feature of several heritable (autosomal dominant) cancer syndromes and has been associated with germline pathogenic and likely pathogenic (PLP) variants in at least 21 susceptibility genes. PLP variants in these genes are clinically actionable in Australia and germline panel sequencing is available to individuals with a family history of kidney cancer or a personal history suggestive of a cancer syndrome. Studies that have screened population-based RCC cases with a panel of cancer predisposition genes have identified a high prevalence of PLP germline variants (6-26%), a proportion of which (1-4%) were identified in established RCC susceptibility genes. This study assessed the prevalence of PLP variants in renal cancer predisposition genes actionable in the Australian setting in a population-based study of RCC.
Method Germline DNA was obtained from 1,034 individuals diagnosed with RCC, recruited through the Victoria and Queensland cancer registries into the population-based Consortium for the Investigation of Renal Malignancies (CONFIRM) study. A custom amplicon-based panel of 21 genes was used to sequence participants. Variants were classified as PLP according to ClinVar (2 stars) or manually classified according to current ACMG recommendations.
Results Eighteen participants (2%) were found to carry a PLP variant. The mean age and standard deviation at diagnosis of RCC was 58 ± 10 years and eleven (61%) were men. The gene with the most PLP variants identified was *MITF* with five individuals carrying c.952G>A;(p.Glu318Lys). Other genes with PLP variants included *FLCN*, *FH*, *MSH6*, *VHL*, *BAP1*, *SDHB* and *TSC1*. No PLP variants were identified in *CDC73*, *EPCAM*, *MET*, *MLH1*, *MSH2*, *PMS2*, *POLD1*, *POLE*, *PTEN*, *SDHC*, *SDHD*, *TP53* or *TSC2*.
Discussion PLP variants were identified in eight established RCC predisposition genes. Our findings are consistent with previous studies that, when looking only at genes with a proven association with RCC, identified PLP variants in 1-4% of a population-based sample of RCC. This is likely an underestimation of the true prevalence of PLP variants due to the difficulty in interpreting the many variants of uncertain significance identified in this study, the experimental design not enabling the detection of deep intronic, UTR, structural and copy-number variants, and it being likely that further RCC susceptibility genes remain to be identified. Further exploration of the clinical utility of gene panel susceptibility testing for all RCCs is warranted.

Session Title: Cancer Poster Session III

PB5139 Prevalence of Germline Pathogenic Variants in Patients with Pancreatic Cancer: Results from a Single-Center Prospective Cohort Study in Mexico

Authors:

M. Mendizábal Rodríguez[1], B. Aceves-Poveda[2], P. Nuñez-Martínez[2], Y. Sánchez-Contreras[2], V. Fragoso-Ontiveros[2], M. Herrera-Martínez[3], M. De la Fuente-Hernández[2], R. Álvarez-Gómez[2,4]; [1]Natl. Inst. of Pediatrics, Mexico city, Mexico, [2]Hereditary Cancer Clinic, Natl. Cancer Inst., Mexico city, Mexico, [3]Med. Oncology Dept., Natl. Cancer Inst., Mexico city, Mexico, [4]Fellow of general directorate of quality and education in Hlth., Ministry of Hlth., Mexico City, Mexico

Background/Purpose. Pancreatic cancer (PC) represents the thirteenth cause of cancer in the Mexican population, with a reported incidence of 2.5%. According to other series, between 4-20% of the patients with PC harbor a germline pathogenic variant (PV) in genes associated with hereditary cancer susceptibility syndromes. Prevalence of PV in PC is unexplored in the Mexican population, for which we aimed to describe the clinical features and prevalence of PV in a group of Mexican patients with PC.

Methods. At the Hereditary Cancer Clinic, of the National Institute of Cancerology in Mexico City, we conducted an observational prospective study. We included patients with PC from February 2016 to October 2022 who meet criteria for genetic testing. A complete cancer risk assessment was conducted, highlighting the oncology background. A peripheral blood sample was obtained, and next generation sequencing (NGS) was performed on the Illumina commercial platform (Illumina, SD, USA), using two panels of 263 and 322 genes, associated with increased risk of cancer. In this study, deleterious or suspected deleterious variants were classified as PV. Variants for which the clinical significance could not be determined were classified as a Variant of Uncertain Significance (VUS). Data was collected in SPSS v25, and descriptive statistics was applied to describe the characteristics of the population. The mean was calculated with the statistical program R (version 3.6.3).

Results. A total of 38 patients were included in our cohort, 26(68.4%) were females and 12(31.6%) were males. Diagnosis of pancreatic cancer was present in 30(78.9%) patients, 5(13.2%) bile duct and 3(7.9%) ampullary. The most common histology was adenocarcinoma in 33 patients (81.3%). The mean age at diagnosis was 56 years. Positive cancer family history was present in 24 patients (59.4%). Molecular testing identified 7 patients (18.4%) harboring a PV, of which 4(57%) were doble heterozygous with an additional VUS. In addition, in 18(47.4%) of patients a VUS was identified and 13(34.2%) had negative results for both; PV and VUS. Among the patients with a PV, 4(57.1%) were females, with family history of cancer. The distribution by gene in the patients with a PV was: *BRCA2* in 3(42.9%) cases, *BRCA1* in 2(28.6%), *MRE11* and *ATM* 1(14.2%).

Conclusions. In 18.4% of the cases, we identified a PV in genes associated to an increased risk of PC. This finding is similar to the rates reported in other countries. To our knowledge, this is the first descriptive study that explores this area in the Hispanic Population and thus provide significative evidence of the prevalence of PV in underrepresented populations.

Session Title: Cancer Poster Session I

PB5140 Prioritizing candidate cancer driver mutations in ovarian cancer susceptibility regions.

Authors:

P-H. Lai[1], J. Tyrer[2], P. Pharaoh[1], S. A. Gayther[3], M. R. Jones[3], P-C. Peng[1]; [1]Dept. of Computational and Biomedicine, Cedars Sinai Med. Ctr., West Hollywood, CA, [2]CR-UK Dept. of Oncology, Univ. of Cambridge, Strangeways Res. Lab., Cambridge, United Kingdom, [3]Ctr. for Bioinformatics and Functional Genomics, Dept. of BioMed. Sci., Cedars Sinai Med. Ctr., Los Angeles, CA

Ovarian cancer genetics research has traditionally focused on germline or somatic mutations independently. Emerging evidence suggests that the somatic mutational landscape can be shaped by the germline genetic background. However, the extent of the interplay between somatic mutations and germline variants is not well understood. In this study, we aim to unravel the role of somatic mutations within the ovarian cancer susceptibility regions, by incorporating functional annotations.

We curated 33 ovarian cancer germline susceptibility regions identified from genome-wide association studies (GWAS) and integrated them with somatic mutation data from the Pan-Cancer Analysis of Whole Genomes (PCAWG) project (n=110). We first asked whether ovarian cancer susceptibility genes are associated with somatic mutational signatures. We found significant associations of genes such as RCCD1 and PANK1 with age-related mutational signatures (p-value=0.004). Since most germline risk regions are located in the non-coding genome, we annotated the genome with 28 cell-type agnostic annotations from the ENCODE and Roadmap Epigenomics projects. To identify frequently mutated elements in these germline risk regions, we applied a statistical model called ActiveDriverWGS. Our results revealed that among the 33 risk regions, the 8q24.21 locus has frequently mutated elements in the coding region (p-value=0.011), while the 8q21.13 locus shows significant somatic mutation enrichment in the promoter region (p-value=0.028). Given that tissue-specific annotations better elucidate disease pathogenesis in the non-coding genome, we incorporate seven chromatin states (active/weak promoter/enhancer, active region, insulator, transcribed region) from seven ovarian precursor or cancer cell lines. We identified that weak enhancer chromatin state is frequently mutated in 14 out of 33 ovarian cancer risk regions (average p-value=0.007). In conclusion, by combining germline and somatic mutations with regulatory and epigenomics data, we provide insights into the genetic basis of somatic mutations in the ovarian cancer susceptible regions, and prioritize candidate driver mutations.

Session Title: Cancer Poster Session II

PB5141 Prostate cancer polygenic risk score associated with risk of upgrading and prostate tumor features in men on active surveillance.

Authors:

L. B. Goss[1,2], M. Liu[2], Y. Zheng[2], B. Guo[1], D. V. Conti[3], C. A. Haiman[3], W. J. Catalona[4], J. S. Witte[5], D. W. Lin[1,2], L. F. Newcomb[1,2], B. F. Darst[2]; [1]Univ. of Washington, Seattle, WA, [2]Fred Hutchinson Cancer Ctr., Seattle, WA, [3]Univ. of Southern California, Los Angeles, CA, [4]Northwestern Univ. Feinberg Sch. of Med., Chicago, IL, [5]Stanford Univ., Stanford, CA

BACKGROUND: Recent attention has been drawn to the potential clinical utility of polygenic risk scores (PRS) in prostate cancer screening and disease management. However, the impact of PRS on prostate tumor features is not well characterized. In a case-only study of men on active surveillance, we investigated the implications of PRS on prostate tumor features, particularly risk of upgrading.
METHODS: This study included 1,041 non-Hispanic White men with prostate cancer on active surveillance in the Canary Prostate Active Surveillance Study (PASS). A previously developed multi-ancestry PRS of 451 prostate cancer risk variants was constructed in PASS as a weighted sum of the number of risk alleles carried. We also evaluated this PRS after excluding 51 variants previously associated with prostate specific antigen (PSA). Associations between PRS and prostate tumor features, including grade reclassification at follow up and prostate size, PSA, PSA density, AUA symptom score, quality of life urinary score, percent of biopsy cores with cancer (as a proxy measure of tumor multifocality), and Gleason grade at diagnosis, were conducted using regression models adjusted for age and principal components of ancestry.
RESULTS: Each SD unit increase in the 451 PRS was associated with a 20% increased risk of grade reclassification (95% CI = 1.05 - 1.36, P = 6.1x10-3), while the 400 PRS was associated with a 25% increased risk (95% CI = 1.10 - 1.42, P= 6.2x10-4). The 451 PRS was also associated with 1.37 higher percentage points of biopsy cores with cancer (95% CI = 0.74 - 1.99, P = 1.8x10-5), 0.26 ng/mL higher PSA (95% CI=0.08-0.44, P=6.0x10-3), and increased PSA density (beta= 0.01; 95% CI=0.01- 0.02, P = 1.9x10-5). The 400 PRS was associated with a larger percentage of positive biopsy cores (beta = 1.43; 95% CI = 0.82 - 2.05, P = 6.0x10-6) and had a weaker association with PSA (beta = 0.18; 95% CI = -0.01 - 0.36, P = 0.06) and PSA density (beta = 0.01; 95% CI = 0.01 - 0.02, P = 7.3x10-5). Although the 451 PRS was not associated with prostate size (P = 0.27), the 400 PRS was associated with a 1.63 cc smaller prostate size (95% CI = -3.03 - -0.23, P = 0.02). Neither PRS was associated with AUA symptom scores, quality of life urinary scores, or Gleason grade at diagnosis (P>0.12).
CONCLUSION: In men on active surveillance, high PRS was associated with risk of upgrading and possibly tumor multifocality. Excluding PSA variants from the PRS enhanced these associations and revealed an association with smaller prostate size, which has been previously associated with more aggressive tumors in men with comparable serum PSA levels. Our findings suggest that men with high PRS may benefit from more intensive surveillance.

Session Title: Cancer Poster Session III

PB5142 Rare genetic determinants of clonal hematopoiesis and progression to hematologic malignancies in 479,117 individuals

Authors:

J. Liu[1], B. Wiley[1], I. Chan[1], D. Tran[1], A. Batchi-Bouyou[1], X. Zong[1], L. Ding[1], Y. Cao[1], M. Walter[1], N. Chatterjee[2], K. Offit[3], L. Godley[4], D. Link[1], Z. Stadler[3], K. Bolton[1]; [1]Washington Univ. Sch. of Med. in St. Louis, St. Louis, MO, [2]Johns Hopkins Univ., Gaithersburg, MD, [3]Mem. Sloan Kettering Cancer Ctr., New York, NY, [4]The Univ. of Chicago, Chicago, IL

Clonal hematopoiesis (CH), while common with aging, confers a high relative risk of hematologic malignancies (HM). Genome-wide association studies have identified common germline predisposition loci for CH. Here we sought to characterize the contribution of rare germline mutations to CH and HM. We used the UK Biobank (UKBB) as a discovery cohort (N=454,859) with validation in the Memorial Sloan Kettering (MSK) IMPACT and Cancer Genome Atlas (TCGA) cohorts (N=24,258). We profiled whole exome sequencing (WES) for pathogenic/likely pathogenic germline variants (PGVs) in 241 known/hypothesized cancer predisposition genes. We simultaneously analyzed WES for CH driven by single nucleotide variants/indels and SNP array data for copy number events (mCAs). Overall, 8.9% of individuals in the UKBB harbored PGVs (5.2% if restricted to HM-related genes). Heterozygous PGVs in 15 genes (8 of which are novel) conferred a significantly increased risk of CH. Prominently featured pathways include DNA damage repair/sensing (*TP53, ATM, CHEK2, NBN*), telomere maintenance (*TINF2, CTC1, POT1*) and RAS, JAK/STAT signaling (*KRAS, PTPN11, SAMD9L, SOS1, MPL*). Through pairwise analysis, we identified an additional 102 germline genes that predispose to CH in specific genes or genetic regions; 30 of these CH-specific associations were also significant in our replication cohort ($p<0.05$). Of the remaining 31 with at least one CH positive germline carrier, all but six were directionally consistent. Overrepresented pathways include immune regulation (heterozygous PGVs in *UNC13D* and *DOCK8* associated with several mCAs and *ADA* associated with CH in *DNMT3A*), homologous recombination deficiency (heterozygous PGVs in *FANCA*, *RAD51D*, *BRIP1* and *BRCA1* associated with multiple mCAs) among others. During 14-years of follow-up, 4,596 UKBB participants developed HM. Several germline genes that predispose to CH also were significantly associated with HM (N=18), including several genes not been previously linked to HM in the heterozygous state (*CTC1*, *BLM*, and *WNR*, among others). We observed synergistic effects between germline PGVs and CH on the risk of HM whereby CH carriers with PGVs have a significantly higher risk of developing HM (HR: 1.31; $p<0.001$) compared to those without PGVs. CH in genes/genetic regions that are moderately/strongly associated with specific PGVs confer higher risk of progression to HM than CH showing weak or no association with PGVs. In summary, we identify several novel genes where PGVs contribute to CH predisposition and to its progression to HM. Our findings highlight the importance of germline-CH interactions in determining the risk of CH progression to HM.

Session Title: Cancer Poster Session I

PB5143 Re-classification of *PMS2* Kozak sequence variants using a fast and easy in vitro assay.

Authors:

E. Matoy, J. Plowman, B. Michael, H. Feser Stessman; Creighton Univ., Omaha, NE

Lynch Syndrome (LS) is the most common hereditary cancer syndrome affecting 1 in 279 people. Heterozygous loss-of-function variants in *PMS2* are linked to LS. While these variants are not directly cancer initiating, PMS2 functions in the DNA mismatch repair pathway such that reduced function results in the accumulation of somatic variants and increased cancer risk over time. We recently identified a suspected hereditary cancer family carrying a 5' UTR variant in *PMS2* (ENST00000265849.7:c.-7T>C; MAF: 0.00032 across all races) that had conflicting interpretations of pathogenicity in ClinVar. This variant lies within the Kozak sequence of *PMS2*, the well-described protein translation initiation motif in eukaryotes. This sequence is highly conserved among higher eukaryotes and is defined as the 9 basepairs upstream of the translation start codon through the first 4 bases of the translated sequence (5'-GTTGCATCCATGG-3'; human *PMS2* NM_000535.7). Variants in the Kozak sequence have been shown to reduce the translation of genes such as those necessary for heart development resulting in congenital heart disease. However, all Kozak sequence variants in *PMS2* in ClinVar are currently classified as variants of undetermined significance (VOUSs) due to a paucity of research on this non-coding region. We hypothesized that variants that significantly disrupt the Kozak sequence motif of *PMS2* would decrease PMS2 protein expression contributing to increased cancer risk over time. Using the psiCHECK™-2 Vector (Promega) and site-directed mutagenesis, we modified the endogenous Renilla luciferase Kozak sequence (but not the Firefly luciferase sequence) to the human *PMS2* sequence. A second round of mutagenesis was similarly used to generate ClinVar *PMS2* Kozak sequence variants. Plasmids (wild-type human *PMS2* or *PMS2* Kozak variants) were individually transfected into HEK 293 TK cells, grown for 48 hours, and translation efficiency was determined as the ratio of Renilla/Firefly expression using the Dual-Glo® Luciferase Assay (Promega). In the case of the c.-7T>C Kozak sequence variant, we identified no significant difference in activity between this variant and the wild-type human *PMS2* sequence supporting a benign clinical classification. In summary, we present a novel method for the high-throughput classification of human *PMS2* Kozak sequence variants that can contribute to the re-classification of VOUSs identified in human patients.

Session Title: Cancer Poster Session II

PB5144 Risk of second primary cancers after a diagnosis of first primary cancer: a pan-cancer analysis and mendelian randomization study

Authors:

Y. Na[1], R. Xiaohao[2], H. Da[2]; [1]The Univ. of Hong Kong, Hong Kong, China, [2]Shanghai Jiao Tong Univ. Sch. of Med. affiliated Ruijin Hosp., Shanghai, China

Background: The risk of second primary cancers (SPC) is increasing after the first primary cancers (FPC) are diagnosed and treated. The underlying causal relationship remains unclear. **Methods**: We conducted a pan-cancer association (26 cancers) study in the Surveillance, Epidemiology, and End Results (SEER) database. The standardized incidence ratio (SIR) was estimated as the risk of SPCs in cancer survivors based on the incidence in the general population. Furthermore, the causal effect was evaluated by two-sample Mendelian Randomization (MR, 13 FPCs) in the UK Biobank (UKB, n=459,136) and robust analysis (radial MR and Causal Analysis Using Summary Effect estimates, CAUSE). **Results**: We found 11 significant cross-correlations among different cancers after harmonizing SIR and MR results. Whereas only 4 of them were confirmed by MR to have a robust causal relationship. In particular, patients initially diagnosed with oral pharyngeal cancer would have an increased risk of non-Hodgkin lymphoma (SIR_{SEER}=1.18, 95%Confidence Interval [CI]:1.05-1.31, $OR_{radial\text{-}MR}$=1.21, 95%CI:1.13-1.30, *P*=6.00×10-3; OR_{cause}=1.17, 95%CI:1.05-1.31, *P*=8.90×10-3). Meanwhile, ovary cancer was identified to be a risk factor for soft tissue cancer (SIR_{SEER}=1.72, 95%CI:1.08-2.60, $OR_{radial\text{-}MR}$=1.39, 95%CI:1.22-1.58, $P=1.07\times10^{-3}$; OR_{cause}=1.36, 95%CI:1.16-1.58, *P*=0.01). And kidney cancer was likely to cause the development of lung cancer (SIR_{SEER}=1.28, 95%CI:1.22-1.35, $OR_{radial\text{-}MR}$=1.17, 95%CI:1.08-1.27, *P*=6.60×10-3; OR_{cause}=1.16, 95%CI:1.02-1.31, *P*=0.05) and myeloma (SIR_{SEER}=1.54, 95%CI:1.33-1.78, $OR_{radial\text{-}MR}$=1.24, 95%CI:1.21-2.45, *P*=0.02; OR_{cause}=1.49, 95%CI:1.04-2.34, *P*=0.02). **Conclusions**: A certain type of primary cancer may cause another second primary cancer, and the profound mechanisms need to be studied in the future.

Session Title: Cancer Poster Session III

PB5145 scLongTree: a novel highly-accurate and scalable computational tool to infer the longitudinal tree from longitudinal scDNAseq data

Authors:

R. Khan, X. Mallory; Florida State Univ., Tallahassee, FL

Cancer grows by acquiring new mutations at different stages. At each stage, there could be multiple subclones of tumor cells coexisting in one tumor, a notoriously difficult-to-treat problem called intra-tumor heterogeneity (ITH). It is important to gain the knowledge of which set of mutations is acquired at each specific time point so that personalized treatment plans, prognosis, and diagnosis can be made. Single-cell DNA sequencing (scDNAseq) makes it possible to decipher ITH. Moreover, the decreasing cost of scDNAseq in recent years makes it possible to sequence cancer patients once every few years, which eventually renders longitudinal data, the data corresponding to one sample but is available for multiple time points.

Given the scDNAseq longitudinal data, we aim to recover the longitudinal tree, on which the nodes represent subclones of cells, and edges represent the new mutations gained on the child nodes. As far as we know, LACE is the only existing method that can infer such a longitudinal tree given the scDNAseq longitudinal data. However, LACE is limited in terms of scalability. It does not scale up to the case when the number of timepoints > 4.

We developed a novel highly-accurate and scalable computational tool, scLongTree, that infers a longitudinal tree from scDNAseq. Different from LACE, scLongTree clusters single cells across all time points before inferring the tree, greatly reducing the computational complexity of inferring the tree. The cell-clustering results are further refined based on the knowledge of time points using a maximum likelihood approach. We then infer the parent-child relationship among the clusters of cells, now viewed as subclones, between two consecutive time points. For the case when two sequencing timepoints are far away from each other and an important subclone has not been sequenced before it evolved into different subclones, scLongTree will also infer such unobserved subclones.

We compared scLongTree with four other methods, LACE, SiCloneFit, SCITE and CALDER on nine simulated datasets, each with a varying parameter. On the default dataset when the number of timepoints is 4, scLongTree achieved the highest accuracy (0.979), to be compared with LACE (0.648), SiCloneFit (0.737), SCITE (0.778), and CALDER (0.558). Such a highest accuracy is gained by using the least running time. In addition, unlike LACE, scLongTree is scalable to the datasets with > 4 timepoints. We also applied scLongTree to real data samples, one of which is SA501, a triple-negative breast tumor that has longitudinal targeted scDNAseq data. ScLongTree was able to correctly place 17 mutations out of 20 mutations on the inferred longitudinal tree.

Session Title: Cancer Poster Session I

PB5146 Search for high-risk melanoma susceptibility genes: a GenoMEL project combining data from 731 families

Authors:

A. Goldstein[1], W. Zhu[1], P. A. Johansson[2], F. Wang[3], F. M. Ortuño[4], S. Li[3], V. Iyer[5], K. Jain[1,6], R. van Doorn[7], V. Höiom[8,9], P. A. Kanetsky[10], D. Pjanova[11], V. Suybeng[12], B. Bressac-de Paillerets[12,13], X. R. Yang[1], T. Dębniak[14], P. Aguilera[15,16,17], O. Cabaret[12], A. E. Cust[18,19], D. E. Elder[20], N. Gruis[7], H. Helgadottir[8,9], K. M. Hernandez[3], E. A. Holland[18,19,21], M. Kreishmane[22], M. Potrony[17,23], M. R. Sargen[1], H. Schmid[18,19,21], M. A. Tucker[1], Z. Zhang[3], S. Puig[15,16,17], E. Nagore[24], D. Calista[25], P. Ghiorzo[26], C. Menin[27], M. C. Fargnoli[28], K. Peris[29], G. J. Mann[19,21,30], B. Hicks[1,6], C. McCabe[31], E. Stahlberg[31], R. L. Grossman[3,32,33], M. T. Landi[1], N. K. Hayward[2], D. J. Adams[5], GenoMEL Exome Sharing Consortium; [1]Div. of Cancer Epidemiology and Genetics, Natl. Cancer Inst., Bethesda, MD, [2]QIMR Berghofer Med. Res. Inst., Brisbane, Australia, [3]Ctr. for Translational Data Sci., Univ. of Chicago, Chicago, IL, [4]Univ. of Granada, Granada, Spain, [5]Experimental Cancer Genetics. Wellcome Sanger Inst., Cambridge, United Kingdom, [6]Cancer Genomics Res. Lab., Leidos BioMed. Res., Inc., Frederick Natl. Lab. for Cancer Res., Frederick, MD, [7]Dept. of Dermatology, Leiden Univ. Med. Ctr., Leiden, Netherlands, [8]Dept of Oncology and Pathology, Karolinska Inst.t, Stockholm, Sweden, [9]Karolinska Univ. Hosp., Stockholm, Sweden, [10]Dept. of Cancer Epidemiology, Moffitt Cancer Ctr. and Res. Inst., Tampa, FL, [11]Riga Stradins Univ., Riga, Latvia, [12]Gustave Roussy, Département de Biopathologie, Villejuif, France, [13]INSERM U1279, Tumor Cell Dynamics, Villejuif, France, [14]Dept. of Genetics and Pathology, Intl. Hereditary Cancer Ctr., Pomeranian Med. Univ., Szczecin, Poland, [15]Dermatology Dept., Hosp. Clínic, Barcelona, Spain, [16]Univ. of Barcelona, Barcelona, Spain, [17]CIBERER , Inst. de Salud Carlos III, Barcelona, Spain, [18]The Daffodil Ctr., The Univ. of Sydney, Sydney, Australia, [19]Melanoma Inst. Australia, The Univ. of Sydney, Sydney, Australia, [20]Dept. of Pathology and Lab. Med., Hosp. of the Univ. of Pennsylvania, Philadelphia, PA, [21]Ctr. for Cancer Res., The Westmead Inst. for Med. Res., Westmead, Australia, [22]Latvian BioMed. Res. and Study Ctr., Riga, Latvia, [23]Biochemistry and Molecular Biology Dept., Hosp. Clínic, Barcelona, Spain, [24]Inst. Valenciano de Oncología, Valencia, Spain, [25]Maurizio Bufalini Hosp., Cesena, Italy, [26]Univ. of Genoa, Genoa, Italy, [27]Veneto Inst. of Oncology, Padua, Italy, [28]Univ. of L'Aquila, L'Aquila, Italy, [29]Catholic Univ. of Rome, Rome, Italy, [30]John Curtin Sch. of Med. Res., Australian Natl. Univ., Canberra, Australia, [31]Ctr. for BioMed. Informatics and Information Technology, Natl. Cancer Inst., Bethesda, MD, [32]Dept. of Med., Univ. of Chicago, Chicago, IL, [33]Dept. of Computer Sci., Univ. of Chicago, Chicago, IL

CDKN2A is the major high-risk cutaneous melanoma (CM) susceptibility gene with pathogenic variants (PVs) detected in 20-40% of CM-prone families. Multiple much rarer high-risk CM genes have also been discovered (*CDK4*, *BAP1*, *POT1*, *ACD*, *TERF2IP*, *TERT*). Identifying new CM genes has been challenging, thus, we initiated a cloud-based project within the Melanoma Genetics Consortium GenoMEL to combine raw next-generation sequencing (NGS) data and use uniform workflows to identify novel CM risk genes. Groups (n=17) from nine countries (Australia, France, Italy, Latvia, Netherlands, Poland, Spain, Sweden, USA) contributed data. After quality control, there were 1255 subjects (731 families) including 315 multiple-case families (839 CM patients) of which 119 families had >=3 sequenced CM cases. To search for high-risk CM genes, we combined evidence for co-segregation in the families with the web-based tool VarElect (ve.genecards.org), a comprehensive phenotype-dependent variant/gene prioritizer, which leverages the wealth of information in GeneCards and its affiliated databases. In the 315 multiple-case families, there were 233,550 non-synonymous variants with minor allele frequency (MAF)<0.0005. We used families with >=3 CM cases sequenced to identify potential candidate genes and detected 1023 genes based on evidence for co-segregation that required variant(s) to be coded as high/moderate impact and MAF<0.001. There was no new *CDKN2A*-like gene observed; that is, no new high-risk gene at a frequency as is seen with *CDKN2A*. Based on a VarElect score >6.5 and multiple families (any size) with supportive evidence for co-segregation, 16 candidate genes were selected for further evaluation of association with CM in the United Kingdom Biobank (UKBB) database. For UKBB, variants (MAF<0.01, 0.005, or 0.001) were classified as pathogenic based on ClinVar and/or InterVar, or if predicted to be damaging by most in silico algorithms inspected. Among the 16 genes examined using logit regression, gene-based rare PVs in *TYR* (p<0.0003), *POLH* (p<0.001) and *FLCN* (p<0.02) were associated with CM in UKBB. Further, individual variants in *FLCN* (c.611C>T, p.Ala204Val), *TYR* (c.1118C>A, p.Thr373Lys), and *BMPR1A* (c.943G>A, p.Gly315Arg) also showed associations with CM in UKBB (p<0.001). In summary, we investigated NGS data from 731 CM families using uniform workflows. Several candidate high-risk CM genes were uncovered and replicated in UKBB but no new

frequent *CDKN2A*-like gene was identified. Further study is required to confirm the findings and to assess whether the genetic risks in many of these CM families result from moderate- or low-risk genes rather than high-risk genes.

Session Title: Cancer Poster Session II

PB5147 Severus: a computational tool to characterize complex structural variations from long-read sequencing.

Authors:

A. Keskus[1], A. Tanveer[1], A. Donmez[1], A. Bryant[1], I. Rodriguez[1], N. Rossi[1], Y. Xie[1], B. Yoo[2], R. Milano[1], H. Lou[1], J. Park[3], J. Gardner[3], B. McNulty[3], K. Miga[3], M. Farooqi[2], B. Paten[3], M. Dean[1], M. Kolmogorov[1]; [1]NIH, BETHESDA, MD, [2]Children's Mercy Hosp, Kansas City, MO, [3]UC Santa Cruz Genomics Inst., Santa Cruz, CA

Most current cancer genomic studies capitalize on small genomic variations (point mutations or small indels) in protein-coding regions to identify diverse, evolutionary trajectories, classify cancers into subtypes, or inform treatment.

Pan-cancer whole-genome sequencing studies revealed diverse classes of structural variants (SVs) in cancer genomes, which may contribute to carcinogenesis and tumor progression through direct modification of coding region as well as alterations in transcriptional regulation, i.e., copy number alterations, enhancer hijacking, or topological domain modification.

Short-read sequencing data excel in small variant discovery but faces limitations in repetitive regions of the genome and the discovery of SV that are longer than the read length. In contrast, long-read sequencing has superior SV sensitivity and precision. However, most current long-read tools were designed for “healthy” genomes, and fail to capture somatic variants, complex rearrangements, and heterogeneity in cancer genomes. Here we present Severus to detect and annotate a wide range of SVs, from simpler indels to complex rearrangements with multiple breakpoints using long-read sequencing. Severus is designed to work in single-sample, somatic (disease vs. normal), and multi-site modes. The algorithm takes bam file(s) as input and, using split alignments, detects a confident set of SVs. Next, SV calls are used to build a breakpoint graph in which complex events with multiple breakpoints form connectivity clusters to characterize the structure of the derived cancer genome. Severus can also use phased bam files to generate haplotype-aware SV calls.

Severus outperforms existing tools in SV recall measured against the multiplatform validated somatic SV set for the COLO829 cell line. Further, we sequenced five cancer cell lines and their matching normals with ONT and compared them to Illumina SV calls. As expected, long-read sequencing substantially improved SV detection. We also show that in addition to indels, Severus can accurately detect known complex rearrangements, chromothripsis, chromoplexy, breakage-fusion-bridge (BFB), inversions with amplifications/deletions, and intrachromosomal long insertions.

We also sequenced HPV-infected cervical cancer cell lines (CaSki, SNU1000, SCC152, HT3), revealing mosaic amplifications near the HPV integration sites. We also detected at least one BFB event per sample, which interestingly were exclusive to chromothripsis events in almost all cases.

Finally, matching tumor/normal sequencing of three clinical cases of two pediatric AML and one LCL with Pacbio HiFi revealed novel SVs in coding sequences.

Session Title: Cancer Poster Session I

PB5149 † Sex differences in transcriptomes: Pan-cancer analysis of The Cancer Genome Atlas (TCGA)

Authors:

C. Printzis[1], L. Pesce[1], C. Stanhope[1], R. Huang[2], B. Stranger[1]; [1]Northwestern Univ., Chicago, IL, [2]Univ Minnesota, Minneapolis, MN

Cancer is a leading cause of death worldwide and biological sex differences exist in incidence, treatment response, and mortality across cancer types. We characterized sex differences between male- and female-derived tumor transcriptomes for 21 non-sex-specific cancers from The Cancer Genome Atlas (TCGA). This enabled the discovery of functions impacted by sex in tumors. We assessed sharing of transcriptomic sex-differences across cancers and compared results to matching tissues from the Genotype-Tissue Expression (GTEx) project. We analyzed 6,430 transcriptomes (N = 3,754 males; N =2,676 females) across 21 TCGA cancers, testing for gene expression sex differences (sex-DE). We identified sex-DE genes ranging from 100 genes in uveal melanoma to 5,103 in liver hepatocellular carcinoma and grouped cancers by the number of identified sex-DE genes into “less-differentiated” and “more-differentiated”. Across the 7 “more-differentiated” cancers, more sex-DE genes are shared, but few are common across all 21. Of 19 sex-DE genes shared across all cancers, 16 are X chromosome genes, e.g., XIST. Sharing of sex-DE genes was quantified with the metric $\pi 1$ to estimate the true positive rate in a replication set. A median $\pi 1$ of 0.45 was found between 7 independent cancer datasets and their matching TCGA cancer. Additionally, $\pi 1$ was estimated for 39 TCGA cancer and normal GTEx tissue pairs. Unsurprisingly, we observed that some of the sex effects on gene expression in normal tissues are not maintained in the matched tumor (median $\pi 1$ GTEx to TCGA; 0.18). Interestingly, we observed many sex-DE genes that were specific to the tumor, i.e., not DE in the matching normal tissue (median $\pi 1$ TCGA to GTEx; 0.26). Gene set enrichment analysis (GSEA) of the sex-DE genes revealed distinct functions and processes enriched in male- and female- biased genes. Furthermore, when comparing GSEA results between TCGA and GTEx, we observed tumor-specific functional enrichment, indicating tumor sex differences in gene expression do not simply reflect the tissue of origin. To explore the relationship between sex, gene expression, disease progression (PFI), and overall survival (OS), we used Cox proportional hazards models to test for gene expression and sex-by-gene expression effects on PFI and OS. We identified 60 unique genes across 4 cancer types with significant sex by gene effects on OS and 13 unique genes across 3 cancer types with significant sex by gene effects on PFI. In summary, our results provide a diverse characterization of sex differences in TCGA tumors which provide novel insights into disease, including progression and survival, and may enable development of new therapies.

Session Title: Cancer Poster Session II

PB5150 Single-cell Transcriptomic Profiling of Colorectal Cancer Tumors Reveals Potential Biomarkers

Authors:

O. Awe[1], B. Kamau[2], R. A. Serage[3], H. Ndugwa[2], B. Onyango[4], J. Wachira[5]; [1]Univ. of Ibadan, Ibadan, Nigeria, [2]Dept. of Biochemistry and Biotechnology, Sch. of Pure and Applied Sci., Pwani Univ., Nairobi, Kenya, [3]South African Natl. Bioinformatics Inst., Faculty of Natural Sci., Univ. of the Western Cape, Cape Town, South Africa, [4]Pwani Univ., Nairobi, Kenya, [5]Dept. of Biochemistry, Faculty of Sci. and Technology, Univ. of Nairobi, Nairobi, Kenya

Colorectal cancer (CRC) is a major cause of death worldwide and its extreme heterogeneity makes it difficult to diagnose in its early stages. Cancer-related biomarkers are crucial for their early detection and management but traditional cancer diagnostic techniques can be biased. Recent advancements in next-generation technologies have led to high-dimensional datasets being generated whose analysis depends on machine learning. In this study, we utilized single-cell transcriptomic datasets coupled with machine learning in identifying CRC biomarkers.
Three scRNA-seq datasets related to CRC were downloaded from the Gene Expression Omnibus (GEO) database and merged for further analysis. Data preprocessing and count matrix generation were performed using the Scanpy toolkit. Two machine learning algorithms (LightGBM and XGBoost) were explored as classifiers to identify genes that can be used as CRC biomarkers. The accuracy, F1 score, and Matthew's Correlation coefficient metrics were used in evaluating the performance of the different models. The Shap Explainer was finally used to extract information about the key genes acting as major contributors to the trained machine learning models and their weights. All methods were automated and made reproducible using Nextflow programming.
LightGBM slightly outperformed XGBoost. We identified two novel biomarkers from eleven potential biomarkers for colorectal cancer. These genes can be further analyzed to identify CRC hub genes.
This study further highlights the combined use of machine learning and scRNA-seq data as an alternative method for finding CRC-related genes. These biomarkers can be used for developing drug targets or therapies for colorectal cancer.

Session Title: Cancer Poster Session III

PB5151 † SoMAS: Finding somatic mutations associated with alternative splicing in human cancers

Authors:

H. Tan[1], D. O. Holland[1], V. Gotea[1], L. Elnitski[2]; [1]NIH, Bethesda, MD, [2]NHGRI, Rockville, MD

Aberrant alternative splicing (AS) is prevalent in cancer and affects most cancer hallmarks involving proliferation, angiogenesis, and invasion. Somatic point mutations, recognized as the major driving force of human oncogenesis, can exert their functions via splicing disruption. We propose SoMAS (Somatic Mutation associated with Alternative Splicing), a computational pipeline that explores the role of somatic mutations in shaping the landscape of alternative splicing at the transcriptome level.

Different from existing methods that examine the direct, one-to-one association between each single nucleotide variant (SNV) and a specific AS event, SoMAS measures the association between each eligible SNV and the overall expression pattern of the whole transcriptome consisting of ~73K annotated transcripts, using the principal component analysis (PCA) technique. We show that SoMAS is more efficient and versatile than existing methods in identification of AS-associated SNVs on several aspects: (1) Reduces the computational complexity significantly; (2) Detects both cis- and trans-regulatory SNVs simultaneously; (3) Lowers the false positive rate incurred by the traditional gene-by-gene association study.

Applying SoMAS to 33 cancer types consisting of 9,738 tumor samples in The Cancer Genome Atlas (TCGA), we identified 7,140 SoMAS genes (i.e., genes whose mutation is significantly associated with differential isoform expression of genes located in either cis- or trans-loci) in human cancers (with trans-associations dominating), among which 908 are present in three or more cancer types. These SoMAS genes include many well-known oncogenes (e.g., the Ras family), tumor suppressors (e.g., TP53, PTEN), RNA binding proteins (e.g., RBM10, RANBP2), splicing factors (e.g., SF3B1, SPEN) and transcription regulators (e.g., CTNNB1, SMARC4A).

Functional analysis further confirmed the biological and clinical significance of these identified SoMAS genes and associated genes. Specifically, the 908 “pan-cancer” SoMAS genes are significantly enriched in cancer cell growth and metastasis related pathways/gene sets, and better stratify patients in survival rate when integrating their mutation status and the isoform expression level of their associated genes. SoMAS genes were widely corroborated to affect gene splicing by previous studies with independent cohorts and/or methodologies.

With SoMAS, we for the first time demonstrate the impact of somatic mutations on the overall splicing and transcription profiles in human cancers and bridge the genetic and epigenetic regulation of human oncogenesis in an innovative and biologically meaningful way.

Session Title: Cancer Poster Session I

PB5152 Stable Overexpression of HDM2 Differentially Modulates the Molecular Phenotype of Non-Small Lung Cancer Cells of Varying p53 Genotypes.

Authors:

Y. Wang, K. Ramos; Texas A&M Inst. of BioSci.s and Technology, Houston, TX

Retrotransposons, particularly Long Interspersed Element-1 (LINE-1), have emerged as significant drivers of genome plasticity in humans. LINE-1, the only autonomous retrotransposon that remains active in the human genome, has been closely linked with poor prognosis of non-small cell lung cancers (NSCLC). Previous studies by our group have reported a marked increase in HDM2 (human homolog of mouse double minute 2) mRNA levels in transformed human bronchial epithelial carrying an intronic LINE-1 insertion in the NACC2 locus. HDM2 possesses E3 ubiquitin ligase activity towards p53 and functions as a negative regulator of p53 via polyubiquitination in transformed cells. However, a recent study reported that following DNA damage, HDM2 can switch from a negative to a positive regulator by binding to p53 mRNA to promote translation of the p53 mRNA. This change in HDM2 activity appears to be governed by ataxia telangiectasia mutated kinase-mediated phosphorylation leading to structural changes that impact the HDM2 interactome. To further define the critical genetic interactions regulated by HDM2 in NSCLC, we stably overexpressed HDM2 in NCI-460 (p53 wt) and NCI-1299 (p53 mut) lung cancer cell lines and subsequently examined the expression of critical molecular targets. HDM2 overexpressionincreased LINE-1 ORF-1p levels in NCI-460 cells, but reduced expression in NCI-1299 cells. p53 levels increased in NCI-460 cells and were undetectable in NCI-1299 cells. RB selectively decreased in NCI-460 cells but remained unchanged in NCI-1299 cells. Together, these findings highlight key molecular interactions of the HDM2 oncogene involved in the regulation of lung cancer phenotypes. Understanding of the intricate relationships between and within the HDM2/p53/RB axis and LINE-1 retrotransposons may pave the way for thedevelopment of therapeutic strategies tailored to specific NSCLC molecular phenotypes.

Session Title: Cancer Poster Session II

PB5153 Structural and functional insights from single cell transcriptional profiles of pituitary tumors

Authors:

M. Brunner, F. Santoni, M. Messerer, R. Daniel, A. Messina, J. Meylan-Merlini, M. Muriset; CHUV, Lausanne, Switzerland

The pituitary gland is a main component of the endocrine system and a master controller of hormones production and secretion. Unlike neoplastic formation in other organs, Pituitary Neuroendocrine Tumors (PitNETs) are frequent in the population (15%) and are almost exclusively benign. Here, we present the single cell transcriptome analysis on 12 independent tumors for a total of ~70'000 single cells: 4 bi-hormonal (GH-PRL) tumor, 3 corticotropic macro-adenoma, 4 non-secreting adenomas and one prolactinoma. Characterization of all tumors showed heterogeneous cell populations. We discover that G*ATA2*, a transcription factor for early gonadotrophs differentiation is exclusively expressed in non-secreting adenoma. GATA2 inhibits GNRH receptor, leading to a decrease in *LHB* and *FSHB* production possibly explaining the non-secreting phenotype.
Furthermore, *GATA2* expression profile in single nuclei RNAseq from healthy pituitary glands shows restricted transcription in prepubescent individuals suggesting a come-back to a pre-differentiated state for gonadotroph cells in non-secreting adenomas. In three tumors we identified an unexpected small population of proliferative cells (*MKI67+*, *TOPA2+*, *BIRC5+*, *PBK+*). Intriguingly, *IQGAP3*, a gene already known to be a bad prognostic marker in different type of carcinomas, was expressed in this cluster, suggesting a correspondence between proliferative markers in PitNETs and malignant adenomas. Moreover, taking advantage of WGS data from both patient and tumor, we were able to detect somatic variants and, using VarTrix, to isolate clear subsets of tumor and normal cells with a limited efficiency due to a lack of coverage in scRNA-seq data. To overcome this issue, a machine learning approach (scANVI) was used to predict the nature of the remaining cells. Pseudobulk differential gene expression analysis revealed a decreased ribosomal activity as well as an increase activity of mitochondria in tumorigenic cells. Additionally, in non-secreting adenomas, tumor cells were the ones expressing *GATA2* and, to a minor extent, *CD274* (PD-L1). Taken together, our results give a new perspective on the comprehension of the structural composition and the dynamic progression of pituitary tumors and of adenomas in general.

Session Title: Cancer Poster Session III

PB5154 Structural variation in *CYP2A6* is associated with risk of lung, but not ovarian, cancer in the UK Biobank

Authors:

A. Langlois[1,2], A. El-Boraie[1,2], J. Pouget[2,3], J. Knight[4], M. Chenoweth[1,2,3], R. Tyndale[1,2,3]; [1]Dept. of Pharmacology and Toxicology, Univ. of Toronto, Toronto, ON, Canada, [2]Campbell Family Mental Hlth.Res. Inst., Ctr. for Addiction and Mental Hlth., Toronto, ON, Canada, [3]Dept. of Psychiatry, Univ. of Toronto, Toronto, ON, Canada, [4]Data Sci. Inst. and Med. Sch., Lancaster Univ., Bailrigg, United Kingdom

CYP2A6 is the primary nicotine-metabolizing enzyme. *CYP2A6* is polymorphic, with structural variants (SV) including gene deletions (e.g. *CYP2A6*4*), duplications (e.g. *CYP2A6*1x2*), and hybrids with the pseudogene *CYP2A7* (e.g. *CYP2A6*12, *34, *53*). Two recent genome-wide studies (one in *BRCA1* pathogenic variant carriers) explored the associations of copy number variants (CNV) with ovarian cancer (OC) risk; both found significant associations between putative *CYP2A7* deletions and OC risk (albeit in opposite directions). We hypothesized that these putative *CYP2A7* deletions may be known *CYP2A6* SVs, misidentified by CNV prediction software (PennCNV), and demonstrated this using PennCNV in an internal dataset with known *CYP2A6* SV diplotypes. Using a validated SV imputation reference panel, we sought to investigate associations of *CYP2A6* SVs with OC and lung cancer (LC) in the UK Biobank. Furthermore, we investigated the association of *CYP2A6* SVs with squamous cell carcinoma (SCC), an LC subtype strongly associated with cigarette smoking.
To impute *CYP2A6* SV diplotypes for UK Biobank participants, we used an internal reference panel composed of array-genotyped and imputed SNP haplotypes from European-ancestry (EUR) individuals with known *CYP2A6* SV diplotypes (n=935). Our case-control design used the presence of at least one deleterious *CYP2A6* SV (*CYP2A6*4*, **12*, **34*, or **53*) as the exposure, testing for an association with case status using multiple logistic regression. Analyses were limited to genetically identified EUR UK Biobank participants, controlling for age and the first ten principal components. OC analyses were limited to females and additionally adjusted for smoking status, while LC analyses were stratified by smoking status and additionally adjusted for sex.
Deleterious *CYP2A6* SVs were not associated with OC risk (adjusted OR=1.1; 95% CI: 0.80-1.37; p=0.7). In current smokers, deleterious *CYP2A6* SVs were associated with lower overall LC risk (adjusted OR=0.4; 95% CI: 0.29-0.64; p<0.0001) and lower SCC risk (adjusted OR=0.2; 95% CI: 0.08-0.58; p<0.01). No significant associations were observed for LC or SCC in former or never smokers.
In conclusion, we found that deleterious *CYP2A6* SVs were not associated with OC risk. However, deleterious *CYP2A6* SVs conferred protection against overall LC and SCC risk within current smokers. Further analyses could examine OC risk in BRCA1 pathogenic variant carriers, and investigate whether decreased smoking exposure explains the protective effect of deleterious *CYP2A6* SV on LC and SCC risk.

Session Title: Cancer Poster Session I

PB5155 *TGFBR1**6A and risk for colorectal cancer

Authors:

A. M. Johansen[1], K. Agyemang[1], M. Pennison[1], H. Jimenez[1], Colon CFR, A. Cristofano[2], J. Ziegler[1], J. A. Baron[3], C. Langefeld[1], B. C. Pasche[1]; [1]Wake Forest Univ. Sch. of Med., Winston Salem, NC, [2]Albert Einstein Coll. of Med., New York, NY, [3]UNC Chapel Hill, Chapel Hill, NC

Background:
*TGFBR1**6A (rs11466445), a hypomorphic variant of the type 1 TGF-beta receptor, has been associated with breast and ovarian cancer risk. However, the evidence of its role in the risk for colorectal cancer is inconclusive. We have observed that *TGFBR1**6A allele frequency varies by genetic ancestry. Here we are the first to elucidate the *TGFBR1**6A risk association with colorectal cancer after adjusting for population substructure.
Methods:
We developed novel knockin mouse models *of TGFBR1* (9A) and *TGFBR1**6A (6A) crossed with Apc^{min} mice. We assessed intestinal polyp development at 12 weeks in 9A/9A, 9A/6A and 6A/6A mice. We genotyped *TGFBR1* exon 1 in 2,147 patients with a biopsy-proven diagnosis of colorectal cancer and 1,184 healthy controls recruited by the NCI-sponsored Colon Cancer Family Registry (CCFR). Additionally, we imputed the TGF-β region of chromosome 9, including the *TGFBR1**6A genotype, in 3,419 individuals using Impute2 and the 1,000 Genomes phase 3 reference panel. Admixture estimates (Admixture) were computed using the GWAS data with HapMap CEU, YRI, CHB, MEX samples as anchoring populations.
Results:
There were significantly fewer polyps among 9A/6A (21.8±2.21) and 6A/6A (28.2±5.4) mice than among 9A/9A (43.4±4.1) mice (p=0.01). In the CCFR cohort, individuals with the 6A allele were associated with a decreased colorectal cancer risk, OR=0.84 (additive genetic model; 95%CI: 0.78-0.90, p=0.02); 6A allele frequency of 0.109 for cases and 0.130 for controls. When adjusting for siblings only and under an additive genetic model, 6A was associated with an even lower risk for colorectal cancer, OR=0.56 (95%CI: 0.43-0.72 p=0.02).
Conclusions:
*TGFBR1**6A is associated with a decreased risk for colorectal cancer with consistent findings in mice and humans. Together, these results provide important supportive evidence that *TGFBR1**6A is a high frequency, low penetrance colorectal cancer susceptibility allele.

Session Title: Cancer Poster Session II

PB5156 The added value of multi-tissue analysis in the identification of mosaicism in tumor suppressor gene syndromes.

Authors:

M. Broekema[1], L. Damen[1], E. Dubbink[2], W. Dinjens[2], L. Van Hest[1]; [1]Amsterdam UMC, Amsterdam, Netherlands, [2]Erasmus MC, Rotterdam, Netherlands

Introduction: Mosaicism refers to the presence of multiple cell clones with distinct genotypes. Mosaicism arises when certain cell lineages acquire pathogenic variants during postzygotic development. The phenotype of mosaic individuals depends on the extent of mosaicism, ranging from localized to generalized. Here, we report three diagnostically challenging cases with previously unrecognized mosaicism in the tumor suppressor gene syndromes Von Hippel Lindau (VHL) syndrome, neurofibromatosis 1 (NF1) and neurofibromatosis 2 (NF2). **Methods and results:** In a 49-year-old male with multiple paraganglioma and pheochromocytoma, no pathogenic variant was detected in genes associated with an increased risk for paraganglioma and pheochromocytoma in DNA from blood. A targeted hereditary cancer gene panel on DNA retrieved from pheochromocytoma and normal tissue revealed a pathogenic variant in the *VHL* gene c.499C>T; p.R167W at varying allele fractions (VAF; ranging from 6.2-40%). Re-examination of the blood NGS data identified this variant in 7.0% of the allele fraction, confirming VHL mosaicism. The second patient is a 35-year-old female with multiple neurofibromas, multiple café au lait macules, Lisch noduli and freckling, in whom no pathogenic variant in the *NF1* gene was identified in DNA extracted from blood. *NF1* analysis by NGS on multiple neurofibroma identified the c.3826C>T; p.R1276* pathogenic variant at varying VAFs (range: 22-34%). NF1 mosaicism was confirmed by detection of this variant in 5.8% and 5.4% of the reads in DNA extracted from buccal mucosa and blood, respectively. The third patient is a 37-year old female with a clinical diagnosis of NF2. She was diagnosed with bilateral vestibular schwannoma, multiple meningioma, and a schwannoma. No pathogenic variant was identified in the *NF2* gene in DNA from blood. *NF2* analysis by NGS on two meningioma and the schwannoma revealed the c.592C>T; p.R198* pathogenic variant in VAFs ranging from 45%-70%. In retrospect, low-grade NF2 mosaicism was confirmed in blood (1% of the reads). **Conclusion:** Recent advances in NGS along with the opportunity for multi-tissue sampling is an effective strategy in the detection of low-level mosaicism. Identification of previously undiagnosed cases allows for more precise and accurate genetic counseling. More cases of mosaic tumor suppressor gene syndromes will likely be identified.

Session Title: Cancer Poster Session III

PB5157 The age-dependent dynamics of healthy clonal haematopoiesis in UK Biobank

Authors:

H. MacGregor, D. Easton, J. R. Blundell; Univ. of Cambridge, Cambridge, United Kingdom

Clonal haematopoiesis (CH), the expansion of somatic mutant clones under positive selection among haematopoietic stem cells (HSCs), is almost ubiquitous in older adults. CH is a known precursor to several haematological cancers and the presence of CH in blood can confound efforts to find rare germline risk-associated variants. The dynamics of CH are not yet fully understood, and an accurate quantitative model of clonal expansion in healthy people could inform early detection strategies.

We analysed the variant allele frequency (VAF) distributions of common CH driver variants in *DMNT3A, JAK2* and *SF3B1* from the exomes of 400,000 cancer-free UK Biobank participants and estimated mutation rates and clonal fitness parameters that were similar to previously published work. However, when stratified by age, the VAF distributions predicted by a simple model assuming constant exponential growth are inconsistent with the data, underpredicting the frequencies at younger ages and implying a significant deceleration of clonal expansion with age.

We considered several alternative models of clonal dynamics to that might be able to explain this conundrum. First we show that variation in fitness of variants among individuals (e.g., due to genetic susceptibility) is unable to resolve the inconsistency. Second, we considered an ageing model where HSC division slows down with increasing age in all people. Third, we considered a clonal competition model in which clonal expansion in later life is inhibited by competition with an unobserved background of other expanding clones. The observed age-specific VAF distributions are consistent with both the ageing and clonal competition models with plausible parameters. To further interrogate these two models, we considered a large data set of longitudinal growth rates obtained from serial samples of blood. These data suggest a likely role for both clonal competition and general slowdown of HSC division rates in shaping the genetic diversity of ageing blood.

Session Title: Cancer Poster Session I

PB5158 † The ClinGen-InSiGHT *MUTYH* Variant Curation Expert Panel: Lessons learned and a call to action.

Authors:

E. Nadeau[1], C. Rioja[2], M. Pineda[3], M. Thet[4], A. Laner[5], D. Buchanan[6], J. Del Valle[2], E. Borras[7], F. Hernandez[8], C. Heinen[9], I. Frayling[10], K. Mahmood[11], M. Johnson[12], P. Georgeson[4], R. Urban[13], S. David[14], S. Farrington[15], T. Hansen[16], W. Shen[13], X. Shi[17], S. Yin[4], D. Ritter[18], P. Mur[2], I. Quintana[19], A. Latchford[20], F. Macrae[21], S. Tavtigian[22], G. Capellá[2], M. Greenblatt[23]; [1]Univ. of Vermont, Hyde Park, VT, [2]Catalan Inst. of Oncology, Barcelona, Spain, [3]Catalan Inst. of Oncology, L'hospitalet De Llobregat, Spain, [4]Univ. of Melbourne, Melbourne, Australia, [5]MGZ, Munich, Germany, [6]The Univ. of Melbourne, Parkville, Australia, [7]Invitae, Houston, TX, [8]Ambry Genetics, Aliso Viejo, CA, [9]Univ Connecticut Hlth.Ctr., Farmington, CT, [10]Inst. of Med. Genetics, Cardiff, United Kingdom, [1112]Ambry Genetics, Los Angeles, CA, [13]Mayo Clinic, Rochester, MN, [14]Univ. of California Davis, Davis, CA, [15]Univ. of Edinburgh, Edinburgh, United Kingdom, [16]RigsHosp.et, Copenhagen, Denmark, [17]GGC, Greenwood, SC, [18]Baylor Coll. of Med., Houston, TX, [19]IDIBELL, Barcelona, Spain, [20]Polyposis Registry, St Mark's Hosp., London, United Kingdom, [21]Human Variome Project, Melbourne, Australia, [22]Huntsman Cancer Inst, Salt Lake City, UT, [23]Univ Vermont, Burlington, VT

The ClinGen-InSiGHT Hereditary Colon Cancer/Polyposis Variant Curation Expert Panel (VCEP) was formed in 2021 to create variant curation specifications for the known genes causing these conditions.
The *MUTYH* subcommittee of this VCEP is nearing the release of gene-specific recommendations for ACMG variant classification. *MUTYH*-associated polyposis (MAP) is an autosomal recessive disorder caused by germline biallelic variants in the base excision repair gene *MUTYH*. *MUTYH* encodes a glycosylase that identifies and excises adenines mispaired with the oxidation product 8-oxo-deoxyguanosine (OG), which, if left incorporated, lead to somatic G>T transversions. Through developing ACMG criteria for *MUTYH*, experts from functional, clinical, and computational fields have identified key gaps in understanding transcript expression, functional studies, and clinical features, plus inadequate communication among researchers in these fields. We first identified nomenclature discrepancies because of confusion in application of the Matched Annotation from NCBI and EMBL-EBI (MANE) *MUTYH* transcript. The *MUTYH* MANE Select transcript (NM_001048174) encodes a 521 AA protein. However, transcript NM_001128425, which encodes a 549 AA protein, is the most used by clinical labs. Our VCEP successfully had this second transcript officially recognized as MANE Plus Clinical, harmonizing historical data with current clinical reports. Another issue the VCEP addressed was establishing specific phenotypic descriptions of affected individuals. These are needed since MAP patients present with a variable number of polyps, age of onset, extracolonic features, and association with colorectal carcinoma. Somatic genomic analysis for G>T transversions, mostly done in colorectal cancer, provides strong support for MAP diagnosis, but these tests are rarely done, so data for MAP-specific cohorts are lacking. For ClinGen, *MUTYH* is the first autosomal recessive cancer gene assessed which requires additional ACMG criteria. Per ACMG guidelines, only evidence from “well-established” functional assays can be used for applying criteria PS3 and BS3. Satisfactory functional assays are lacking due to inadequate use of controls and inattention to statistical principles that establish the strength of evidence, despite years of research on *MUTYH*. Improved communication among academics, clinicians, and industry labs is needed. Here, we identify the knowledge gaps in developing *MUTYH*-specific ACMG recommendations in the hopes of galvanizing these communities to generate the needed data for this routinely assessed cancer susceptibility gene.

Session Title: Cancer Poster Session II

PB5159 The development of Korea Cancer Omics Research (K CORE) web portal

Authors:

J. Hwang, H. Koo, S. Kong, K. Choi, H-J. Kim; Natl. cancer center, goyang-si, Korea, Republic of

Background Multi omics analysis in cancer research is an important approach to integrate high dimensional datasets to better understand the molecular biology of cancer. In addition, integrating individual clinical information with omics profile may provide significant opportunities for precision oncology. Therefore, this study aims to design and develop a user-friendly web portal that enables comprehensive analysis of clinical features and omics data of cancer patients. **Methods** The web portal development was carried out as follows: 1) Derivation of requirements for integrated analysis of multi-dimensional data via available related portals; 2) Acquisition of sample data for clinical and omics information from the Korea National Cancer Center; 3) Design of system through data structure and interface discussion; 4) Development and validation of an integrated analytics portal using R or Python; and 5) Open beta testing with multiple stakeholders **Results** An open access web portal called Korea Cancer Omics REsearch (K CORE) was developed to provide visualization and download of analysis results through multidimensional linkage of omics data and clinical information. This web portal allows users to upload their data to the portal for integrated analysis of various clinical and omics profiles. Users can implement a variety of integrated visualization analysis within this portal (e. g. Circos Plot, Oncoprint, Lollipop Plot, Heatmap, Kaplan-Meier, Cox-regression, Correlation Plot, CNV Plot, and Boxplot). Patient-specific drug information can be also checked through the report function. In addition to integrated analysis, we provide a single analysis for each dimension of data and plan to continue to expand. **Conclusion** K CORE (https://cancerdata.re.kr/k-core) may be used as a meaningful web analysis tool in precision oncology by enabling integrated analysis of high dimensional information including clinical and omics data in cancer patients.

Session Title: Cancer Poster Session III

PB5160 The genetic breast cancer risk factors in Ghanaian women.

Authors:

C. Anyigba[1,2], L. N. Amenga-Etego[1,2], l. Paemka[1,2]; [1]Univ. of Ghana, Accra, Ghana, [2]West African Ctr. for Cell Biology of Infectious Pathogens, Accra, Ghana

Breast cancer genetics is largely unexplored in the Ghanaian population despite the increasing associated mortality and prevalence of early-onset breast cancer (EOBC) and aggressive presentations. The only available genetic study investigated the germline drivers using a Eurocentric 34-gene known and putative breast cancer panel. This study essentially highlighted the contribution of some of the known gene variants genes to EOBC and identified some novel variants in Ghanaian women, however, it was limited in discovering potentially novel genes and variants. Therefore, this study sought to profile the genetics of breast cancer in Ghanaian women using whole exome sequencing. With informed consent, exomes from 86 neoadjuvant-naïve breast cancer patients were sequenced and analyzed for germline and somatic variants. Novel pathogenic germline variants were detected in *BRCA2* and three potentially novel genes. Interestingly, the *IRF5* variant (rs2004640), which was previously associated with systemic lupus erythematosus (SLE), was detected in 27% of patients and no controls. While this association remains debated, it is worth noting that the loss of *IRF5* expression reportedly contributes to breast cancer metastasis, suggesting that this variant may drive breast tumorigenesis. A well-powered study will help determine the relative risks and penetrance of the identified pathogenic variants. Notably, the known somatic drivers were not detected in this study but the somatic events observed were mediated by defective homologous recombination, DNA mismatch repair, and *APOBEC* cytidine deaminase activity. Somatic interaction analysis detected 3 mutually exclusive and 19 co-occurrence interactions. Of note is a co-occurrence interaction between *SCUBE2* and *UGT2A1* in 10 patients. Previous studies suggest that *SCUBE2* may be a novel breast tumor suppressor and an independent disease-free breast cancer survival prognosticator. However, *SCUBE2* overexpression has been associated with triple-negative breast cancer (TNBC; the most aggressive form) aggression and metastasis; a high TNBC prevalence is observed in Ghanaian women. *UGT2A1* variants have been implicated in drug resistance. Thus, *SCUBE2* and *UG2TA1* cancer driver mutations may confer poor breast cancer prognosis.

Session Title: Cancer Poster Session I

PB5161 The GOBACK study: High molecular diagnostic rate in individuals with pediatric cancer and birth defects.

Authors:

H. Gu[1], Y. Yu[2], S. Dushyant Sisoudiya[1], P. Mishra[1], H. Li[1], J. M. Schraw[1], M. E. Scheurer[1,3], D. Muzny[1], D. Mitchell[1], O. Taylor[1], R. A. Gibbs[1], C. Huff[4], S. E. Plon[1,3], P. J. Lupo[1,3], A. Sabo[1]; [1]Baylor Coll. of Med., Houston, TX, [2]UT MD Anderson Cancer Ctr., Houston, TX, [3]Texas Children's Hosp., Houston, TX, [4]U.T. MD Anderson Cancer Ctr, Pearland, TX

Background: While having a birth defect is one of the strongest risk factors for developing pediatric cancer, the genetic underpinnings of these associations remain unclear. We sought to characterize germline genetic susceptibility in children with both birth defect(s) and cancer diagnosis.
Methods: Children with birth defects and cancer were recruited through the Genetic Overlap Between Anomalies and Cancer in Kids (GOBACK) Study. For this analysis, the cohort included 39 probands, as well as biological parents and siblings. We conducted whole-genome sequencing (WGS) and evaluated single nucleotide variants, indels, and structural variants. Variants were excluded if the minor allele frequency was >1% or if classified as intergenic, intronic, or synonymous without dbscSNV scores and/or any reports in clinical variant databases. Finally, pathogenic, and likely pathogenic (P/LP) variants were identified using American College of Medical Genetics (ACMG)/Association of Molecular Pathology (AMP) criteria and evaluated in the context of the proband's phenotype.
Results: We identified P/LP variants in 25% (10/39) of the probands undergoing sequencing. Among the 10 probands: 1) four had P/LP variants in genes known to underlie both the birth defect and cancer diagnosed in the individual (*WT1, USP9X, PTPN1, LZTR1*); 2) four had P/LP variants in genes associated with the birth defect (*MMUT, FBN1, COL3A1, and KAT6B*); and 3) two had P/LP variants in *TP53*, an established cancer predisposition gene. The loss of function (LOF) variant in the *KAT6B* gene was found in one proband with genitopatellar syndrome (GPS) and neuroblastoma. Notably, a previous study reported a different LOF variant in *KAT6B* in an individual with GPS and neuroblastoma (Knight et al., 2018). To further characterize the role of germline *KAT6B* variants in neuroblastoma risk, we conducted gene-based test using WGS data from 409 neuroblastoma cases and 952 control samples from the Gabriella Miller Kids First Pediatric Research Program and found a significant enrichment of rare, predicted deleterious variants in *KAT6B* in neuroblastoma cases (p-value = 0.017).
Conclusion: Our study demonstrates a higher molecular diagnostic yield (25%) in individuals with birth defects and pediatric cancer than cancer cohorts alone. The results elucidate the contribution of genetic variants in explaining either cancer, birth defects, or both, which can provide valuable insights for both research and clinical applications. Additionally, our study adds to a body of evidence suggesting *KAT6B* is a novel neuroblastoma predisposition gene.

Session Title: Cancer Poster Session II

PB5162 The Interaction Between TCE Permissiveness and HLA Class I Non-coding Mismatches May Impact Outcomes of Hematopoietic Cell Transplantation

Authors:

M. Askar[1,2], T. Mosbruger[3], G. Shaw[3], Y. Duan[4], A. Allen[4], J. Duke[3], T. Olson[3,5], D. Monos[3,5], T. Hayeck[6,5]; [1]1QU Hlth.and Dept. of Immunology, Coll. of Med., Qatar Univ., Doha, Qatar, [2]Be The Match / 2Natl. Marrow Donor Program (NMDP), Minneapolis, MN, [3]Div. of Genomic Diagnostics, Dept. of Pathology and Lab. Med., Children's Hosp. of Philadelphia, Philadelphia, PA, [4]Dept. of Biostatistics and Bioinformatics, Duke Univ., Durham, NC, [5]Dept. of Pathology and Lab. Med., Perelman Sch. of Med., Univ. of Pennsylvania, Philadelphia, PA, [6]Div. of Genomic Diagnostics, Dept. of Pathology and Lab. Med., Children's Hosp. of Philadelphia, Philadephia, PA

Hematopoietic cell transplantation (HCT) is the only curative treatment for many hematological malignancies, such as leukemia, immunodeficiencies and other disorders. However, HCT from carries a significant risk of acute graft-versus-host disease (aGVHD) and treatment-related mortality (TRM). Matching at HLA loci in transplantation settings is crucial to improve outcomes. The primary focus is matching coding variants, specifically the antigen recognition domain (ARD), of both Class I and Class II HLA molecules. Furthermore, strategies have been developed to detect compatible T-cell epitopes (TCEs) and distinguish between permissive and non-permissive mismatches in HLADPB1. The role of non-coding variation in HCT outcomes remains unclear.We performed a retrospective study with ultra-high resolution (UHR)HLA typing in 5,109 HCT patient-donor pairs transplanted between 2008 and 2017 and reported to CIBMR. In this study we examine the impact of mismatches at non-coding sequences of HLA class I alleles compared to UHR matches. UHR matches referred to patient-donor pairs that exhibited a match on all 12 class I and class II HLA genes at the highest available resolution. To account for potential confounding factors, Cox proportional hazard models were weighted by inverse probability of matching. We investigated whether such impact is influenced by the TCE permissiveness of HLA-DPB1 mismatches. We found that TCE non-permissive mismatches combined with ≥1 non-coding mismatches (excluding individuals with coding mismatches) was associated with increased risk of acute graft vs. host disease (aGVHD) in T cell replete (HR: 1.6, p = 0.0003) but not T cell deplete and increased risk of transplant related mortality (TRM) in the T-cell-deplete (HR: 1.8, p = 0.006) but not in T cell replete transplants. There were no significant association between TCE non-permissive mismatches combined with ≥1 non-coding mismatches and risk of chronic GVHD (cGVHD) or between any of the studied outcomes and non-coding sequencing mismatches in TCE permissive transplants.These results suggest there may be an interaction between the impact of non-coding mismatches, TCE permissiveness, and the graft manipulation and/or composition on transplant outcome. This is consistent with the literature reporting association with non-coding mismatches and level of expression of different HLA loci that correlates with HTC outcomes. These results need to be validated in larger studies, ideally with a mechanistic component.

Session Title: Cancer Poster Session III

PB5163 The mammalian female reproductive tract at single-cell spatial resolution

Authors:

A. Goncalves, I. Winkler, A. Tolkachov, P. Lacour, F. Lammers, K. Daugelaite, N. Schneider, M-L. Koch, A. Schneider, D. Odom; DKFZ, Heidelberg, Germany

The study of female-specific biology has historically been overlooked and undervalued. In this presentation, I will showcase a large-scale project that involves generating and analyzing single-cell and spatial transcriptomics data from 50 different cell types found in 5 reproductive organs. This data was collected throughout 5 phases of the estrous cycle and during pregnancy in young mice, as well as during 6 different time points as the female reproductive tract (FRT) ages. Our findings provide a comprehensive atlas of the FRT throughout the lifespan and shed light on the consequences of unresolved recurrent inflammation and tissue repair.
During each reproductive cycle, the FRT undergoes significant remodeling, which is regulated by systemic changes in sex hormones. However, it remains unknown whether this recurrent remodeling influences the aging process of specific organs within the FRT. To address this question, we conducted a systematic analysis at the single-cell level to examine the morphological and transcriptional changes that occur in the ovary, oviduct, uterus, cervix, and vagina at each phase of the mouse estrous cycle, during decidualization, and throughout the aging process. We investigated whether the cyclic inflammation and remodeling that naturally occur during the reproductive lifespan of young mice lead to age-related chronic inflammation and fibrosis. In-depth characterization of fibroblasts and their communication networks revealed that transcription factors and communication pathways involved in fibroblast activity during the estrous cycle and aging are enriched in extracellular matrix (ECM) remodeling and inflammation. Remarkably, these findings were consistent between human and mouse uteruses. Our research establishes a direct link between the intensity of inflammation and ECM activity during the estrous cycle and the severity of chronic inflammation and fibrosis in old age. Our data supports a model where incomplete resolution of inflammation and ECM remodeling over multiple cycles gradually leads to the development of fibrosis and chronic inflammation, making organs more susceptible to disease. Furthermore, using a mouse model of premature menopause, we isolated the effects of aging and cycling on fibrosis development, confirming the contribution of cycling. Our single-cell atlasing efforts implicate fibroblasts in maintaining a "memory" of past inflammation and propose a unifying mechanism to explain the epidemiological association between various factors that modify the number of cycles and the risk of endometrial cancer in humans.

Session Title: Cancer Poster Session I

PB5164 The molecular landscape of myeloid malignancies with ring chromosomes

Authors:

Y. Zou, J. B. Murry; Johns Hopkins Univ., Baltimore, MD

Background: Ring chromosome (RC) is a cytogenetic abnormality reported infrequently in myeloid malignancies, and we sought to evaluate the incidence internally. As the molecular profiles of cases with concomitant myeloid-associated RC are not well-known, we sought to evaluate the gene mutations in these RC. **Methods:** We reviewed 13,124 hematological malignant specimens that underwent routine chromosome analysis performed internally from 2014 to 2022. 83 (~0.6%) hematological malignant specimens had RC. Identified myeloid malignancies (n=58) in decreasing frequency included acute myeloid leukemia (n=36)(62.1%), myelodysplastic syndromes (n=19)(32.8%), chronic myeloid leukemia (n=2)(3.4%), and chronic myelomonocytic leukemia (n=1)(1.7%). 52 identified myeloid RC also had Next-Generation-Sequencing-based results. **Results: Of the** 58 myeloid RC, 55 (94.8%) had complex karyotypes (ck) with >three independent cytogenetic abnormalities, and copy number variants (CNVs) in these cases included monosomy 5/5q deletion (5q-)(67.9%), monosomy 7/7q deletion (7q-)(60.4%), a gain of marker chromosome(s) with unknown origin (+mar)(56.6%), monosomy 17/17p deletion/abnormality (17p-)(37.7%), monosomy 13/13q deletion (13q-)(20.7%), monosomy 20/20q deletion (20q-)(15.1%), and trisomy 8 (+8)(9.4%). Ten (18.9%) had multiple RCs (rr), and 9 out of 10 had both rr and +mar of unknown origin. Myeloid RC with ck had pathogenic mutations
in *TP53* (77.6%), *NRAS* (10.2%), *TET2* (10.2%), *SETBP1* (8.2%), *ASXL1* (6.1%), *CEBPA* (6.1%), *FLT3* (6.1%), *DNMT3A* (6.1%), *ABL1* (4.1%), *NF1* (4.1%), *JAK2* (4.1%), *NPM1* (4.1%), *KRAS* (4.1%), etc. Three (5.2%) lacking ck showing the following karyotypes: 1) 46,XY,-7,+r, 2) 46,XY,-18,+r, and 3) 48,XX,+4,+r. All three myeloid RC without ck had a pathogenic mutation in *TET2* (100%), two had two pathogenic mutations in *TET2* with a *DNMT3A* mutation (67%), and none had a *TP53* mutation. **Conclusions:** A subset (~0.6%) of myeloid RC was identified in our institution. Most myeloid RC were associated with ck and had recurrent myeloid-associated CNVs, with 5q- and 7q- being the most common, followed by 17p- abnormalities. Over two-third of RC with ck had pathogenic mutations in *TP53*. All RC without ck had pathogenic mutations in *TET2* and no *TP53* mutation. Most RC with rr also had +mar, suggesting genomic instability secondary to RC formation. Overall, this study identified specific molecular profiles that may be associated with distinct RC subtypes in myeloid malignancies. These data indicate that future study of the molecular profiles and their impact on the outcomes of the different myeloid RC subtypes are needed to inform clinical decision-making.

Session Title: Cancer Poster Session II

PB5165 The mutational landscape defines the proteome and spatial organization of tumor, stroma and immune cells in ovarian cancer

Authors:

M. Jones[1], K. Dabke[2], N. Gull[1], S. kreimer[1], P-C. Peng[3], N. Ng[1], K. Lawrenson[3], B. Rimel[1], B. Karlan[4], S. Knott[3], S. Gayther[2], S. Parker[3]; [1]Cedars Sinai Med Ctr, Los Angeles, CA, [2]Cedars Sinai Med. Ctr., Los Angeles, CA, [3]Cedars-Sinai Med. Ctr., Los Angeles, CA, [4]David Geffen Sch. of Med. at UCLA, Los Angeles, CA

High-grade serous ovarian cancer (HGSOC) is highly aggressive and lethal, with clinical challenges to both diagnosis and treatment . The genomic instability of HGSOC, further complicated by homologous recombination deficiency (HRD), leads to heterogeneity in the HGSOC tumors and patient response to treatment. Proteogenomic studies have provided some insight into HGSOC biology, but a deeper understanding of the effects of chemotherapy and HRD on the HGSOC proteome is needed. We have profiled the complete proteome of tumors from 32 HGSOC patients before and after multiple rounds of chemotherapy. We identified a stable methylome, transcriptome and proteome throughout progression and the development of resistance, despite an increasing somatic mutational burden. Tumors with germline or somatic loss of BRCA1 or BRCA2 had increased expression of proteins related to immune pathways, more shared TCR CDR3 repertoires of tumor-infiltrating T cells, increased neoantigen counts, and enriched immune clusters within HGSOC tissue after chemotherapy. Single cell spatial profiling of 24 paired primary and recurrent tumors was performed with Akoya FUSION to capture tumor, stroma and immune cell populations (27 protein panel). We identified increased immune cell infiltration in recurrent tumors and defined specific cell population differences based on homologous recombination deficiency that correlated with survival. Through proteo-genomic and spatial analysis, this work has shown that the immune landscape of HGSOC tumors is influenced by homologous recombination status and identified candidate drivers of HGSOC biology.

Session Title: Cancer Poster Session III

PB5166 The performance characteristic of the low inputtagmentation-based Whole Genome Sequencing in high quality somatic variant calling

Authors:

C. Dalgard, X. Zhang, G. Sukumar, M. Wilkerson, L. Jiang; USUHS - TAGC, Bethesda, MD

Whole genome sequencing (WGS) is widely used for cancer diagnostic and therapeutic applications in clinical practice and clinical trials. Currently, the Truseq PCR-free library preparation methods are routinely used in genomics laboratories. However, microgram amounts of DNA input is a limitation due to limited materials often from clinically-derived specimens. Here, we evaluated a novel low input (100-300 ng) PCR-free tagmentation (TAG) based library preparation for WGS. Replicates of TAG- and ligation-based (Illumina TruSeq)sequencing libraries were prepared from 3 pairs of breast cancer-derived tumors (HCC1195, HCC1143 and HCC1187) and matched B-lymphocyte-derived normal (HCC1195BL, HCC1143BL and HCC1187BL) cell lines. Libraries were sequenced on Illumina Novaseq 6000 platforms to target 30x and 90x mean coverage for normal and tumor samples respectively. Raw sequencing data were aligned to the hg38 by Isaac Aligner before variant analysis by Strelka2. Technical sequencing metrics demonstrated high similarity between TAG- and ligation-based workflows, including passing filter reads, Q30%, aligned reads, mean coverage. Germline variants of HCC1395BL showed greater than 98% precision and sensitivity using ligation-based variants as a reference. For somatic mutation calling from TAG-based libraries, HCC1395, HCC1143 and HCC1187, 88%, 85% and 86% precision and 81%, 70%, 68% sensitivity was observed, respectively, compared to the ligation-based method. Furthermore, using the SEQC2 Consortium high-confidence mutation set, TAG- and ligation- based variant calls of HCC1395 had 90% and 93% precision and 74% and 73% sensitivity, respectively with this high confidence reference. Further investigation showed that a high fraction of false negative calls was associated with low variant allele frequency (<10%). Compared to an available reference dataset for benchmarking somatic calling pipelines from New York Genome Center, TAG- and ligation-based variant calls of HCC1143 yielded 85% and 83% precision and 63% and 63% sensitivity, respectively; TAG- and ligation-based variant calls of HCC1187 resulted in 91% and 92% precision and 81% and 79% sensitivity, respectively. In summary, the low input TAG-based WGS protocol produced highly reproducible variant calls with highly concordant somatic variant calls compared to the commonly-used PCR-free ligation-based methods and to reference callset. Further evaluation is warranted to broaden clinical application from minimal starting material.

Session Title: Cancer Poster Session I

PB5167 The relation between vitamin D receptor (rs2228570) polymorphism and prostate cancer.

Authors:

P. Khademi; Yeditepe Univ., Istanbul, Turkey

Most chronic diseases are associated with vitamin D deficiency in the body. The sun helps our body produce vitamin D, which is the main source of this vitamin. Vitamin D shortage is accompanied by diseases such as bone pain, increased blood pressure, and depression. You may know that the main supply of vitamin D is out of your home and through sunlight. Muscle growth and strengthening bones, lowering blood pressure, reducing pain caused by fibromyalgia, and slowing the progression of MS symptoms are all associated with adequate levels of vitamin D inside the body, however, a lack of this vitamin causes symptoms such as muscle weakness, rising blood strain, bone pain, and depression. It has even been said that the prevalence of MS is higher in regions where there is not enough sunlight. However, you may not know that many chronic diseases are the result of low surfaces of vitamin D, and prostate cancer has been proven throughout the world. Prostate cancer is directly related to vitamin D shortage in the body. This occurs mainly in older men and, according to the American Cancer Society, The most common disease and a leading cause of death in men is carcinoma. Moreover, the antiproliferative effect on prostate cells is mediated by the vitamin D receptor active form of vitamin D. In a case-control study, we reviewed whether the vitamin D receptor (VDR) gene (RS2228570)polymorphism could influence PC sensitivity. Keywords: Vitamin D, deficiency, Prostate cancer, connection between, (VDR) gene(RS2228570) polymorphism

Session Title: Cancer Poster Session II

PB5168 The Sherlock Lung cancer study: A Pilot Study of Lung Cancer in Never Smokers in Nigeria

Authors:

O. Omonijo[1], S. Isreal-Isah[1], M. Kojusola[1], E. Jonathan[1], A. Ndubuisi[2], P. Jibrin[3], J. Ahmed[3], I. M. Ibrahim[3], A. Omotosho[4], E. Jedy-Agba[1]; [1]Inst. of human Virology Nigeria, Abuja, Nigeria, [2]Natl. Hosp. Abuja, Abuja, Nigeria, [3]Aminu Kano Teaching Hosp., Kano, Nigeria, Kano, Nigeria, [4]Univ. of Calabar Teaching Hosp., Nigeria, Calabar, Nigeria

Background: Lung cancer is a leading cause of mortality from cancer globally. It occurs in approximately 10 - 20% of never smokers worldwide (National Institutes of Health, 2021). The various risk factors for lung cancer in never smokers were explored in the Sherlock Lung Cancer Study. **Methodology:** This was a pilot cross-sectional study conducted at 3 participating hospitals in Nigeria: The National Hospital Abuja (NHA), Aminu Kano Teaching Hospital Kano (AKTH) and the University of Calabar Teaching Hospital (UCTH). All patients who presented with symptoms highly suggestive of lung cancer at each of these hospitals were invited to participate in the study. All patients included in the study were never smokers who had no previous cancer diagnosis. The sample size was 15 based on a feasibility assessment of annual cases of lung cancer reported in these hospitals over a period of 3 years. Questionnaires were administered, blood samples and lung tissue biopsy samples were collected from consented participants. Data was entered into REDCap. **Results:** A total of 21 suspected cases of lung cancer were enrolled into the study,13 participants were confirmed lung cancer cases in never smokers. 7 (53.85%) of the cases were male, 6 (46.15%) were female respectively. Most cases have the same risk factor of being exposed to air pollution from cooking with smoky coal and wood, passive smokers make up for 11 (84.62%), 11 (84.62%) wood burning and 6 (46.15%) coal burning being major contributors. Moreover, there are varying degrees of comorbidities in participants such as: 8 cases of peptic ulcer disease (61.54%), 5 cases of HBP (38.46%), 2 cases of asthma (15.39%), 2 cases of COPD (15.39%), 1 case of tuberculosis (7.69 %), and 1 case of pneumonia (7.69%), respectively. 7 (53.84%) were shown to have at least two types of comorbidities, while 3 (23.07%) showed no presence of any form of comorbidity. Mortality rate is 53.84% (1 in every 2 cases). **Conclusion:** The relationship between environmental factors, comorbidities & lung cancer cannot be fully established suggesting a need for more elaborate study. However, there is need for more awareness on lung cancer in LMICs to avoid late presentations to hospitals and lower mortality rate.

Session Title: Cancer Poster Session III

PB5169 The somatic landscape of low-grade epilepsy-associated tumors across 275 surgically accessible human epileptogenic brain lesions

Authors:

C. Leu[1], C. Boßelmann, M[2], L. Hoffmann[3], S. Baldassari[4], R. Coras[5], K. Kobow[5], H. Hamer[5], S. Brandner[5], K. Rössler[6], C. Bien, G[7], T. Kalbhenn[7], T. Pieper[8], T. Hartlieb[8], K. Becker[9], J. Altmüller[10], L. Ferguson[2], R. Busch[2], S. Baulac[4], P. Nürnberg[9], I. Najm[2], I. Blümcke[5], D. Lal[1]; [1]McGoven Med. Sch., UTHlth.Houston, Houston, TX, [2]Cleveland Clinic, Cleveland, OH, [3]Univ. clinic Erlangen, Erlangen, Germany, [4]Pitié-Salpêtrière Univ. Hosp., Paris, France, [5]Univ. Hosp. Erlangen, Erlangen, Germany, [6]Med. Univ. of Vienna, Vienna, Austria, [7]Bielefeld Univ., Bielefeld, Germany, [8]Schoen Clinic Vogtareuth, Vogtareuth, Germany, [9]Univ. of Cologne, Cologne, Germany, [10]Charité Med. Univ. Berlin, Berlin, Germany

Our understanding of the molecular mechanisms involved in the etiology of epileptogenic brain tumors is essential for improving the treatment of drug-resistant focal epilepsy. Recent large-scale studies made significant contributions, yet our knowledge is incomplete. We aggregated 275 brain tissues with pathology-defined low-grade epilepsy-associated tumors (LEAT) and 89 brain control tissues across multiple centers in Europe and the Cleveland Clinic in the US. LEAT samples were screened with >350x whole-exome sequencing (N=154) and >1500x targeted sequencing of 122 genes selected from a previous study (N=121). Brain control tissues were screened with the 122 gene panel. Somatic variant enrichment was tested using the dNdScv model that tests for the ratio of non-synonymous (missense, nonsense, and splice site) to synonymous somatic mutations while accounting for variations in the background local mutation rate along the human genome. We used stringent Bonferroni correction after the number of genes captured by the whole-exome screens to identify genome-wide significant somatic variant-enriched genes. We also tested for association with LEAT against the brain control tissues using only panel-sequenced tissues and the burden/variant collapsing method (121 LEAT vs. 89 controls). Somatic variant enrichment analysis confirmed three genes known to cause LEAT (*NF1*, *BRAF*, and *FGFR1*) and identified one novel gene. *NF1* was enriched in 28% of dysembryoplastic neuroepithelial tumors (DNET, P=2.39x10^{-6}) and 22% of gangliogliomas (GG, P=7.30x10^{-7}). The enrichment of *BRAF* was driven by the GG pathology (31% carriers, P=3.34x10^{-10}), while *FGFR1* was mainly enriched in DNET (26% carriers, P=7.72x10^{-6}). All known genes identified in the enrichment analysis were also significantly associated with LEAT in the somatic burden analysis against controls (P_{BURDEN}<1.4x10^{-4}). Also, the burden analysis identified two additional novel genes. All three identified novel candidate genes were previously reported as potential tumor suppressor genes in various cancers. A subset of individuals carried somatic variants in >1 of identified genes. Interestingly, all individuals with multiple hits and a *BRAF* variant had the second hit in *NF1*. Our study suggests that LEAT pathologies not explained by the main known driver genes are highly heterogenic. A burden test against a small control cohort was powered enough to identify several associated genes, improving the perspective for case/control somatic association studies in brain tissues.

Session Title: Cancer Poster Session I

PB5170 The ungracious guest: how the human papillomavirus (HPV) changes the local host epigenome and transcriptome to promote tumorigenesis.

Authors:

M. M. Hoffman[1,2,3], M. Karimzadeh[2], E. Collier[2], C. Arlidge[1], A. Rostami[2], J. Zou[1], Z. Zhao[1], Y. Zheng[1], E. Leung[4,2], K. Han[1,2], S. Lheureux[1,2], M. Lupien[1,2], S. V. Bratman[1,2]; [1]Univ. Hlth.Network, Toronto, ON, Canada, [2]Univ. of Toronto, Toronto, ON, Canada, [3]Vector Inst. for Artificial Intelligence, Toronto, ON, Canada, [4]Sunnybrook Hlth.Sci. Ctr., Toronto, ON, Canada

Background: Human papillomavirus (HPV) drives almost all cervical cancers and up to 70% of head and neck cancers. Frequent integration into the host genome occurs predominantly in tumorigenic types of HPV. We hypothesize that changes in chromatin state at the location of integration can result in changes in gene expression that contribute to the tumorigenicity of HPV.
Results: We find that viral integration events often occur along with changes in chromatin state and expression of genes near the integration site. We investigate whether introduction of new transcription factor binding sites due to HPV integration could invoke these changes. Some regions within the HPV genome, particularly the position of a conserved CTCF binding site, show enriched chromatin accessibility signal. ChIP-seq reveals that the conserved CTCF binding site within the HPV genome binds CTCF in 4 HPV^+ cancer cell lines. Significant changes in CTCF binding pattern and increases in chromatin accessibility occur exclusively within 100 kbp of HPV integration sites. The chromatin changes co-occur with out-sized changes in transcription and alternative splicing of local genes. Analysis of The Cancer Genome Atlas (TCGA) HPV^+ tumors indicates that HPV integration upregulates genes which have significantly higher essentiality scores compared to randomly selected upregulated genes from the same tumors.
Conclusions: Our results suggest that introduction of a new CTCF binding site due to HPV integration reorganizes chromatin state and upregulates genes essential for tumor viability in some HPV^+ tumors. These findings emphasize a newly recognized role of HPV integration in oncogenesis.

Session Title: Cancer Poster Session II

PB5171 Three-Dimensional Nuclear Organization of Telomeres in Breast Cancer

Authors:

O. Samassekou[1], S. Mai[2]; [1]Univ. of Sci., Technics and Technologies, Bamako, Mali, [2]Univ. of Manitoba, Winnipeg, MB, Canada

Introduction:Breast cancer is the most common cancer among women, exhibiting clinical heterogeneity and diverse molecular subtypes. Within each molecular subtype, further heterogeneity exists. Breast cancer cells often demonstrate shorter telomeres compared to normal cells, and telomere alterations may serve as indicators of genomic instability. This study investigates whether the three-dimensional nuclear organization of telomeres is altered in different molecular subtypes of breast cancer. Methods:We examined the telomeric nuclear organization in 50 breast cancer patients representing various molecular subtypes. Telomere fluorescent in situ hybridization (FISH) was performed on tissue biopsies using Peptide Nucleic Acid (PNA). Three-dimensional microscopy and a quantitative method called TeloView^{TM} were utilized to assess different parameters defining telomeric nuclear organization. Results:Our findings revealed notable differences among the molecular subtypes. HER2-positive patients exhibited a higher presence of telomeric aggregates and signals compared to luminal A and basal subtypes. Conversely, luminal A and basal subgroups displayed longer telomeres than the HER2 subgroup. Additionally, we observed heterogeneity within each molecular subtype concerning telomeric aggregates, signal number, and signal intensities. This heterogeneity was particularly pronounced in the luminal A subtype. Conclusion:This study proposes a categorization of breast cancer molecular subtypes based on the nuclear organization of telomeres. The identified differences in telomeric nuclear organization provide a new approach to refine the current molecular classification of breast cancer. This knowledge may contribute to improved understanding and management of breast cancer based on its distinct molecular subtypes.

Session Title: Cancer Poster Session III

PB5172 To investigate the mutational features of the immunoglobulin heavy chain variable region gene in patients with chronic lymphocytic leukemia using immunophenotypic and molecular genetic methods

Authors:

E. Coskunpinar[1], K. Gunduz[1], S. Erdem[2], M. Yenerel[2]; [1]Univ. of Hlth.Sci., Turkey, Istanbul, Turkey, [2]Istanbul Univ., Istanbul, Turkey

CD38 expression and Zeta chain (TCR) associated Protein kinase 70 kDa (Zap-70) has an importance in prognosis of chronic lymphocytic leukemia (CLL). Furthermore, studies have shown that the mutation status of non-mutated immunoglobulin heavy chain variable (IgVH) region can be used in order to predict the prognosis of CLL cases. Generally, the cases which are positive for Zap-70 and/or CD38, indicate a worse prognosis for CLL. However, IgVH mutation is not accompanied with positive results of Zap-70 and/or CD38. The aim of this study was to examine the IgVH mutation status and CD38, an immunophenotypic marker, of 10 patients diagnosed with CLL through whole exome sequencing (WES). In this research, presence of mutation of CD38 and IgVH, and specific IgVH fragments were examined in 10 CLL cases. CD38 was analysed with flow cytometric method, whereas IgVH mutations were analysed through WES (whole exome sequencing). For detection of the pathogenic variant and association of other variants to the disease, ClinVar database was used. The results were analysed with Iliom Bioscience program. As a result, 2 out of 10 cases were CD38 positive and 8 were negative. The cut-off value in CD38 evaluation was taken as 10. In 7 patients who were included in the research, 10 distinct pathogenic variants (RARS1, OBSCN, FAH, GAA, C11orf80, SLC7A7, PEPD, TEX14, TAF4B and STRA6) were detected. Moreover, in four patients, 4 possible pathogenic variants (AK2, SERPINF2, GNB5 and CD36) were detected. In 8 out of 10 cases, mutation was detected in different IgVH segments whereas in 2 cases, no IgVH mutation was found.

Session Title: Cancer Poster Session I

PB5173 *TP53* exon 6 and intron 6 microdeletions spanning splicing regulatory elements lead to exon skipping.

Authors:

D. Canson[1], I. Llinares[2], C. Fortuno[1], L. Sanoguera[2], M. De La Hoya[3], E. Bueno[2], A. B. Spurdle[1], E. A. Velasco[2]; [1]QIMR Berghofer Med. Res. Inst., Brisbane, Australia, [2]Inst. de Biología y Genética Molecular (CSIC-UVa), Valladolid, Spain, [3]IdISSC (Inst. de Investigación Sanitaria del Hosp. Clínico San Carlos), Hosp. Clinico San Carlos, Madrid, Spain

Most studies on splicing for *TP53* genetic variation have focused on variants that affect the donor and acceptor conserved dinucleotide positions. None have experimentally explored variants that disrupt splicing regulatory elements (SREs). We undertook a minigene construct-based study as a model for confirming the role of SREs in native *TP53* splicing. *TP53* exon 6, with a weak native donor (reference MES 2.6) flanked by putative exonic and intronic splicing enhancer elements, was prioritized for study.
Native and cryptic donor and acceptor splice motifs, and putative SRE-rich regions, in *TP53* exon 6 and intron 6 were mapped using HEXplorer. Bioinformatic prediction was performed for microdeletions spanning 3 exonic and 2 intronic SRE clusters, located outside of the native acceptor and donor splice motifs and upstream of a predicted strong cryptic donor (reference MES 9.6) positioned 64 bp downstream of the native donor. Microdeletions affecting SREs were designed such that none were predicted to create a new donor or acceptor motif. All deletions were predicted to disrupt splicing (SpliceAI ≥0.2 threshold), namely, exon 6 skipping resulting from exonic deletions and +64 cryptic donor activation resulting from intronic deletions. There is evidence for rare usage of the +64 cryptic donor as observed from 8/199,336 RNAseq samples in SpliceVault.
A 3.5 kb-insert with *TP53* exons 2 to 9 was cloned into the pMAD splicing vector (minigene mgTP53_ex2-9); 6 exonic and 4 intronic microdeletions were introduced into the minigene for assay in SKBR3 breast cancer cells. Minigene construct assays showed that all microdeletions abrogating enhancers without creating silencers led to exon 6 skipping (8-78%), with minor impact for intronic deletions (8-12%), demonstrating the utility of HEXplorer and SpliceAI in designing SRE-rich microdeletions for construct assays. In contrast to all predictions, none of the intronic deletions activated the +64 cryptic donor site, suggesting strong silencer activity affecting the cryptic site.
In summary, we effectively mapped and confirmed 5 SRE-rich regions between *TP53* exon 6 acceptor and the +64 cryptic donor in intron 6. The SREs surrounding the weak native donor explain its preferential usage over the predicted strong intronic cryptic donor. These findings demonstrate the importance of in vitro assays to confirm predictions of variant impact on splicing in SRE-rich regions. We are now using this SRE map to prioritize single nucleotide variants located within the microdeletions for splicing assays.

Session Title: Cancer Poster Session II

PB5174 Transcriptome-wide association study and causal inference analysis identifies colorectal cancer susceptibility genes with therapeutic potential.

Authors:

T. O'Mara[1], E. Hazelwood[2], P. Kho[3,1], X. Wang[1], T. Bishop[4], A. Chan[5,6,7,8], S. Gruber[9], J. Hampe[10], L. Le Marchand[11], M. Woods[12], Colon CFR, E. Vincent[13], D. Glubb[1]; [1]QIMR Berghofer Med. Res. Inst., Brisbane, Australia, [2]Univ. of Bristol, BRISTOL, United Kingdom, [3]Stanford Univ., Sunnyvale, CA, [4]Univ. of Leeds, Barnet, United Kingdom, [5]Massachusetts Gen. Hosp. and Harvard Med. Sch., Boston, MA, [6]Brigham and Women's Hosp. and Harvard Med. Sch., Boston, MA, [7]Broad Inst. of Harvard and MIT, Cambridge, MA, [8]Harvard T.H. Chan Sch. of Publ. Hlth., Harvard Univ., Boston, MA, [9]City of Hope Natl. Med. Ctr., Duarte, CA, [10]TU Dresden, UKD, Dresden, Saxony, Germany, [11]Univ Hawaii, Honolulu, HI, [12]Mem. Univ Newfoundland, St John's, NL, Canada, [13]Univ. of Bristol, Bristol, United Kingdom

Colorectal cancer is the third most common cancer worldwide. However, the biological mechanisms underlying disease development are unclear, posing challenges in the development of effective therapeutic and preventive interventions. Previous transcriptome-wide association studies (TWAS) have revealed several genes potentially implicated in colorectal cancer development. Nevertheless, existing TWAS studies have not yet stratified by anatomical subsite or sex, and the lack of robust causal inference analysis raises concerns about spurious associations among the identified genes.

We conducted a comprehensive investigation using two multi-tissue TWAS methods (S-MultiXcan and Joint Tissue Imputation) to maximise power to detect associations. Specifically, we integrated data from a large colorectal cancer genome-wide association study (GWAS; 52,775 cases and 45,940 controls) with expression quantitative trait loci data from six relevant tissues from GTEx v8 (transverse colon, sigmoid colon, subcutaneous adipose, visceral adipose, whole blood and EBV-transformed lymphocytes). Analyses were repeated stratifying the colorectal cancer GWAS by sex and anatomical subsite. In total, we identified 78 genes at 28 loci that significantly associated with colorectal cancer risk at a stringent Bonferroni threshold. Genes were further assessed for a causal relationship using Mendelian randomization and colocalization analyses. We also assessed the causal relationship between colorectal cancer risk and genes encoding 1,263 actionable proteins that are targeted by approved drugs or are in clinical development (i.e. the “druggable genome”).

Our workflow prioritized 35 genes with evidence for a causal relationship with colorectal cancer risk, including six genes encoding proteins that are targeted by available drugs. Prioritized genes were further evaluated for therapeutic potential by performing phenome-wide MR to identify possible side effects and exploring drug repurposing opportunities using the Connectivity Map and Open Target platforms.

In summary, our study provides valuable insights into the molecular underpinnings of colorectal cancer and identifies promising candidate genes with a causal role in disease etiology. Furthermore, our findings offer potential avenues for therapeutic intervention, including the repurposing of existing drugs and the development of novel treatments.

Session Title: Cancer Poster Session III

PB5175 Transcriptome-Wide Association Study of Mammographic Density Phenotypes

Authors:

J. Rothstein[1,2], A. Sistig[2], X. Song[2], S. E. Alexeeff[3], L. C. Sakoda[3], N. Achacoso[3], E. Jorgenson[4], V. McGuire[5,6], J. A. Lipson[5], D. L. Rubin[5], M. J. Yaffe[7], Markers of Density Consortium, Breast Cancer Association Consortium, D. F. Easton[8], R. M. Tamimi[9], C. M. Vachon[10], S. Lindström[11,12], N. Risch[6,3], A. S. Whittemore[5], R. J. Klein[2], L. A. Habel[3], P. Wang[2], W. Sieh[1,2]; [1]Univ. of Texas MD Anderson Cancer Ctr., Houston, TX, [2]Icahn Sch. of Med. at Mount Sinai, New York, NY, [3]Div. of Res., Kaiser Permanente Northern California, Oakland, CA, [4]Regeneron Genetics Ctr., Tarrytown, NY, [5]Stanford Univ. Sch. of Med., Stanford, CA, [6]Univ. of California, San Francisco, San Francisco, CA, [7]Univ. of Toronto, Toronto, ON, Canada, [8]Univ. of Cambridge, Cambridge, United Kingdom, [9]Weill Cornell Med., New York, NY, [10]Mayo Clinic, Rochester, MN, [11]Univ. of Washington, Seattle, WA, [12]Fred Hutchinson Cancer Res. Ctr., Seattle, WA

Background: Mammographic density (MD) phenotypes are highly heritable and strongly associated with breast cancer risk. Genetic variants identified by genome-wide association studies (GWAS) explain only a small fraction of the heritability, and the responsible genes remain largely unknown. Transcriptome-wide association studies (TWAS) can improve the power of GWAS and identify genes associated with MD through their genetically regulated expression levels.
Methods: The study population included 24,158 women of European ancestry who underwent screening with Hologic (n=20,282) or GE (n=3,876) digital mammography and participated in the Research Program on Genes Environment and Health (RPGEH) at Kaiser Permanente. Dense area (DA), non-dense area (NDA), and percent density (PD) were measured centrally using Cumulus6. Gene expression was estimated using PrediXcan models for mammary tissue, fibroblast cells, subcutaneous and visceral adipose tissues, and assessed for their associations with MD in linear regression models adjusted for age, BMI and other covariates. Tissue-specific results were combined, and genes that were significant at a false-discovery rate of 0.05 were carried forward for replication ($p<0.05$) in an independent GWAS of MD in up to 27,900 European ancestry women.
Results: In the discovery sample, 58 genes in 36 distinct regions were associated with MD. In the replication sample, a subset of 32 genes in 21 regions was associated with MD, including 8 novel genes in 7 regions. *LRRC17*, *PPP2R3A*, and *TNFSF12* were novel genes for DA. *KCNN4*, *NKX6-1*, *MYEOV* and *RP11-211G23.2* (both at 11q13.3) were novel genes for NDA. *SNX16* was a novel gene for PD. Among the replicated MD genes, 17 genes in 12 regions also were associated with breast cancer risk.
Conclusion: This TWAS identified novel genes for MD and breast cancer risk, and prioritized genes at known GWAS loci that may be causally associated with MD phenotypes through their expression levels.

Session Title: Cancer Poster Session II

PB5177 Tumor cellularity bias on interpretationof genomic data in pancreatic ductaladenocarcinoma

Authors:

D. Kim[1], W-G. Yun[2], J-Y. Jang[2], D. Park[1]; [1]Ajou Univ., Suwon, Korea, Republic of, [2]Seoul Natl. Univ. Coll. of Med., Seoul, Korea, Republic of

Backgroud: Tumor cellularity, the proportion of cancer cells within a sample, is a primary consideration before performing next-generation sequencing (NGS) in cancer research. Low cellularity, a majority of non-cancer cells within a tissue, can lead to misinterpretation of cancer genomics and transcriptomics. However, little perspective exists on how low cellularity decreases the accuracy of omics data. Here, we aimed to investigate the bias of tumor cellularity on genomic interpretation from NGS data in pancreatic ductal adenocarcinoma (PDAC), a representative cancer type with low cellularity (mean < 40%). **Methods:** We performed single nucleotide polymorphism (SNP) array, whole exome sequencing (WES), and RNA-seq on 196 surgically treated patients with PDAC between 2009 and 2016. In addition, *KRAS*-targeted sequencing data were generated from the 193 matched samples. A tumor cellularity was calculated from the SNP profile per tissue. Then, we divided the patients into high (≥ 0.26) and low cellularity groups (< 0.26) and compared these groups. **Results:** First, cellularity significantly affected the sensitivity to identify *KRAS* mutation in WES; a detection rate of 86.3% (88/102) and 54.3% (51/94) in high and low cellularity groups, respectively ($p < 0.001$). This low sensitivity issue was mostly resolved by *KRAS*-targeted sequencing as *KRAS* variants were present in 40 low cellularity patients out of 43 who had no *KRAS* mutation in WES. Second, discordant tumor cellularity between DNA and RNA based methods showed that various composition of non-cancer cells within the tissues resulted in heterogeneous gene expression by low cellularity. There was a high correlation between tumor purity estimated from SNP profiles and RNA-seq in the high cellularity group (*Pearson's correlation coefficient* $r = 0.45$, $p < 0.001$). In contrast, the low cellularity group had no positive correlation ($r = -0.03$, $p = 0.77$). Third, profiling cell type abundance through RNA-seq revealed the association of molecular subtypes with cellularity. More normal acinar and immune cells were detected in low cellularity samples, whereas malignant ductal cells were more abundant in high cellularity samples. As a result, aberrantly differentiated endocrine exocrine (ADEX) subtype, originating from normal acinar cells, accounted for 23% (23/102) and 50% (47/94) of high and low cellularity samples, respectively. **Conclusion:** Low tumor purity leads to misunderstanding the nature of cancer by disturbing mutation identification, accurate cellularity assessment, and molecular subtyping. Consequently, it is necessary to interpret cancer omics data considering cellularity.

Session Title: Cancer Poster Session III

PB5178 Tumor Heterogeneity in breast cancer by cellular Telogenomics

Authors:

C. Ongoiba[1], O. Samassekou[1], G. Landoure[2], M. Traore[1], M. Keita[1], M. Ly[1]; [1]USTTB, Bamako, Mali, [2]CHU du Point G, Bamako, Mali

Introduction: Tumor heterogeneity is a prominent characteristic of breast cancer, observed at the genomic, transcriptomic, and proteomic levels. The identification and understanding of tumor heterogeneity in breast cancer can serve as valuable diagnostic, prognostic, and therapeutic biomarkers, particularly for breast cancer subgroups such as the HER2-equivocal subgroup. However, accurately assessing tumor heterogeneity at the cellular level, while considering tumor architecture and cell morphology, remains challenging due to a lack of established methodologies. Therefore, our current study aims to develop a methodology using telomere nuclear organization, referred to as cellular telogenomics, to determine tumor heterogeneity at the cellular level in breast cancer.
Objective: The objective of this study is to determine cellular heterogeneity by employing cellular telogenomics.
Methods: Breast cancer samples were collected from patients presenting with the HER2-equivocal form. Three-dimensional (3D) fluorescence in situ hybridization using telomeric probes was performed on tissue sections. Subsequently, 3D microscopy was conducted, and six parameters (telomere numbers, telomere intensities, telomere aggregates, telomere distribution, telomere volume, and nuclear volume) defining cellular telogenomics were assessed in 400 cells from various regions of the sample. Finally, a heterogeneity score was developed through statistical analysis, phenotyping each cell based on the six parameters of cellular telogenomics.
Results: Cells were successfully grouped based on their telogenomics phenotype. It was observed that certain low-grade tumoral regions exhibited a higher index of tumor heterogeneity, whereas some high-grade tumoral regions demonstrated greater genomic homogeneity. Furthermore, cell-by-cell analysis revealed that telogenomics phenotyping can determine tumor cell evolution based on their telogenomic changes. Additionally, specific telogenomic phenotypes were associated with varying levels of tumor aggressiveness, with the aggressiveness being dependent on the cellular regions exhibiting a higher heterogeneity index.
Conclusion: This study introduces a novel approach for determining tumor heterogeneity at the cellular level. Our findings suggest that a higher index of cellular heterogeneity, as defined by cellular telogenomics, can serve as an indicator of tumor aggressiveness.

Session Title: Cancer Poster Session I

PB5179 Tumor Mutation Burden calculation from a targeted gene panel

Authors:

C. O'Connor[1], R. J. Martinez[1], M. Schomaker[2], R. LaRue[1], S. A. Munro[1], S. Yohe[1], A. C. Nelson[1]; [1]Univ. of Minnesota, Minneapolis, MN, [2]Fairview Molecular Diagnostics Lab., Minneapolis, MN

Tumor mutation burden (TMB) is an emerging biomarker for identifying patients likely to respond to immune checkpoint therapies. While the least biased approach to TMB calculation is to use whole exome sequencing (WES) of paired tumor-normal samples, time and cost constraints have led to a preference for using tumor-only targeted sequencing panels to calculate TMB. To calculate TMB from a targeted gene panel we need to quantify how reductions in panel size affect TMB accuracy. Friends of Cancer Research has published recommendations for evaluating the effect of panel size on TMB; most published examples of this use a large panel (at least 1 Mb in total gene sequence length). Decisions about which variants are counted as part of the TMB calculation has several considerations including removal of artifacts, driver mutations and putative germline mutations. Due to the emerging nature of TMB as a biomarker there are still a variety of methods used to remove putative germline variants and other sequence artifacts.

We used publicly available data from The Cancer Genome Atlas (TCGA) for four types of cancer to quantify the effect of a targeted gene panel on TMB estimates compared to WES, and utilized a TMB calibration panel to refine our variant filtering process. To develop our own TMB calling pipeline, we calculated TMB from two targeted gene panels: 127 genes (0.38 Mb) and 511 genes (1.37 Mb), and studied the effects of filtering variants based on their presence in population databases, presence in databases of common cancer driver mutations, presence in an in-house variant database, synonymous/non-synonymous status, variant allele frequency and ClinVar annotation status on TMB values. We found that our targeted gene panels tended to overestimate TMB, although there was high agreement (82% for the 127 gene panel) between WES and targeted gene panel TMB rates at a threshold of greater than or less than 10 mutations/Mb. Using the TMB calibration samples and TCGA samples for validation we developed a TMB calculation algorithm that accurately calculates TMB from only a targeted gene panel.

Session Title: Cancer Poster Session II

PB5180 Two-sample Mendelian randomization study of circulating metabolites and prostate cancer risk in Hispanic populations

Authors:

H. Fuller[1], R. Rhode[2], H. Highland[2], J. Shen[3], B. Yu[4], E. Boerwinkle[4], M. Grove[4], K. E. North[2], D. V. Conti[3], C. A. Haiman[3], K. Young[2], B. F. Darst[1]; [1]Fred Hutchinson Cancer Ctr., Seattle, WA, [2]Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, [3]Univ. of Southern California, Los Angeles, CA, [4]Univ. of Texas Hlth.Sci. Ctr. at Houston, Houston, TX

While prostate cancer (PCa) is highly heritable, the mechanisms underlying PCa risk are not well understood, particularly in underserved populations. Here, we conducted a two-sample Mendelian randomization (MR) to assess whether serum metabolites are causally associated with PCa risk in Hispanic/Latino men. MR analyses were performed using the inverse variance weighted (IVW) method on GWAS summary statistics for 711 metabolites quantified by untargeted mass spectroscopy via the Metabolon platform for 3,166 Hispanic Community Health Study/ Study of Latinos participants and PCa GWAS summary statistics for 3,931 cases and 26,405 controls from Hispanic populations in PRACTICAL. SNPs associated with metabolites at the genome-wide significance level ($P<5x10^{-8}$) were included in the instruments after removing rare SNPs (minor allele frequency≤0.01) and pruning by linkage disequilibrium ($R^2=0.2$), which was calculated in European and African ancestry TOPMed populations in TOP-LD. A false discovery rate ($\alpha=0.05$) was implemented to account for multiple testing. A range of sensitivity analyses were utilized to assess violations of MR assumptions (IVW multiplicative, weighted mode, weighted median, MR egger, and MR PRESSO). In total, 22 metabolites were significantly associated with PCa risk, including 3 amino acids, 1 carbohydrate, and 18 lipids, which included 4 polyunsaturated fatty acids. All 4 polyunsaturated fatty acids were associated with 14-19% reduced odds of PCa: n3 DPA (22:5n-3) (OR=0.81, 95% CI=0.73-0.90, $P_{IVW\ random}=1.7x10^{-4}$), n6 DPA (22:5n-6) (OR=0.86, 95% CI=0.78-0.95, $P_{IVW\ random}=2.2x10^{-3}$), EPA (20:5n-3) (OR=0.81, 95% CI=0.77-0.85, $P_{IVW\ random}=2.8x10^{-15}$), and arachidonate (20:4n-6) (OR=0.85, 95% CI=0.82-0.88, $P_{IVW\ random}=3.3x10^{-20}$). Furthermore, increased levels of acyl choline arachidonoylcholine, also involved in fatty acid metabolism, were associated with 16% reduced PCa odds (OR=0.84, 95% CI=0.79-0.89, $P_{IVW\ random}=3.5x10^{-9}$). The most significant associations ($P_{IVW\ random}\leq3.9x10^{-40}$) observed were phosphatidylcholines 1-stearoyl-2-arachidonoyl-GPC (18:0/20:4) and 1-palmitoyl-2-arachidonoyl-GPC (16:0/20:4n-6), both of which were associated with ~15% reduced odds of PCa. Results were largely robust to sensitivity analyses. This large-scale MR study provides evidence of causal associations between a range of metabolites and PCa risk in individuals of Hispanic ethnicity. Work is ongoing to replicate these findings in individuals of European and African ancestry, conduct multi-ancestry MR for discovery, and utilize multivariate MR to further assess the potential for metabolites to serve as PCa biomarkers.

Session Title: Cancer Poster Session III

PB5181 Unified somatic calling and machine learning-based classification enhance the discovery of clonal hematopoiesis of indeterminate potential

Authors:

S. Tian[1], G. Jenkinson[2], A. Ferrer[3], H. Yan[1], J. Morales-Rosado[3], K. Wang[1], T. Lasho[4], S. Baheti[1], J. Olson[5], L. Baughn[6], M. Patnaik[4], W. Ding[3], K. Lazaridis[7], E. Klee[1]; [1]Div. of Computational Biology, Mayo Clinic, Rochester, MN, [2]Div. of Computational Biology, Mayo Clinic, ROCHESTER, MN, [3]Div. of Hematology, Mayo Clinic, Rochester, MN, [4]Div. of Hematology, Mayo Clinic, ROCHESTER, MN, [5]Div. of Epidemiology, Mayo Clinic, ROCHESTER, MN, [6]Dept. of Lab. Med. and Pathology, Mayo Clinic, Rochester, MN, [7]Div. of Gastroenterology & Hepatology, Mayo Clinic, Rochester, MN

Clonal hematopoiesis (CH) of indeterminate potential (CHIP), driven by somatic mutations in leukemia-associated genes, confers increased risk of hematologic malignancies, cardiovascular disease, and all-cause mortality. In blood of healthy individuals, small CH clones with competitive advantage can expand over time to reach 2% variant allele frequency (VAF), the current threshold for CHIP. Nevertheless, reliable detection of low-frequency CHIP mutations requires deep targeted sequencing which is costly and not scalable. In particular, no bioinformatics pipeline has been specifically developed for the discovery of CHIP mutations. Herein, we present a streamlined variant detection and refinement workflow, **UNI**fied **SO**matic calling and **M**achine learning-based classification, or UNISOM for short, to enhance CHIP discovery from whole-genome and whole-exome sequencing data that are underpowered, especially for low VAFs, due to insufficient coverage and inherent sequencing error. UNISOM utilizes a meta-calling strategy to achieve high sensitivity in variant detection, which is an ensemble of three sensitive tools benchmarked over simulated genomes with CHIP spike-in, in couple with XGBoost-based machine learning models built with variant-associated features. The model classifies annotated variants into CHIP, germline and artifact, thus minimizing the time-consuming manual inspection needed for CHIP refinement. UNISOM achieved good recall in a cohort of 25 patients with whole-exome data in Mayo Clinic CHIP clinic, recovering nearly 80% of the CHIP mutations identified in the same cohort via deep targeted sequencing. To demonstrate the utility of UNISOM in large-scale CHIP discovery, we applied UNISOM to whole-genome data from 979 individuals in the Mayo Clinic Biobank, UNISOM recapitulated the patterns previously established in much larger cohorts, including the most mutated genes, predominant mutation types and signatures, as well as strong associations of CHIP with age, TERT polymorphisms and smoking status. Most notably, 30% of the identified CHIP mutations had VAFs below 5%, demonstrating its high sensitivity toward small mutant clones. This workflow is applicable to the CHIP screening in population genomics studies. In clinic practice, implementing sensitive tools like UNISOM will improve the early detection of low-VAF CHIP mutations prior to diagnosis, enabling monitoring of the carriers and possible intervention. In particular, UNISOM is highly sensitive to CHIP with low VAFs, such as those occurring in *TP53*.

Session Title: Cancer Poster Session I

PB5182 Unraveling inequality: UCSF 500's path to universal precision oncology.

Authors:

A. Pinard, C. Badduke, L. Zhang, C. Onodera, R. Sono, J. McGuire, A. Griffin, P. Devine, J. Van Ziffle, A. Rajkovic; Univ. of California, San Francisco, San Francisco, CA

Since 2015, the University of California, San Francisco (UCSF) Clinical Cancer Genomics Laboratory (CCGL) has provided the UCSF500 cancer genetic test, a capture-based next-generation sequencing (NGS) assay in its third version currently targeting 529 cancer genes and select introns of 73 genes. These results can aid clinicians in personalizing cancer treatments, identifying suitable clinical trials, or clarifying a patient's exact cancer type. In pursuit of equitable patient care, we examine its utilization across racial and ethnic groups among cancer patients treated at UCSF to uncover and eradicate any potential disparities in accessing UCSF500. Over the past 5 years (07/01/2016-06/30/2021), we collected 3,002 UCSF500 cases, and analyzed them based on patients' race/ethnicity and juxtaposed them to the UCSF Cancer Registry 29,052 active cases. Patients' characteristics were summarized by descriptive statistics. In general, frequency and percentage were used to summarize categorical variables. Chi-squared test was applied to test if there was a statistical association between two categorical variables. The statistical significance was declared at alpha <0.05. All the analysis was done by R. Approximately 10% of the patients from the cancer registry had UCSF500 performed. Significant disparities were identified. Overall, in male population, White Non-Hispanic (WNH) were undertested than expected whereas Asian/Pacific Islander and Hispanic were over tested but no difference in female population. When examining each cancer type independently, it was in respiratory system (African American (AA) underrepresented), urinary system (Hispanic underrepresented), bone and soft tissue (Hispanic overrepresented), blood and bone marrow (Hispanic overrepresented) and genital cancers (AA underrepresented) in male and in skin (WNH underrepresented and Hispanic overrepresented), blood and bone marrow (WNH underrepresented and Hispanic overrepresented) and genital (AA overrepresented) cancers in female. The project also revealed a complexity of data analysis, different tumor types and prevalence of cancer by racial/ethnic groups and need for age stratification. We strongly believe that algorithmizing of when to perform tumor testing and to whom it should be offered is essential to reduce waste, to standardize care, and to prevent racial/ethnic disparities. Consequently, we developed two reflex testing algorithms for lung and pelvic cancers and one pilot testing of NGS approach for Lynch syndrome screening and molecular classification of endometrial cancers. Future analysis will assess the utility of such an approach.

Session Title: Cancer Poster Session II

PB5183 Unraveling the impact of variants of unknown significance in breast cancer risk modulation within the context of polygenic background

Authors:

y. jeong[1], J. Sung[1], J. Hopper[2], T. L. Nguyen[2], J. Lee[3], Y. Lee[3], T. Yoo[3], S. Lee[3]; [1]Genome & Hlth.Data Lab, Sch. of Publ. Hlth., Seoul Natl. Univ., Seoul, Korea, Republic of, [2]Ctr. of Epidemiology and Biostatistics, Melbourne Sch. of Population & Global Hlth., The Univ. of Melbourne, Melbourne, Australia, [3]Dept. of Surgery, Asan Med. Ctr., Univ. of Ulsan Coll. of Med., Seoul, Korea, Republic of

Background : Growing evidence suggests an interaction between high-penetrance genetic variants and individual's polygenic scores (PGS) in determining breast cancer (BC) risk. However, the effects of variants of unknown significance (VUS) in conjunction with PGS remain poorly understood. Goal : This study aims to investigate the modification of BC risk conferred by VUSs in five hereditary BC predisposing genes (*BRCA1/2, ATM, PALB2, CHEK2*) as identified by the Breast and Ovarian Analysis of Disease Incidence and Carrier Estimation Algorithm (BOADICEA), in the context of different PGS backgrounds. Methods & Results : We analyzed a dataset comprising 18,066 BC cases and 137,852 controls with no history of cancer from the UK Biobank's cohort of white females. The UK Biobank-enhanced version of BC-PGS was derived using 9,553,208 variants, revealing a significant distribution difference between cases and controls (AUROC 0.8033). VUSs were defined using ClinVar and ENIGMA for *BRCA1/2*, while missense and splice region variants were utilized for *ATM, PALB2*, and *CHEK2*. All genes exhibited higher proportions of VUSs
(*BRCA1* 4.26%, *BRCA2* 7.34%, *ATM* 15.12%, *PALB2* 2.23%, *CHEK2* 2.47%) compared to pathogenic or loss-of-function (frameshift, start lost, stop gain) variants
(*BRCA1* 0.1%, *BRCA2* 0.3%, *ATM* 1.83%, *PALB2* 0.17%, *CHEK2* 0.57%). To evaluate the increase in BC risk associated with VUSs based on BC-PGS, we calculated the BC proportion of VUS carriers in each BC-PGS quantile (Low: 0-5%, Intermediate: 5-95%, High: 95-100%), and compared it to the BC proportion of non-carriers. Individuals in the highest BC-PGS quantile (95-100%) showed significantly elevated BC risk increase due to VUSs (*BRCA1* 4.57%, *BRCA2* 4.29%, *ATM* 2.19%, *PALB2* 6.46%, *CHEK2* 6.51%) compared to the low (0-5%) and intermediate (5-95%) quantiles (*BRCA1* 0.25%, 1.7%; *BRCA2* 0.42%, 1.79%; *ATM* 0.36%, 1.50%; *PALB2* 3.63%, 3.01%; *CHEK2* 3.1%, 4.21%). Discussion : Our findings demonstrate that, similar to Tier1 variants, the risk of VUSs interact with the polygenic background. Given the higher frequency of VUSs and substantial increase in risk in highest PGS (95-100%) group, our findings suggest the importance of broadening the current list of pathogenic variants and incorporating polygenic background in prevention and genetic counselling of breast cancer.

Session Title: Cancer Poster Session III

PB5184 † Unveiling Hidden Associations: Leveraging Continuous Tumor Dynamics to Discover Novel Somatic Mutations Influencing Tumor Progression and Treatment Response in Precision Oncology

Authors:

E. Petter[1], J. LoPiccolo[2], S. Groha[2], K. L. Kehl[2], B. Pasaniuc[1], A. Gusev[2], N. Zaitlen[1]; [1]UCLA, Los Angeles, CA, [2]Dana-Farber Cancer Inst., Boston, MA

A central goal of personalized oncology is to identify specific tumor mutations affecting prognosis and treatment response. Studies to date have concentrated on single endpoints, such as overall survival and progression-free survival. Here, we present a novel method to model longitudinal tumor dynamics and identify associations between somatic mutations, treatments, and patient outcomes. We hypothesize that integrating continuous tumor progression and treatment response data, rather than solely relying on single endpoints, will enhance the power and resolution of personalized oncology research.

Our study focuses on ~9,000 pan-cancer patients at the Dana Farber Cancer Institute (DFCI) for which we obtained time-dependent outcomes capturing cancer progression and treatment response from radiology reports. Coupled with tumor sequencing, clinical phenotypes, and treatment history, this data provides an unprecedented opportunity to study the impact of somatic mutations on cancer progression and treatment response.

Rather than relying on only ~5,000 patient death dates, our method builds from the Andersen-Gill model for time-dependent outcomes, in order to additionally account for over 40,000 tumor progression and treatment response events in these same patients. We further put into use time-dependent treatment information collected from the Electronic Health Record (EHR), to identify interactions between somatic mutations and treatment. Specifically, we explored interactions with immunotherapy, targeted therapies, and common chemotherapy.

Our method detected 298 associations between somatic mutations and patient outcomes, across all cancer types and treatments, out of which 77 are only detectable when using the time-dependent outcomes for progression or response. This represents a 34% increase in discovered associations compared to single endpoint analysis. Furthermore, when examining time-dependent treatment interactions in lung cancer, we identified 35 significant interactions (5% FDR corrected) with HR ranging from 0.27-1.73 between somatic status and drug group. These interactions reflect somatic mutations that alter the rate of tumor progression or treatment response under a specific regime, and represent a potential route to sequencing guided treatment.

By leveraging longitudinal data and novel time-dependent modeling approaches, our research uncovers additional signals that are often overlooked in traditional approaches. We propose our method as a valuable tool in aiding the development of precision oncology and detecting candidate genes for the development of new companion diagnostic-guided therapies.

Session Title: Cancer Poster Session I

PB5185 Unveiling Inter- and Intra-Tumor Heterogeneity in Ductal Carcinoma *in Situ* (DCIS) and Invasive Ductal Carcinoma (IDC) through Integration of Unified Single-Cell Whole Genome Copy Number and RNA Expression Data

Authors:

T. Tate[1], I. Salas-Gonzalez[1], K. Kennedy[1], J. Remington[1], D. Arvapalli[1], S. Velivela[1], J. R. Marks[2], E. Hwang[2], V. Weigman[1], J. Zawistowski[1]; [1]BioSkryb Genomics, Durham, NC, [2]Dept. of Surgery, Duke Univ. Med. Ctr., Durham, NC

Introduction: Ductal Carcinoma in situ (DCIS) is a precursor to invasive ductal carcinoma (IDC) with complex heterogeneity influenced by genomic alterations and the tumor microenvironment. This study aims to uncover gene dysregulation mechanisms in breast cancer patients with a history of DCIS, using a unified single-cell copy number and RNA expression workflow. Methods: DCIS and IDC samples were collected from 11 patients post-mastectomy. Cells were isolated and either enriched for EpCAM using FACS or analyzed without enrichment. Unified whole genome and transcriptome amplification were performed on approximately 100 cells per sample (1225 cells in total) using ResolveOME™. DNA libraries were enriched for exomes and data was analyzed using the BaseJumper™ platform. Results: Our analysis of 11 patient biopsies provide evidence of potential characteristics underlying tumoral evolution in IDC. We observed vast heterogeneity among the tumor profiles, with some displaying a more quiescent nature. Across the 11 profiles, a range of 1.5-17% copy number events were identified, carrying potential genomic consequences. Importantly, these findings confirmed known DCIS copy number profiles, including 1q gains in 7 out of 11 cases and 16q loss in all cases. We identified additional copy signatures occurring in low proportions of patients. Notable alterations included 17p loss in 5 of 11 patients, chromosome 7 gain in 4 of 11 patients and 11q loss in 6 of 11 patients. These characteristic copy number alterations are associated with key genes, involved in cell cycle regulation and progression (CCND1, CKS1B), regulation of cell growth and proliferation (BRAF, NF1), migration (BRAF, PTK2), immune response (JAK1), and tumor invasion and metastasis (MMP7). Conclusions: Our findings emphasize the assorted copy number profiles associated with invasive breast cancer and highlight the potential involvement of key genes in tumoral evolution. By elucidating these mechanisms at the single-cell level, we gain valuable insights into the underlying processes driving tumoral evolution from DCIS to IDC, providing implications for personalized treatment strategies by identifying aggressive clonal cell populations and expanding treatment options.

Session Title: Cancer Poster Session II

PB5186 Unveiling Susceptible Transcription Factors and Novel Risk Genes in Colorectal Cancer: Integrating Large-Scale Genetic and Omics Data Analysis

Authors:

Z. Chen[1], W. Song[1], W. Wen[1], M. Devall[2], C. Dampier[2], F. Moratalla-Navarro[3], X-O. Shu[1], J. Long[1], J. Huyghe[4], S. Schmit[5], the Asia Colorectal Cancer Consortium (ACCC), L. Van Kaer[1], L. Wu[1], L. Hsu[4], Q. Cai[1], M. Dunlop[6], S. Gruber[7], R. Houlston[8], I. Tomlinson[9], U. Peters[4], V. Moreno[3], G. Casey[2], W. Zheng[1], X. Guo[1]; [1]Vanderbilt Univ. Med. Ctr., Nashville, TN, [2]Univ. of Virginia, Charlottesville, VA, [3]Catalan Inst. of Oncology (ICO), Barcelona, Spain, [4]Fred Hutchinson Cancer Ctr., Seattle, WA, [5]Cleveland Clinic, Cleveland, OH, [6]Univ. of Edinburgh, Edinburgh, United Kingdom, [7]City of Hope Natl. Med. Ctr., Duarte, CA, [8]Inst. of Cancer Res., London, United Kingdom, [9]Edinburgh Cancer Res. Ctr., Edinburgh, United Kingdom

Genome-wide association studies (GWAS) and subsequent functional experiments have provided strong evidence that risk regulatory variants can disrupt DNA binding affinities of transcription factors (TFs), and consequently leading to altered gene expression and an increased risk of cancer. Identifying these susceptible TFs is crucial for understanding the mechanisms of transcriptional dysregulation in cancer development. Moreover, our recent transcriptome-wide association studies (TWAS) approach, sTF-TWAS, by integrating genetic variants occupied by susceptible TFs, has demonstrated significant improvement in risk gene detection with increased statistical power and accuracy. However, the investigation of susceptible TFs and the integration of their occupancies with sTF-TWAS for risk gene discovery remains largely unexplored in colorectal cancer (CRC). In this study, we integrated TF ChIP-Seq datasets (n=219) generated from CRC-related cells, along with large GWAS data consisting of 100,204 CRC cases and 154,587 controls of European and East Asian ancestries. We used generalized mixed models (PMID: 34518541) to estimate the associations between the Chi-squared values (from GWAS summary statistics) and the binding status of TFs to genetic variants. At a Bonferroni-corrected $P < 0.05$, we identified 51 susceptible TFs, including CRC development-related TFs VDR, MYC, JUN, CDX2, NIPBL, MED12, and ETV5. Furthermore, we observed significant interactions among 151 pairs of TFs at $P < 3.9 \times 10^{-5}$ (0.05/1,275 possible TF pairs from 51 TFs), where genetic variants co-occupied by these TF-TF pairs exhibited a higher CRC risk compared to variants occupied by single TFs. Using our sTF-TWAS framework, we built prediction models for gene expression, alternative splicing (sp), and alternative polyadenylation (apa) by incorporating genetic variants occupied by the 51 identified susceptible TFs. These models were trained on transcriptomic data from normal colon tissues of European descendants (n=707) from the BarcUVa-Seq and GTEx projects, and East Asian descendants (n=364) from ACCC. At a Bonferroni threshold, we identified 153 CRC putative susceptibility genes, combining results from sTF-TWAS (n=116), sp-sTF-TWAS (n=43), and apa-sTF-TWAS (n=16). Notably, among these genes, 69 were not previously associated with CRC susceptibility, including 24 genes located at 21 loci that were at least 1Mb away from GWAS-identified risk variants. Our study identifies numerous susceptible TFs and their potentially regulated novel susceptibility genes in CRC and provides valuable insights into the molecular mechanisms driving this common cancer.

Session Title: Cancer Poster Session III

PB5187 Von Hippel-Lindau syndrome - a case report

Authors:

H. Singh, P. Kaur, I. Panigrahi, A. kaur, A. Arora; PGIMER, Chandigarh, India

INTRODUCTION: VHL syndrome is characterized by presence of CNS and retinal hemangioblastomas, renal cell carcinoma and renal cyst, pheochromocytoma, and tumors involving endolymphatic sac. The *VHL* gene is a tumor suppressor gene located on the short arm of chromosome 3 and has three exons encoding the VHL protein. **CASE DETAILS:** A 27 year-old male first born out of non consanguineous marriage, presented with hearing loss at the age of 17 years and was diagnosed to have Meniere disease for which grommet insertion was done. At the age of 22 years he had complaints of blurring of vision and was found to have bilateral retinal hemangioblastoma with retinal detachment right side for which pars plana vitrectomy was done. Magnetic resonance imaging of brain showed hyperintense lesion 20X20X30 mm in cortical and subcortical region of right lobe of cerebellum. Contrast enhanced computed tomography of abdomen showed few well defined hypodense non enhancing lesion seen in pancreas largest 30 mm in pancreatic tail and well defined 20X20 milimetres hypodense non enhancing non calcified lesion in mid pole of left kidney. Urine metanephrine screening test for pheochromocytoma showed normal excretion of metanephrines with value being 42microgram/24 hours (< 350). Ultrasonography of scrotum showed bilateral epididymal head cyst and bilateral mild varicoceles. With above constellation of findings suggestive of VHL syndrome clinical exome sequencing was done and a heterozygous missense variation in exon 1 of *VHL* gene on chromosome 3: *VHL*:c.C333G:p.(Ser111Arg) was observed and this has been previously reported as pathogenic for VHL syndrome. **CONCLUSION**: Identification of gene mutations enables appropriate prognostication, genetic counseling and further presymptomatic screening in at risk individuals in family.

Session Title: Cancer Poster Session I

PB5188 Whole exome sequencing of Moroccan prostate cancer samples identifies novel prostate cancer genes and several well-known signaling pathways.

Authors:

C. Samtal[1], M. Bendall[2], A. Wong[2], S. Ambs[3], N. H. Lee[4], **K. A. Crandall**[2], H. Bekkari[1]; [1]Lab. of Biotechnology, Environment, Agri-food and Hlth., Faculty of Sci. Dhar El Mahraz, Sidi Mohammed Ben Abdellah Univ., Fez, Morocco, [2]Computational Biology Inst., Milken Inst. Sch. of Publ. Hlth., George Washington Univ., Washington DC, DC, [3]Lab. of Human Carcinogenesis, Ctr. for Cancer Res., Natl. Cancer Inst., Bethesda, MD, [4]Dept. of Pharmacology and Physiology, George Washington Univ., Washington DC, DC

Prostate cancer (PCa) is the second cause of cancer death among Moroccan men. The data reported by Rabat registry indicate that the incidence rate for PCa similarly increases sharply with age after 50 years. After 75 years, the incidence reaches more than 274 new cases per 100.000 inhabitants per year. However, there are no genomic studies of this disease in the Moroccan population. Our study aims at identifying genomic features of prostate cancer in Moroccan men using Whole Exome Sequencing (WES). This study is the first in Morocco and one of the few attempts in the African continent using WES to analyze PCa. We analyzed the exomes of 12 Moroccan prostate cancer patients. Prostate cancer tumors and non-tumor adjacent tissue samples were collected from 2 sites in Central and Northeast Morocco. DNA was extracted from Formalin-Fixed Paraffin-Embedded (FFPE) and frozen tissue samples. After sequencing, samples were processed using Genome Analysis Toolkit (GATK) best practices. Somatic variant analysis of matched tumor and normal sequencing data lead to the identification of genes known to be associated with prostate cancer, and novel prostate cancer genes that were initially associated with other types of cancers. We identified novel variants that were not found in either of the following databases: COSMIC, dbSNP, and ClinVar. Our analysis identified several well-known signaling pathways that are involved in prostate cancer progression and tumorigenesis and are clinically targetable including *RTK-RAS*, *NOTCH*, *WNT*, *PI3K*, and *MYC* pathways. These findings represent the first and most comprehensive characterization of the Moroccan PCa exome to date. Because North African are underrepresented in genomic studies, our findings address a gap in the contribution of genetic variation to the incidence of prostate cancer and aggressiveness of the disease in the Moroccan and all North African populations.

Session Title: Cancer Poster Session II

PB5189 Yield of integrated DNA and RNA analysis of hereditary cancer associated genes based on cancer diagnosis.

Authors:

N. Kamps-Hughes[1], B. Heald[1], S. Nielsen[2], E. Borras[1], H. Kang[3], K. Nykamp[4], R. Nussbaum[1], E. D. Esplin[1], **B. Johnson**[5], D. Pineda-Alvarez[6], L. Fresard[1]; [1]Invitae, San Francisco, CA, [2]Invitae, Chicago, IL, [3]Invitae Corp., San Francisco, CA, [4]Invitae, Berkeley, CA, [5]Invitae, Fort Lauderdale, FL, [6]Invitae, Wellington, FL

Background RNA sequencing can help classify variants predicted to alter RNA splicing, as well as discover splice-altering variants in regions outside of the standard reportable range of targeted DNA panels. Prior research from our group showed that RNA sequencing impacted variant interpretation for 6.3% of 20,317 patients undergoing multi-cancer gene panel testing. The intronic variant discovery rate was 0.2%. Here we report the resolution of potential splicing variants and the discovery of intronic variants using RNA sequencing stratified by patients' reported cancer history.

Methods RNA sequencing was performed on cDNA from leukocyte mRNA for 63 transcripts from an 84 gene multi-cancer panel. RNA data were used as functional evidence in the Sherloc classification algorithm to reclassify potential splicing variants of uncertain significance to benign/likely benign (B/LB) or pathogenic/likely pathogenic (P/LP). Aberrant splicing was also used to discover variants outside of the reportable DNA range (all coding exons ±20bp of flanking intronic sequence). Cancer history was determined by clinician-reported data on the test requisition form. Data were stratified per no history of cancer, history of any cancer, and history of breast, prostate, colorectal, or pancreatic cancer. Descriptive statistics and Chi-square with Yates' correction were utilized, and significance was set at $p < 0.05$.

Results The cohort consisted of 61,388 patients. No cancer was reported in 43,285 (70.5%) patients, while 18,103 (29.5%) had any type of cancer (including 9,282 [15.1%] with breast, 1,405 [2.2%] with prostate, 2,393 [3.9%] with colorectal, and 1,054 [1.7%] with pancreatic cancer). Overall, RNA sequencing impacted variant interpretation for 1,448 (2.4%) of patients with 1,243 (2.0%) downgraded to B/LB and 213 (0.4%) upgraded to P/LP. There were no differences in the effect of RNA sequencing for variant interpretation based on no reported cancer history (2.4%), any cancer history (2.4%), breast (2.5%), prostate (2.4%), colorectal (2.5%), or pancreatic (2.6%) cancer. Additionally, no differences in variant discovery were observed based on no reported cancer history (0.1%), any cancer history (0.2%), breast (0.2%), prostate (0.1%), colorectal (0.1%), or pancreatic (0.3%) cancer.

Conclusion In this study, the impact of RNA sequencing on variant interpretation was 2.4%, with the majority of variants of uncertain significance being downgraded to B/LB. The effect of RNA sequencing did not differ based on the patient's reported cancer history.

Session Title: Cancer Poster Session III

PB5190 A multi omics classifier for prediction of androgen deprivation treatment response in prostate cancer patients

Authors:

I. Isewon, E. Alagbe; Covenant Univ., Ota, Ota, Nigeria

Clinical decision and treatment strategies have been based on the experience of a care giver as well as the pathophysiology of the disease. This evidence-based medicine strategy focuses on groups of people rather than individuals but is insufficient since treatment response differs between individuals. There has been a general move towards precision in oncology, notably due to the severe cytotoxic side effects of "one size fits all" cancer therapies, and identification of tumor-specific vulnerabilities. There is need to leverage on advances in high throughput sequencing and machine learning (ML) to improve treatment options for each patient. In this study, we developed a two-step classifier to predict the response of prostate cancer (PCa) patients to androgen treatment leveraging on high throughput multi-omics datasets provided by The Cancer Genome Atlas (TCGA). The TCGA PCa phenotype dataset contains records for 623 patients with 120 patients having missing values for primary treatment outcomes. Four PCa omics datasets were used: copy number variation, RNAseq, miRNAseq, and reverse phase protein array. Amongst the 120 patients with missing phenotypes, 68 had records for the four omics considered. Our investigation using ten ML algorithms reveals that tree-based algorithms such as decision tree, random forests, extreme gradient boosting, and gradient boosting machine had better predictive performance than probabilistic models such as naïve bayes and kernel-based methods such as support vector machines for the dataset. We also investigated the performance of all possible omics combinations. Our results show that there is an overall increase in performance when multiple omics datasets are used in contrast to single omics strategies. We have predicted for the first time, possible treatment response outcomes for 68 prostate cancer patients with missing treatment outcome phenotype values in TCGA with the following distribution: complete response (45), partial response (7), progressive disease (5) or stable disease (11).

Session Title: Cancer Poster Session I

PB5191 CRISPR/Cas9 Engineering of 3D Genome Structural Variants: Unveiling Predictive Models of Oncogene Activation in Cancer Genomes

Authors:

Z. Xu[1], D-S. Lee[2], S. Chandran[1], V. T. Le[1], R. Bump[1], J. Yasis[1], S. Dallarda[1], S. Marcotte[1], B. Clock[1], N. Haghani[1], C-Y. Cho[1], K. Akdemir[3], S. Tyndale[1], P. Futreal[3], G. McVicker[1], G. M. Wahl[1], J. R. Dixon[1]; [1]Salk Inst. for Biological Studies, La Jolla, CA, [2]Univ. of Seoul, Seoul, Korea, Republic of, [3]MD Anderson Cancer Ctr., Houston, TX

Higher order chromatin structure plays a critical role in the regulation of genes by distal regulatory sequences. Structural variants that alter 3D genome organization can lead to enhancer-promoter rewiring and human disease, particularly in the context of cancer. However, it remains unclear how widespread structural variants that alter 3D genome structure are in cancer genomes and what genes are affected by such events. Furthermore, recent studies have shown that only in a small minority of structural variants are associated with altered gene expression. Therefore, it is unclear whether or how an individual structural variant may contribute to oncogene activation. To address these questions, we have analyzed Hi-C data and structural variants from 92 cancer cell lines and patient samples representing diverse cancer types. We identified loci affected by recurrent alterations to 3D genome structure, including loci containing oncogenes such as *MYC*, *TERT*, and *CCND1*. We also find evidence that these loci are frequently affected by structural variants predicted to alter 3D genome structure from whole genome sequencing datasets. Using CRISPR/Cas9 genome engineering to generate *de novo* structural variants, we show that "Activity-by-Contact" models predict the likelihood of oncogene activation in the context of structural variants. However, such Activity-by-Contact models are only predictive of specific subsets of genes in the genome, suggesting that different classes of genes engage in distinct modes of regulation by distal regulatory elements. These results indicate that structural variants that alter 3D genome organization are widespread in cancer genomes and begin to illustrate predictive rules for the interpretation of the consequences of non-coding structural variants on oncogene activation.